KAUNAS UNIVERSITY OF TECHNOLOGY

ARNAS SURVYLA

NOVEL SOFT SENSORS FOR BIOPROCESS STATE ESTIMATION

Doctoral dissertation
Technological Sciences, Informatics Engineering (T 007)

2024, Kaunas

This doctoral dissertation was prepared at Kaunas University of Technology, Faculty of Electrical and Electronics, Department of Automation during the period of 2019–2023. The studies were supported by the Research Council of Lithuania.

The doctoral right has been granted to Kaunas University of Technology together with Vilnius Gediminas Technical University (Vilnius Tech).

Scientific Supervisor:
Prof. dr. Renaldas URNIEŽIUS (Kaunas University of Technology, Technological Sciences, Electrical and Electronic Engineering, T 001, Informatics Engineering, T 007).

The dissertation was edited by: English language editor Dr. Armandas Rumšas (Publishing House *Technologija*), Lithuanian language editor Aurelija Gražina Rukšaitė (Publishing House *Technologija*).

The public defense of the dissertation will be held at 9 a.m. on 7 May, 2024 at the public meeting of Dissertation Defense Board of Informatics Engineering Science Field in Rectorate Hall at Kaunas University of Technology.

Address: K. Donelaičio 73-402, Kaunas, LT-44249, Lithuania
Phone: (+370) 608 28 527; e-mail doktorantura@ktu.lt

The doctoral dissertation was sent out on 5 April, 2024.

The doctoral dissertation is available on the internet at http://ktu.edu and at the libraries of Kaunas University of Technology (Gedimino 50, Kaunas, LT-44239, Lithuania), Vilnius Gediminas Technical University (Vilnius Tech) (Saulėtekio 14, Vilnius, LT-10223).

KAUNO TECHNOLOGIJOS UNIVERSITETAS

ARNAS SURVYLA

INOVATYVŪS PROGRAMINIAI JUTIKLIAI, SKIRTI BIOPROCESO BŪSENAI ĮVERTINTI

Daktaro disertacija
Technologijos mokslai, informatikos inžinerija (T 007)

2024, Kaunas

Disertacija rengta 2019–2023 metais Kauno technologijos universiteto Elektros ir elektronikos fakultete, Automatikos katedroje. Mokslinius tyrimus rėmė Lietuvos moklso taryba.

Doktorantūros teisė Kauno technologijos universitetui suteikta kartu su Vilniaus Gedimino technikos universitetu (Vilnius Tech).

Mokslinis vadovas:
prof. dr. Renaldas URNIEŽIUS (Kauno technologijos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija, T 001, informatikos inžinerija, T 007).

Disertaciją redagavo: anglų kalbos redaktorius dr. Armandas Rumšas (leidykla „Technologija“), lietuvių kalbos redaktorė Aurelija Gražina Rukšaitė (leidykla „Technologija“).

Informatikos inžinerijos mokslo krypties disertacijos gynimo taryba:
prof. dr. Rytis MASKELIŪNAS (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija, T 007) – **pirmininkas**;
prof. dr. Arnas KAČENIAUSKAS (Vilniaus Gedimino technikos universitetas (Vilnius Tech), technologijos mokslai, informatikos inžinerija, T 007);
prof. dr. Olga KURASOVA (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija, T 007);
doc. dr. Rui OLIVEIRA (Nova universitetas Lisabonoje, Portugalija, technologijos mokslai, chemijos inžinerija, T 005);
prof. dr. Vidas RAUDONIS (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija, T 007).

Disertacija bus ginama viešame informatikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje 2024 m. gegužės 7 d. 9 val. Kauno technologijos universiteto Rektorato salėje.

Adresas: K. Donelaičio g. 73-402, LT-44249 Kaunas, Lietuva.
Tel: (+370) 608 28 527; el. paštas doktorantura@ktu.lt

Disertacija išsiųsta 2024 m. balandžio 5 d.

Su disertacija galima susipažinti internetinėje svetainėje http://ktu.edu, Kauno technologijos universiteto bibliotekoje (Gedimino g. 50, LT-44239 Kaunas, Lietuva) ir Vilniaus Gedimino technikos universiteto (Vilnius Tech) bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius).

TABLE OF Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Relevance of the Research

Informatics engineering promotes biotechnology by creating improved computer systems tools and algorithms which provide for the effective research and interpretation of complicated biological data and phenomena, resulting in faster drug development and manufacture. This synergy between informatics and biotechnology encourages innovation, efficiency, and quality, alongside the improvement in healthcare and research [1, 2]. The importance of informatics in the biotechnology industry, especially in the field of pharmaceuticals, has been gradually growing in recent decades, and, with the start of the COVID-19 pandemic in 2020–2021, the importance of informatics engineering has grown rapidly to speed up the production of drugs and vaccines as much as possible [3, 4].

In the pharmaceutical sector, as well as biologics, large, complex compounds are derived from live organisms. Microorganisms produce substances, such as proteins and antibodies, that are used to treat diseases like cancer, autoimmune conditions, infections and more [5].

Cell culture, formulation, planning, and product synthesis are a few of the processes in the biologics production process [6]. These cell cultures are cultivated in bioreactors, where the necessary ingredients are also synthesized. The goal of a bioreactor is to create the ideal environment for cell growth and product synthesis. Temperature, pH, dissolved oxygen, mixing, nutrition, and sterility are just a few of the main variables that must be carefully considered to create the optimal environment for cells in a bioreactor. High levels of cell growth and productivity can be attained by researchers by optimizing these variables, which will result in the creation of new treatments and goods [7, 8].

Many different cell types in a bioreactor depend on glucose as a crucial nutrition source. It serves as the primary energy source for cellular metabolism and is utilized to make a variety of cellular products. The growth and productivity of cell cultures are significantly influenced by the glucose feed [9, 10]. Low glucose levels in the bioreactor result in slower growth rates. On the other hand, excessive glucose levels cause harmful buildup of byproducts and alterations in cellular metabolism that may affect the overall cell survival and synthesis [11, 12]. Researchers must carefully balance the supply of glucose with the metabolic needs of the cells to optimize glucose feed in a cell culture. This process requires monitoring the typical bioprocess parameters, such as the glucose level in the bioreactor, the growth rate of specific cells, cell concentration, and the product synthesis rate. Adapting the feed rates regarding those variables is mandatory in case we seek to maintain the optimal level of the cultivation process efficiency [13, 14].

Yet, without pricey hardware sensors, measuring those crucial bioprocess parameters online is essentially impossible. **The lack of information about the process is the major problem in biotechnology, and the feedback control is unattainable without acceptable estimation.** In that circumstance, researchers have begun the development of ‘virtual sensors’, or ‘soft sensors’, which are software-based instruments that estimate the process variables online based on the previously

accrued process data [15]. **This research relies on soft sensors that evaluate the cultivation process from the exhaust gas from the bioreactor and give a feedback signal to the control loop that manages the efficiency of the process.**

1.2. Research Object

The research focuses on three key variables in the cultivation process that cannot be measured online – these are the specific cell growth rate, the viable biomass concentration, and the target product concentration.

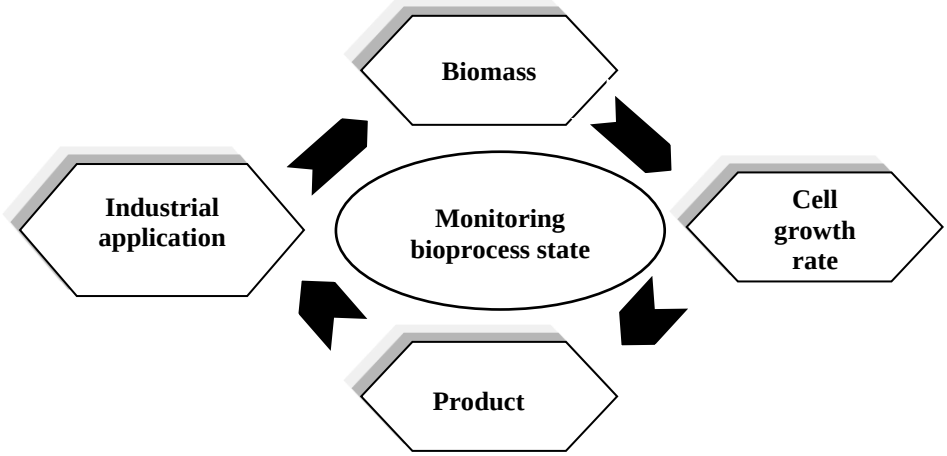


Figure 1.1. Dissertation's main goal: full cultivation process monitoring. It includes biomass estimation, target product estimation, and a specific cell growth rate algorithm

1.3. Aim of the Research

The aim of the research is to increase the productivity of the cultivation process by applying precise, reliable estimation algorithms (soft sensors) based on the bioreactor exhaust gas to monitor the crucial growth process parameters which cannot be assessed directly. To achieve the goals of this research, the following objectives have been set out:

1. Development and investigation of an estimation algorithm for biomass concentration evaluation to monitor the state of a bioprocess.
2. Development and analysis of target product modelling to offer product yield relationships.
3. Development and investigation of an estimation algorithm for specific cell growth rate evaluation to provide feedback information to the control system.
4. Development and investigation of an estimation algorithm for viable cells evaluation of a mammalian cell culture to monitor the state of a bioprocess.

1.4. Dissertation Key Statements

1. In the first half of the cultivation process (the lag and logarithmic phases), the maintenance term of the cells is negligible and has a minimal effect on the precision of estimation procedures.
2. The efficiency of the synthesis of the target products directly depends on the growth rate capacity of specific cells at the moment of induction.
3. The assessment of the bioprocess state based on the oxygen consumption rate conforms to the requirements, and its repeatability is unaffected by *E. coli* bacteria metabolism.
4. The age of the cell culture is a significant variable which improves the precision of viable cell concentration estimations.

1.5. Research Novelty

Each publication of this dissertation provided a scientific innovation of estimation/prediction for significant variables based on fundamental knowledge and mathematical methods. The robustness, simplicity of implementation, highest accuracy as of the date of publishing, and a novel concept which affects the algorithm are the ways in which the papers show their innovativeness.

1. Article 1 (A1)

R. Urniezius, A. Survyla, D. Paulauskas, V. A. Bumelis, V. Galvanauskas
Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate, *Microb Cell Fact* (2019)

The innovation of **A1** paper is the improvement of the Luedeking-Piret model (1959) [16] by expanding the value of the stoichiometric parameter which is responsible for the biomass maintenance term. According to the proposed theory, when the biomass concentration is not high, the maintenance term early in the process is noticeably low and cannot be assessed. Additionally, when induction is being carried out and product synthesis substantially increases, or when the amount of biomass has reached a value leading to appreciable oxygen maintenance, the maintenance term must be evaluated in accordance with the suggested form.

2. Article 2 (A2)

Renaldas Urniežius, Arnas Survyla
Identification of functional bioprocess model for recombinant E. coli cultivation process, *Entropy* (2019).

A2 publication is an expansion of Study **A1**, which has been further developed to incorporate oxygen (resource) consumption for product synthesis throughout the maintenance term. The article also covers the effects of a specific growth rate during induction on the product synthesis and offers relationship to increase the efficiency of cultivation processes.

3. Article 3 (A3)

Arnas Survyla, Donatas Levisauskas, Renaldas Urniezius, Rimvydas Simutis

An oxygen uptake-rate-based estimator of the specific growth rate in Escherichia coli BL21 strains cultivation processes, *Computational and Structural Biotechnology Journal* (2021)

The algorithm's uniqueness lies in its robustness and simplicity because it only has one tuning parameter and is appropriate for the whole cultivation process with various conditions and metabolic pathways. To eliminate uncertainties and recalculation errors, the specific growth rate is calculated directly from the gas data (the oxygen uptake rate) rather than from the evaluated biomass.

4. Article 4 (A4)

A. Survyla, R. Urniezius, R. Simutis

Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional, *Talanta* **254**, 124121 (2023)

The study's primary innovation is the addition of a cell-aging term for use to estimate viable cells. The age of the culture has an impact on the productivity of the cultivation process as well as the dynamic and yields of the synthesis.

1.6. Research Methodology

The primary goal of this study is to assess the crucial cultivation process variables that are impossible to measure online, as shown in Figure 1.1. The Luedeking-Piret model [16] is a fundamental knowledge-based model that was chosen as the origin to ensure the robustness and simplicity of the research algorithms. Since the biomass concentration defines the overall condition of the cultivation process, the first phase in obtaining the necessary parameters is to investigate and create the biomass estimation method (**A1**). The second publication (**A2)** is a sequel of article (**A1)**, which proposed an extended biomass concentration algorithm supplemented by product synthesis. Furthermore, the study also covers the effects of a specific growth rate during induction and offers relationship to increase the efficiency of product synthesis. Since the cells specific growth rate variable is important to the protein synthesis model, the third published article features an estimation algorithm of the specific growth rate (SGR) (**A3**). The final paper (**A4**) describes an improved biomass concentration algorithm capable of estimating a viable cell to cover a wider range of cell cultures. Study (**A4**) inherited the cell culture age variable from an article written by colleagues [17].

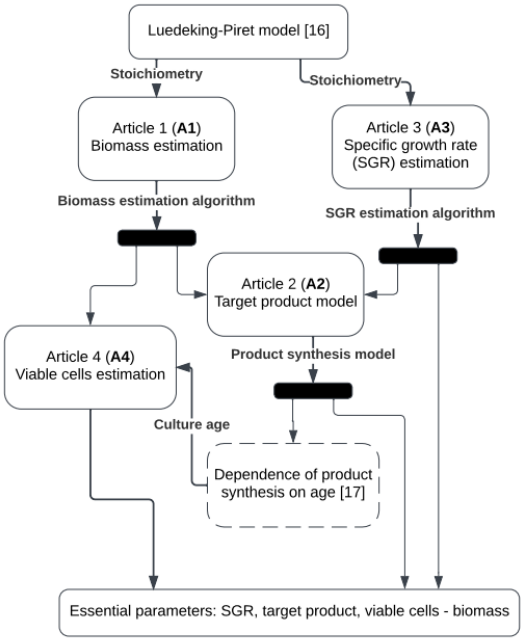


Figure 1.2. Methodology of this research and relationship between articles

1.7. Practical Significance

1. The estimation algorithms of biomass concentrations, specific cell growth rate, target protein can be implemented and operated in the following biotechnology companies in Lithuania:
    a) *Northway Biotech*, UAB,
    b) *Celltechna*, UAB,
    c) *Thermo Fisher Scientific Baltics*, UAB,
    d) *Roquette Amilina*, AB.
2. The methods and algorithms presented in this research have been established in support of 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.

1.8. Approval of the Results

1. The proposed algorithms and models have been published in 4 different international scientific journals referred to in the *Web of Science* database. Three journals are in Q1 quartile, and 1 journal is in Q2 quartile.
2. The essential outcomes have been presented in 3 international conferences.

3. All the estimation algorithms presented in this research have been implemented and used in the laboratory of bioprocessing modelling and management at Kaunas University of Technology.
4. The estimation algorithm of biomass concentrations (**A1**) has been installed and employed in the R&D laboratory of the Centre for Innovative Medicine (IMC).

1.9. Author's Contribution

A1: RU (the corresponding author) created the concept, edited the manuscript, and developed the mathematical model. AS created the concept, edited the manuscript, developed the mathematical model, analysed the data, and validated the algorithm. DP organized the data and validated the model. VB was responsible for funding, and project administration. VG reviewed and edited the manuscript and was responsible for sourcing the funding.

A2: RU (the corresponding author) devised the concept, prepared the manuscript, developed the mathematical model, and submitted the final manuscript. AS devised the concept, prepared the manuscript, developed the mathematical model, analysed the data, and validated the algorithm.

A3: AS: developed and improved the algorithm, was responsible for validation, data curation, and prepared the original draft. DL: developed the algorithm and prepared the original draft. RU (the corresponding author): edited the manuscript, was responsible for funding, and submitted the final version of the manuscript. RS: was the data organizer and supervisor.

A3: AS worked out the concept, developed the soft sensor, analysed and validated the data, and prepared the final draft. RU (the corresponding author) contributed to the model development, edited the manuscript, submitted the final manuscript, and was responsible for sourcing the funding. RS: was the data organizer and supervisor, and also reviewed the manuscript.

1.10. Literature Review

The most important issues in biotechnology are the optimal control of the fed-batch cultivation process and the monitoring of the cell's growth states [18]. As a result, fed-batch cultivation process optimization and the development of a precise monitoring system have been topics of ongoing research for many years, and are still highly relevant today [19]. This dissertation's investigation primarily focuses on the development of fundamental models of the growth process in order to establish an estimation algorithm which would offer crucial information to the control algorithm as a feedback signal and data to the monitoring system [20]. The creation of sustainable engineering solutions should consider using soft/non-invasive sensors [21] to enhance the product quality, to obtain the coefficient values more effectively, to increase safety, and to provide the feedback signal. A feedback signal from soft sensors or estimation algorithms which offer parameters that cannot be directly measured online [22] is necessary for the implementation of a feedback control system. The key parameters of the bioprocess – the biomass concentration and the specific cells growth rate – are taken into account by the control algorithm and the feedback signal [23–25].

Scientific techniques are necessary to gather the required information during cell culture procedures, as critical parameters cannot be examined directly. Consequently, scientists have been working hard on numerous innovations and estimating methods of biomass concentration and specific growth values [26]. As of now, there are numerous methods for achieving the desired outcomes. One of the ways by which on-line measurements are obtained is by using advanced instrumentation which has to be implemented in the bioreactor: dielectric spectroscopy, Raman spectroscopy, and in situ mid-infrared spectrometry. Dielectric spectroscopy is used for monitoring the biomass concentration [27]. The main drawback of this measurement device is its need to be calibrated for each cell strain, and it is not suitable for low concentration measurements. The media composition, density and homogeneity also have a significant impact on the accuracy. The Raman spectroscopy instrument [28] and in situ mid-infrared spectrometry [29] are used for the measurement of glucose concentration in the fed-batch cultivation process. A nonlinear model for biomass concentration estimation from the glucose measurements has the primary drawback of the necessity of interaction with the device containing the medium so that to detect the glucose concentration online. Furthermore, the function of glucose consumption encounters accuracy issues during the culture death phase and metabolic pathways of the cells [30]. As a result, the expensive instruments that must be integrated into the bioreactor must deal with the challenging bioprocess conditions and the rapidly altering composite medium state that affect the results [31].

Another path taken by the researchers is the development of estimation models using bioreactor exhaust gas analysis data. The main data points that are directly related to cell growth and synthesis processes are the *Oxygen Uptake Rate* (OUR) and the *Carbon dioxide Production Rate* (CPR). As estimation algorithms require data that is closely related to the biomass growth rate and the biomass concentration, this path is the most popular among scientists [32, 33]. By using data from exhaust gas as inputs to the artificial neural network (ANN), several researchers have proposed reliable results of the predicted parameters [33, 34]. When the entire model is a black box model, the data requirements for network training are substantial and complex when utilizing ANN alone with OUR and CPR data. Further research demonstrates that the amount of data required for training and the complexity of the training were lowered when ANN and the mass balance equation were combined [35, 36]. The model's accuracy rises rapidly. However, with hybrid models, obtaining sufficient data for model calibration remains challenging to attain adequate precision, and methods are also denoted by high operational demands, and design space maintenance [37]. The developed model also only functions with one particular cell culture.

The primary areas of research that this dissertation focuses on are fundamental knowledge-based models relying on exhaust gas analyses. Robert Luedeking and Edgar L. Piret [16] were the first scientists to propose a stoichiometry equation linking the oxygen uptake to cell development. The two key elements in the suggested model are the oxygen consumption term for biomass maintenance and the cell reproduction. However, since Piret proposed the model in 1959, the cultivation process and the cell cultures have changed since then, and the model no longer satisfies the criteria for accuracy. As a result, Piret's theory has been refined further, and several additional

concepts and models have since been put forth [38, 39]. Fundamental knowledge-based estimation algorithms provide the following benefits: fundamental knowledge-based estimating algorithms use underlying theories, principles, and validated models to produce estimates that are accurate and resilient. When compared to solely data-driven methods, fundamental knowledge-based estimation algorithms are frequently more resistant to noisy or insufficient data. Robustness: mathematical equations that can be easily implemented on any platform serve as the foundation for many key knowledge-based estimation techniques [40, 41].

2. OVERVIEW OF THE PUBLISHED ARTICLES

By addressing the primary issue in biotechnology – a lack of knowledge about critical parameters – this dissertation provides four soft-sensor publications offering the accurate feedback signal that is required:

1. Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate (**A1**) [42] (Section 2.2)
2. Identification of functional bioprocess model for recombinant E. Coli cultivation process (**A2**) [43] (Section 2.3)
3. Oxygen uptake-rate-based estimator of the specific growth rate in Escherichia coli BL21 strains cultivation processes (**A3**) [44] (Section 2.4)
4. Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional (**A4**) [45] (Section 2.5)

2.1. General Overview of the Articles

This dissertation consists of four articles, and each of them offers a soft-sensor technique. The primary objective of this project is to develop an estimating method that is robust, fundamental knowledge-based, and operates on bioreactor exhaust gas. By using stoichiometric knowledge, the presented algorithms transform information from exhaust gases to a general picture of the plant. The stoichiometry form, which connects gas analytics to the state information of the process, was proposed long ago (1959) [16]. Nevertheless, the proposed equation has become outdated because it does not account for contemporary inventions that have emerged on the edge of the 20 and 21 centuries and during the most recent period [46].

Hence, one of the key benefits of the papers is that the proposed technology includes cultivation processes that use induction (e.g., with isopropyl-D-1-thiogalactopyranoside/IPTG) which activates product synthesis [47]. Following induction, cultivation process phenomena quickly alter, which makes process estimation more difficult.

Assessment of fed-batch processes with limited feed rates to the models is another important advantage of the articles. Low feed rates to the bioreactor result in a lack of glucose in the media. With this type of cell culture cultivation, the growth rate of the cells changes dynamically and is based on the feed rate, so the value of the growth rate can range from zero to the maximum passable value immediately [48].

In general, the major objective of the algorithm creation was to examine the impact of induction on the process as well as the variations in the feeding methods and their effects on the efficiency of the cultivation process. Considering these two key factors, this work provides a reliable model to produce the necessary estimate.

2.2. Generic Estimator of Biomass Concentration for Escherichia Coli and Saccharomyces Cerevisiae Fed-Batch Cultures Based on Cumulative Oxygen Consumption Rate

The first published research article is *Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate* written by Renaldas Urniezius, **Arnas Survyla**, Dziugas Paulauskas, Vladas Algirdas Bumelis, and Vytautas Galvanauskas. This work focuses on online biomass concentration estimation in fed-batch cultures. It offers a methodical solution for bioengineering that is investigated for *Saccharomyces cerevisiae* and *Escherichia coli* cells. The experimental analysis of both cultures provides results of experimental validation from the beginning of the bioprocess and assesses the induction effect. The main goal of this study is to provide the possibility to ensure continuous monitoring of the state of the bioprocess and abandon the sampling procedure.

2.2.1. Material and Methods

In this paper, four different strain cultures were examined to confirm the biomass concentration estimation. *S. cerevisiae* (no DY7221) was utilized as a benchmark for yeast cells. Other cell cultures in the research were *Escherichia coli* recombinant strains: *E. coli* BL21(DE3) pET9a-IdeS, *E. coli* BL21 (DE3) pET21-IFN-alfa-5, and *E. coli* BL21(DE3) pLysS.

Yeast cells were cultivated in the Laboratory of Bioprocess Design and Modelling at Kaunas University of Technology. The cultivation processes were of the fed-batch type with a limited feeding rate. The cells were grown in a standard nutrient medium (YPD) with 0 g/kg initial glucose concentration. The feeding solution included 600 g/kg of glucose and was used immediately at the beginning of the process. Online measurements of the bioreactor exhaust gas were monitored by using a *BlueSens* gas analyser (*BCpreFerm*, *BlueSens*, Herten, Germany).

The cell strains *E. coli* BL21 (DE3) pET21-IFN-alfa-5 and *E. coli* BL21 (DE3) pET9a-IdeS were cultivated in industrial R&D laboratories. The growth techniques were focused on the fed-batch type with limited feeding solution. Induction (ITPG) was used at both cell cultures to activate product synthesis. The bioreactor's exhaust gases were analysed online by *BlueSens*.

E. coli BL21(DE3) pLysS growing information was collected from the authors to incorporate different strains and culture techniques from various countries into this research [48]. Cells were cultivated in a minimal mineral medium.

2.2.2. Novelty and comparison of biomass estimation

The main mathematical framework in this paper, which was used to estimate the biomass concentration, is the Luedeking-Piret model developed via the stoichiometric equations for oxygen consumption [16]. As a result, the model for estimating biomass depends on the cells' oxygen uptake rate data. The proposed estimator is compared to the most recently published biomass prediction models that also use oxygen consumption to demonstrate the uniqueness, robustness, and ease of implementation of the algorithms.

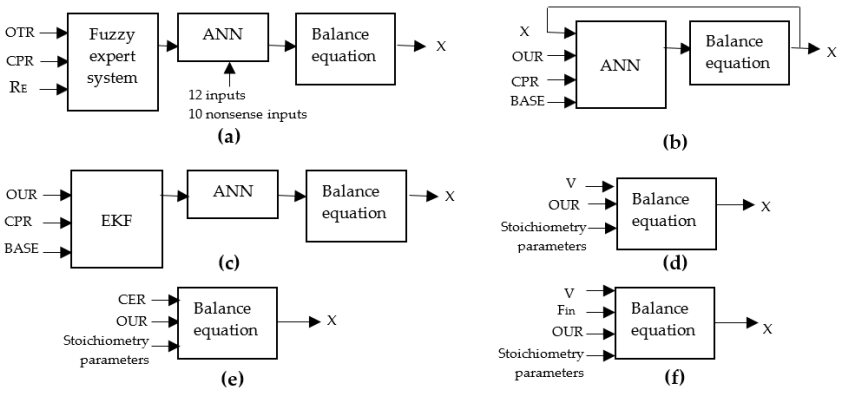


Figure 2.1. Related biomass estimators: (a) A. Lübbert [49], (b) M. Achle [35], (c) R. Simutis [36], (d) biomass estimation of this text, (e) R.A. Davis [50], (f) J.M. Barrigon [51]

The complexity of the chosen algorithms, the number of input signals, and the precondition parameters or the initial conditions needed are depicted in Figure 2.1, along with their main distinctions. The main goal of this paper is to demonstrate that, based on stoichiometry, biomass estimation may be approached from a fundamental point of view, and that the method's format does not have to be complicated.

2.2.3. General mathematical model of stoichiometric parameters estimation

The Luedeking-Piret theory serves as the model's foundation, which links the stoichiometry of cells with oxygen consumption. It illustrates the relation between the oxygen uptake rate in a bioreactor and the biomass concentration growth/maintenance [16, 52]:

$$OUR(t) = \alpha \cdot \mathrm{X}'(t) + \beta \cdot \mathrm{X}(t). \tag{2.1}$$

The stoichiometric coefficients α and β are parameters of the biochemical reactions during the cultivation process, X is the biomass concentration, g/L. In Eq. (2.1), α coefficient denotes a cell's oxygen consumption yield for growing ($\alpha \equiv Y_{O/X}$), and β coefficient denotes oxygen consumption for maintenance ($\beta \equiv m$) [53, 54]. Real-time data obtained from the devices contains interference and disruptions throughout the cultivation process, which might lead to the parameter and estimated value distortion [55]. To reduce uncertainty, cumulative information is used. Furthermore, when biomass and its metabolic byproducts increase during culture research, these masses are more closely connected to the cumulative signals of OUR and CPR [56, 57]. As a result, by integrating the model in Eq. (2.1), the new provided form is protected from disturbances:

$$\int_{t_0}^{t} OUR(t^*)\,\mathrm{d}t^* = \alpha \cdot \int_{t_0}^{t} X'(t^*)\mathrm{d}t^* + \beta \cdot \int_{t_0}^{t} X(t^*)\,\mathrm{d}t^*. \tag{2.2}$$

The proposed maintenance term form, which is a growing variable rather than a constant, especially after induction, is one of the primary novelties of the research. The fact that the formation of the product and other elements are included in the oxygen consumption for biomass maintenance may be used to explain the phenomena

of the parameter's value growing. Such a circumstance most frequently arises when induction is carried out, and the product's synthesis substantially increases near the end of an exponential phase of microbial cultivation (for recombinant protein synthesis) [57, 58]. This paper suggests the following β parameter form of two additive terms:

$$\beta = \frac{1}{Y_{X/O}} + \frac{1}{Y_{P/O}}. \tag{2.3}$$

In Equation (2.3), the oxygen demand for product synthesis is denoted by $Y_{P/O}$, whereas $Y_{X/O}$ represents the oxygen consumption for cell respiration. Consequently, biomass concentration has a linear or polynomial affiliation, which depends on the strain and cultivation process model, connection to the β parameter, as shown in Figure (2.2). *β(tm)* is the maintenance value from the cultivation experimental data at the moments after induction.

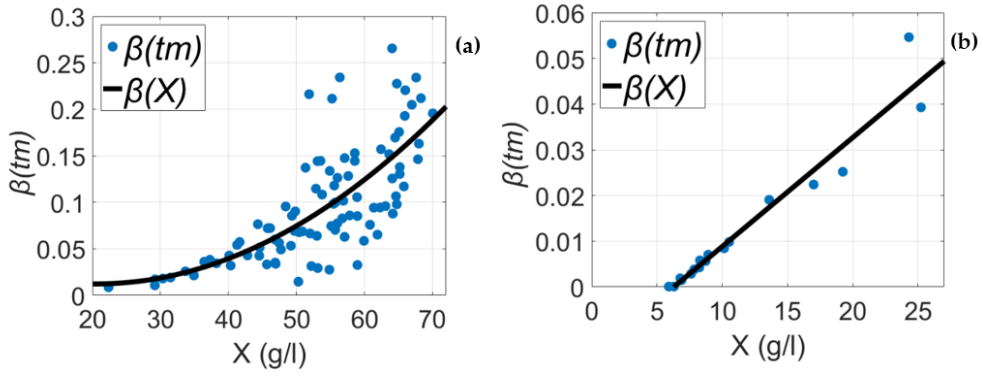


Figure 2.2. Biomass concentration influence on oxygen consumption for maintenance, (a) *E. coli*, (b) *Saccharomyces cerevisiae*

Figure (2.2) illustrates how the biomass concentration affects the maintenance term value. A parabola regression of the biomass concentration is represented by the expression of parameter *β(X)* to the *E. coli* strain. Induction in the growth processes carried out at various periods led to the distribution of the maintenance period shown in the graph.

$$\beta(X) \equiv \beta\big(X(t)\big) = k_{\beta 2} \cdot X^2(t) + k_{\beta 1} \cdot X(t) + k_{\beta 0}. \tag{2.4}$$

The maintenance oxygen requirement of *S. cerevisiae* is linearly related to biomass concentration, $k_{\beta s2} = 0$.

$$\beta_{Saccharomyces}(X) \equiv \beta_{Saccharomyces}\big(X(t)\big) = k_{\beta s1} \cdot X(t) + k_{\beta s0}. \tag{2.5}$$

Accordingly, in view of Figure (2.2) and Equations (2.4–2.5), the cell strain has a particular biomass concentration value $X_{specific}$, at which point the maintenance term became apparent. Equations (2.4) and (2.5) must be set to zero and solved to determine the culture's unique biomass concentration value, $X_{specific}$:

$$\beta(X_i) \equiv \beta\big(X(t)\big) = 0. \tag{2.6}$$

Following the discovery of the stoichiometric cell culture parameters and the introduction of a new form for the maintenance term, the Luedeking-Piret model with the dynamic maintenance term takes the following final revised form:

$$\begin{cases} OUR(t) = \alpha \cdot X'(t) + k_{\beta 2} \cdot X^3(t) + k_{\beta 1} \cdot X^2(t) + k_{\beta 0} \cdot X(t), X(t) > X_{specific}; \\ OUR(t) = \alpha \cdot X'(t), X(t) \le X_{specific}. \end{cases} \tag{2.7}$$

Therefore, stoichiometric parameters and the cumulative oxygen uptake rate (*cOUR*) are used in this work to estimate the biomass concentration. When the oxygen demand for maintenance is very low or non-existent before the specified biomass concentration $X_{specific}$ level has been attained, the biomass state estimation equation is:

$$X_m = \frac{cOUR_m}{\alpha} + X_0. \tag{2.8}$$

The second scenario's stoichiometric parameter β takes effect as a function of the biomass concentration after the concentration of biomass starts exceeding $X_{specific}$, at which point oxygen consumption becomes apparent. The proposed integral form of the biomass concentration estimation is as follows:

$$x_m \cong \frac{cOUR_m - \sum_{l=i}^{m-1} \beta(x_l) \cdot x_l \cdot \Delta t_{l,l-1}}{\alpha} + x_0. \tag{2.9}$$

2.2.4. Result

Three different sorts of indicators were used to assess the results of the biomass estimation and evaluate the statistical method's forecast accuracy [59, 60]:

$$\text{Mean absolute error (MAE): } MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}; \tag{2.10}$$

$$\text{Mean absolute error (MAE): } MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}; \tag{2.11}$$

$$\text{Root mean square error (RMSE): } RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}. \tag{2.12}$$

In Equations (2.10–2.12), n is the number of data counts, $\hat{y}_i$ is the methods' result, and y_i is the measured value from the cultivation process.
Several sources were used to gather the experimental biomass measurements and data on the cumulative oxygen uptake rate (*cOUR*) from fed-batch experiments with *E. coli* and *S. cerevisiae*, including datasets from the study [48], industrial R&D and university laboratories. The diversity of cultivation procedures, shown in Table 2.1 with varying reception and methods, demonstrates the universality of the algorithm.

Table 2.1. Variety of cultivation processes for estimator validation

Cell culture	No. of experiments	Feeding type	Bioreactor size L
E. coli	3	Growth limiting	15
Yeast	2	Growth limiting	5
E. coli	1	Growth limiting	12
E. coli	8	Growth limiting	7
E. coli	7	Growth nonlimiting	7

All the cultivation data was examined before the algorithm validation process began to determine the stoichiometric characteristics of cell cultures. The estimating process neglected both metabolic routes: product synthesis in response to induction (e.g., with isopropyl-D-1-thiogalactopyranoside/IPTG) and acetate metabolic during unlimited dosed substrate feed cultivations [61]. Table 2.2 contains the findings of an offline examination of stoichiometric parameters.

Table 2.2. Stoichiometric parameters of cell strains

Escherichia coli	*Saccharomyces cerevisiae*
$\alpha = 1.01$	$\alpha = 1.35$
Confidence Interval ∓ 0.0186	Confidence Interval ∓ 0.149
$k_{\beta e2} = 7.2 \cdot 10^{-5}$	$k_{\beta s2} = 0$
$k_{\beta e1} = -2.9625 \cdot 10^{-3}$	$k_{\beta s1} = 2.3851 \cdot 10^{-3}$
$k_{\beta e0} = 4.27047\text{d} \cdot 10^{-2}$	$k_{\beta s0} = -1.5014 \cdot 10^{-2}$
$X_{specific} = 20.6$ g/l	$X_{specific} = 6.29$ g/l

The second step is to determine the biomass concentration by using both Equations (2.8) and (2.9) from the *cOUR* signal after the cell culture's stoichiometry parameters have been determined. The biomass estimation method used in this paper included a variety of cultivation experiments with various cell strains, bioreactor volumes, feeding solution types, IPTG induction times, OD levels at IPTG injections, substrate feeding restrictions, and substrate feed start times.

Since the inoculation, the average MAE for biomass concentration estimate was 1.1 g/l, and the average MAPE of biomass estimation was 7.28%. *S. cerevisiae* cultivations had an average RMSE value of 0.5 g/l. *E. coli* cultivations with restricted substrate feeding had an RMSE value of 1.26 g/l, while cultivations with dosed substrate feeding had an RMSE value of 2.44 g/l.

In Figures (2.3), (2.4), (2.5), the graphical representation of the biomass estimate results of the most important cell cultures are contrasted with the measured values offline.

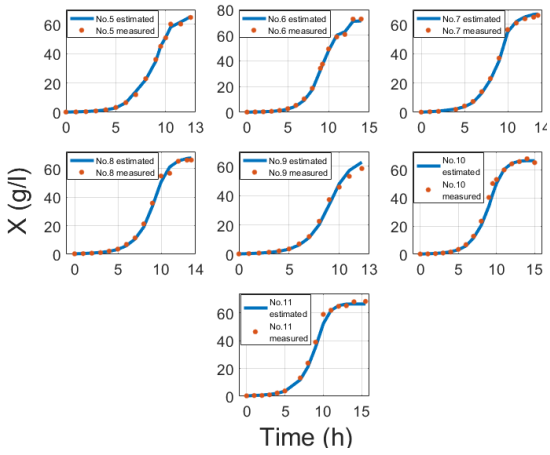


Figure 2.3. Results of this paper's algorithm with recombinant *E. coli* BL21 (DE3) pET21-IFN-alfa-5 strain at 7 L bioreactor with unlimited dosed substrate feeding

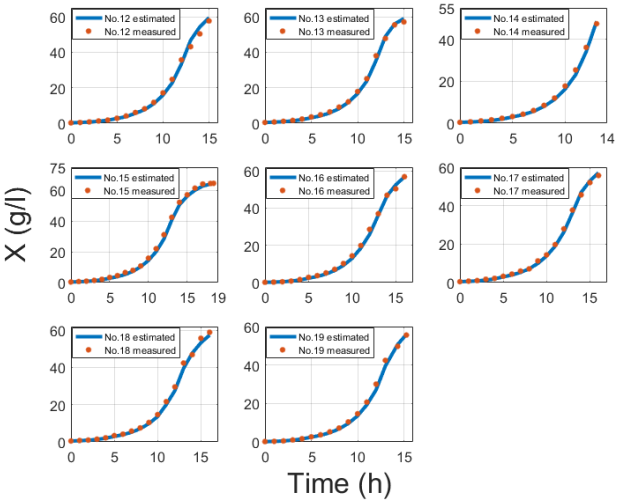


Figure 2.4. Results of this paper's algorithm with recombinant *E. coli* BL21 (DE3) pET21-IFN-alfa-5 strain at 7 L bioreactor with limited substrate feeding

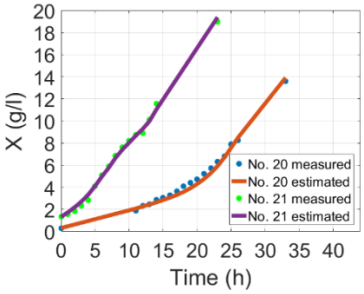


Figure 2.5. Results of this paper's algorithm with *Saccharomyces cerevisiae* (n° DY7221) strain at 5 L bioreactor

2.3. Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process

The second published research article is *Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process* written by Renaldas Urniezius, **Arnas Survyla**. This work focuses on a system identification model which provides accurate biomass concentration and target protein concentration estimates in fed-batch cultivations.

2.3.1. Material and methods

To validate the biomass and protein model fitting in this study, *E. coli* BL21 (DE3) pET28a (*Novagen*) was used as the test object in all experiments [57]. Both active soluble and insoluble forms of *E. coli* BL21 (DE3)'s product were generated as inclusion bodies. In this investigation, inclusion bodies, an insoluble protein, were the intended result. After being induced with 1 mM isopropyl-D-1-thiogalactopyranoside (IPTG), the T7 promoter was in charge of the protein's production.

For the analysis in this work, experimental data [48, 62, 63] was used as the basis of analysis. The BL21 (DE3) pET28a *E. coli* strain was grown in a *B. Braun* 10 L bioreactor. At inoculation, the initial medium volume was 5 L. Mineral salt media was used as the culture medium [62]. Processes for growing plants were implemented by using a fed-batch mode when there was no glucose present in the bioreactor at the moment of inoculation. Therefore, the feed was pumped in from the start. The time picked for induction was 10 hours after the beginning of the process. Online exhaust gassing tracking was done, and a paramagnetic oxygen sensor (*Maihak Oxor 610*) was used to monitor O_2 concentration. SDS-PAGE electrophoresis assisted in determining the concentration of the target protein following the actions of cell disruption, separation of the soluble fraction, and solubilization of inclusion bodies.

2.3.2. System identification and parameter estimation

A previous study [42] shows that a biomass concentration estimator based on oxygen consumption may provide sufficiently accurate results. The novelty of the article [42] was the maintenance terms of the biomass mathematical form, which is dynamic, rather than constant, by assuming that the maintenance term takes into account product synthesis. This study proposes a biomass concentration and protein model fitting based on a mass balance equation when the maintenance term is divided into product synthesis and maintenance. This model was developed on the basis of the Luedeking-Piret model [16]:

$$OUR(t) = \alpha \cdot X'(t) + \beta \cdot X(t) + \gamma \cdot P'(t). \tag{2.13}$$

The target protein for this investigation is the inclusion body, and the protein synthesis yield γ in this study is considered to be a function of the biomass concentration X in a grey box model [64]. Protein productivity is influenced by IPTG and biomass concentrations during the induction period, as demonstrated by Babaeipour *et al*. [65]. The same 1 mM of isopropyl-D-1-thiogalactopyranoside (IPTG) was used in all tests. Furthermore, in each culture operation, the biomass concentration at the induction time varied, and we discovered that it significantly

affected product synthesis. According to our investigation, the biomass concentration at the induction time is a function of the product synthesis parameter $\gamma(X)$ [66]:

$$\gamma(X) = k_\gamma \cdot (X(t) - X_{ind}), \tag{2.14}$$

here, k_γ is the product synthesis yield, which is assumed to be constant, and X_{ind} is the biomass concentration at the moment of induction. In conclusion, the estimator's final model form consists of the product synthesis parameter $\gamma(X)$, the maintenance term which is described in the research [42], and the biomass reproduction term α:

$$\begin{aligned} OUR(t) = \alpha \cdot X'(t) + k_\gamma \cdot (X(t) - X_{ind}) \cdot P'(t) + \\ \left(k_{\beta 2} \cdot X^2(t) + k_{\beta 1} \cdot X(t) + k_{\beta 0}\right) \cdot X(t). \end{aligned} \tag{2.15}$$

The original signal is converted to the cumulative information, which is an effective noise filter, to protect the model from disturbances [36]:

$$\begin{aligned} \int_{t_0}^{t} OUR(t^*)dt^* = \alpha \cdot \int_{t_0}^{t} X'(t^*)dt^* + k_\gamma \cdot \int_{t_0}^{t} (X(t^*) - X_{ind}) \cdot \\ P'(t^*)dt^* + \int_{t_0}^{t} \left(k_{\beta 2} \cdot X^2(t^*) + k_{\beta 1} \cdot X(t^*) + k_{\beta 0}\right) \cdot X(t^*)dt^*. \end{aligned} \tag{2.16}$$

Regarding the following model analysis and calculations, the results indicate that, during the whole culture process, the stoichiometric parameter $\beta(X)$, the oxygen maintenance term for biomass concentration, is significantly lower than the other stoichiometric values. The fact is that the biomass concentration at the induction moment is relatively low (about 30 g/L). The biomass maintenance is unnoticeable until the induction and after IPTG injection, whereas product synthesis covers the maintenance term. This phenomenon explains why the biomass maintenance component is absent in this approach. The model simply takes into account two terms of oxygen consumption:

$$\int_{t_0}^{t} OUR(t^*)dt^* = \alpha \cdot \int_{t_0}^{t} X'(t^*)dt^* + k_\gamma \cdot \int_{t_0}^{t} (X(t^*) - x_{ind}) \cdot P'(t^*)dt^*. \tag{2.17}$$

By simplifying Equation (2.17) and applying the Riemann sum [67], the final form of oxygen consumption is:

$$cOUR_m = \alpha \cdot (X_m - X) + k_\gamma \cdot (P_m \cdot (X_m - X_{ind}) - \sum_{l=1}^{m} (X_l - X_{l-1}) \cdot P_l). \tag{2.18}$$

In Equation (2.18), $\mathrm{cOUR_m} \equiv \int_{t_0}^{t} \mathrm{OUR(t^*)}\, dt^*$, $X_m \equiv X(t)$, $X_0 \equiv X(t_0)$ is the biomass concentration at the inoculation moment, and $\mathrm{m} \in [1, \mathrm{n_m}]$ is the measurement of offline samples. The following algorithm is the final formula for biomass concentration model fitting:

$$X_m = \frac{cOUR_m + \alpha \cdot X_0 + k_\gamma \cdot P_m \cdot X_{ind} + k_\gamma \cdot \sum_{l=1}^{m} (X_l - X_{l-1}) \cdot P_l}{\alpha + P_m \cdot k_\gamma}. \tag{2.19}$$

The second proposal of this study is a product estimation algorithm. According to the work by Levisauskas *et al*. [68], which asserts that protein synthesis is a function of the biomass growth rate:

$$\frac{dP_x}{dX} = q_{px}(\mu, P_x). \tag{2.20}$$

In Equation (2.20), P_x is protein concentration divided by biomass concentration $P_x(t) = P(t)/X(t)$, q_{px} is a particular protein accumulation rate (U/g/h), μ is a specific biomass growth rate (1/h) [68]. Data analysis and research demonstrated that product synthesis is inhibited by product concentration and is linearly dependent on the specific growth rate (SGR) of the biomass [69]:

$$\frac{dP_x}{dX} = q_{px}(\mu, P_x) = P_{max}(\mu, X) - k_t \cdot P_X(t). \tag{2.21}$$

In Equation (2.21), parameter k_t is a time constant value that is expected to be a form of the self-inhibition outcome [70]. P_{max} is a maximal specific product concentration, which depends on μ (SGR). The maximum particular product concentration, expressed as the highest potential protein concentration in the current process stage, is determined by the specific biomass growth rate and biomass concentration:

$$\begin{aligned} P_{max}(\mu, X) &= \mu(t) \cdot \left(k_{m0} + k_{m1} \cdot (X(t) - X_{ind})\right), \ t \geq t_{induction}, \\ P_{max}(\mu, X) &= \mu(t) \cdot k_{m0}, \qquad t < t_{induction}. \end{aligned} \tag{2.22}$$

In Equation (2.22), k_{m0} and k_{m1} are empirical parameters suggested by this study, k_{m0} relates to SGR and protein synthesis, and k_{m1} links the biomass concentration at the induction time and productivity [64], $t_{induction}$ is the induction moment (hours). The product estimation technique also serves as the biomass estimation model in the cumulative form to protect against disturbances, and by applying the left-hand Riemann sum, the protein model's final formula is as follows:

$$P_m = \frac{\left(\sum_{j=i}^{m} P_{max,j} \cdot \Delta t_{j,j-1} - k_t \cdot \sum_{j=1}^{m-1} P_{x,j} \cdot \Delta t_{j,j-1}\right) \cdot X_m}{1 + \Delta t_{m,m-1} \cdot k_t}. \tag{2.23}$$

The paper also presents a parameter identification process using a convex optimization method and proposes new criteria to find parameter values that are appropriate from the beginning of the cultivation process, when the biomass concentration is low, to the end, when the parameter values are much higher:

$$\sigma_{X,m}^2 \sim \frac{X_m^2}{1 - K_{exp} + X_m^2 \cdot K_{exp}}, \tag{2.24}$$

Here, K_{exp} is a tuning coefficient which is required to adjust the uncertainty. A value of '1' for K_{exp} ($0 \leq K_{exp} < 1$) replicates the least squares method, which has a heavier penalty for larger criteria values. In the meanwhile, the squared MAPE criteria are produced when the value is zero.

2.3.3. Results

Three assessment criteria are applied, the same as in Section 2.2.4, to compare our results with those of other researchers: MAE, MAPE, and RMSE, as shown in Equations (2.10–2.12). The experimental data of *Escherichia coli* fed-batch

cultivation processes was obtained from the study [62]. Data from 19 cultivation experiments was investigated in the system identification study to verify and evaluate the suggested model parameters in this paper. Various formulations were used in numerous experiments, including prior hypotheses on polynomial maintenance [42]. The goal was to identify the best formula for describing the cell stoichiometry. The best-achieved coefficient values for the fitted model are shown in Table 2.3.

Table 2.3. Comparison of biomass concentration estimating results. MAPE, mean absolute percentage error; MAE, mean absolute error

Model	α	$k_{\beta 0}$	$k_{\beta 1}$	$k_{\beta 2}$	k_{γ}	MAE	MAPE
Study [3]	0.996	0.07	0.00084	0	—	1.1	7.28%
This study	0.997	0	0	0	2.705	0.68	7.09%

The average value of the investigated experiments was represented by the MAE and MAPE values. The findings indicate that the maintenance factor in the stoichiometry equation is not as successful as product synthesis. Table 2.4 presents the outcomes of the protein model fitting, whereas Table 2.5 represents the model's accuracy of biomass and product estimation.

Table 2.4. Protein model parameters in accordance with Equation (2.23)

***E. coli* BL21 (DE3) pET28a**
$k_{m0} = 0.2346$
$k_{m1} = -0.0172$
$k_t = 0.0687$

Table 2.5. Estimation results of this study

	Dry Biomass Concentration (DCW)			Product concentration		
No.	**MAE (g/L)**	**MAPE (%)**	**RMSE (g)**	**MAE (g/L)**	**MAPE (%)**	**RMSE (g)**
1	0.769	8.594	5.279	0.128	11.947	0.722
2	0.481	7.39	2.916	0.0813	6.565	0.491
3	0.843	8.107	6.354	0.0563	7.86	0.397
4	0.727	5.25	5.975	0.05	4.996	0.323
5	0.596	7.199	4.17	0.134	8.715	0.821
6	0.402	6.033	2.768	0.149	9.26	1.185

The average MAE of biomass concentration from the beginning of inoculation is 0.636 g/L, while that of the product is 0.099 g/L, according to the validation results displayed in Table 2.5. Since the beginning of inoculation, the biomass concentration average MAPE was 7.09%, and the product average MAPE was 8.22%. Since the beginning of inoculation, the average RMSE for the biomass concentration was 4.577%, whereas the average RMSE for the product was 0.656%. These results are also shown graphically in Figure 2.6 presenting the estimation of biomass concentration, and Figure 2.7 featuring the product estimation.

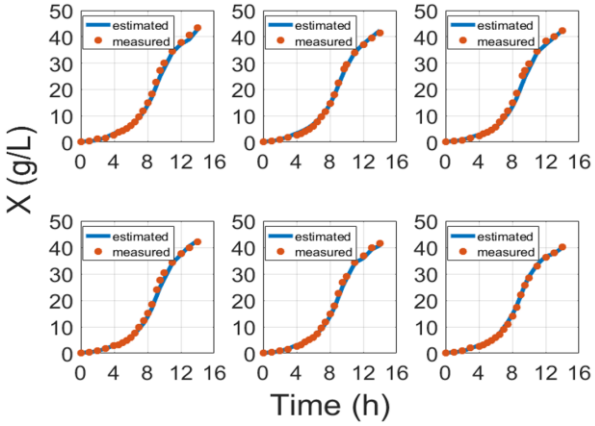


Figure 2.6. Results of the biomass concentration estimation algorithm

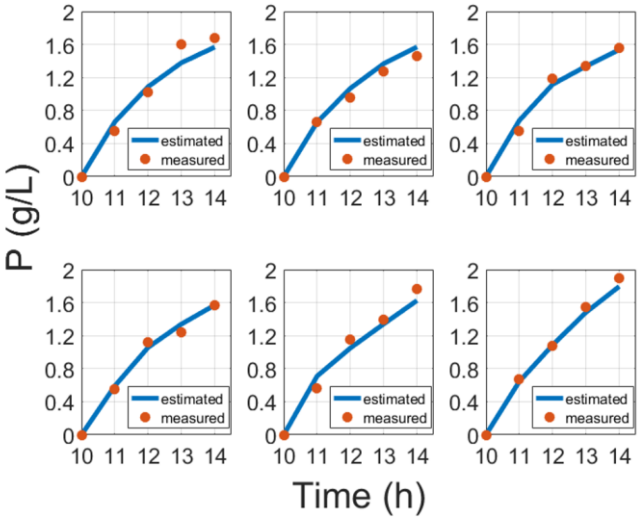


Figure 2.7. Results of target product concentration estimation

Table 2.6. Comparison of estimating model accuracy outcomes amongst research projects focusing on product prediction

Paper works	RMSE (g)		
	Total Biomass	Total Soluble Protein	Total Insoluble Protein
Conventional model from Gnoth *et al.* [62]	10.81	1.78	0.87
Hybrid network from Gnoth *et al.* [62]	4.71	1.28	0.62
Model in this study	4.577	—	0.656

The main objective of this paper is to provide evidence that biomass and protein model fitting can be handled from a fundamental point of view based on Stoichiometry Equation (2.1) and protein synthesis Equation (2.20), which would

enrich hybrid approaches using the *Artificial Neural Network* (ANN) or other hybrid black box systems requiring data training as results are compared in Table 2.6 [62, 70. 71]. In contrast to the usual requirement for enormous quantities of training data, this study suggests a strategy that aids in the identification of the parameters only once per strain.

2.4. An Oxygen-Uptake-Rate-Based Estimator of the Biomass-Specific Growth Rate in Microbial Cultivation Processes

The third published research article is *An oxygen-uptake-rate-based estimator of the biomass-specific growth rate in microbial cultivation processes* written by **Arnas Survyla**, Donatas Levisauskas. Renaldas Urniezius, and Rimvydas Simutis. This work focuses on the development of an SGR specific cells growth rate from the oxygen uptake rate by applying the fundamental knowledge of stoichiometry. It provides the desired value (SGR) directly calculated from the exhaust gases measurements taken online without the need for any further modelling or biomass estimation. This model is the most suitable in control algorithms where the feedback signal is SGR.

2.4.1. Material and methods

Three different forms of *Escherichia coli* cell-strain culture data were examined in this study to validate the SGR estimations and assess their dependability and adaptability owing to the availability of data. The *E. coli* BL21(DE3) pET9a-IdeS, pET21-IFN-alfa-5, and pLysS strains were chosen as the study's test organisms. In various separate R&D facilities, all three cell strains were grown. A 7 L bioreactor was used to grow the *E. coli* BL21 (DE3) pET21-IFN-alfa-5 cell strain. The culture medium included merely trace amounts of minerals. In a 15 L bioreactor, the *E. coli* BL21 (DE3) pET9a-IdeS cell strain was grown. The introduction of the cultivation media followed the minimal specifications for a mineral medium. A *BlueSens BlueInOne Ferm* gas analyser with a measurement range of 0 to 100% was used to monitor the oxygen content in the exhaust gas from the bioreactor while the *E. coli* BL21 (DE3) pET21-IFN-alfa-5 and *E. coli* BL21 (DE3) pET9a-IdeS cell growth procedures were running. The *E. coli* (BL21(DE3) pLysS) cell strain was grown in a minimum mineral medium. All cultivations started out with a mass of 5 kg. The *Biostat C* apparatus, manufactured by *Sartorius Stedim Biotech*, has a working volume of 15 L and a stirrer speed range of 100 to 1400 rpm. A paramagnetic oxygen sensor, installed in the reactor's vent line beneath the exhaust gas cooler by *Sidor*, Sick-Maihak, Hamburg, was used to monitor the Oxygen Uptake Rate (OUR) online.

2.4.2. Development of the estimation method

This investigation, as well as other papers included in the dissertation, is based on oxygen consumption combined with stoichiometry. The fundamental idea is drawn from the traditional Luedeking-Piret approach [16, 72] and the specific growth rate expression, which is taken from the biomass growth dynamics:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu \cdot X(t). \tag{2.25}$$

In Equation (2.25), X is the biomass concentration value in the broth; μ is the specific growth rate; t is time. A combination of Piret model Equation (2.1) and the specific growth rate approach Equation (2.25) gives us the expression of oxygen consumption, which is mostly dependent on SGR [73]:

$$\frac{1}{OUR(t)} \cdot \frac{\mathrm{d}OUR(t)}{\mathrm{d}t} = \frac{1}{\mu + \beta/\alpha} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}t} + \mu. \tag{2.26}$$

In Equation (2.26), the cell metabolism of oxygen consumption is described by the stoichiometric coefficients α and β. The goal of this study is to provide a model that would be robust with the least amount of complexity feasible when using the final equation, which only contains the oxygen uptake rate, stoichiometry, and SGR, without any biomass expression [74]. The model needs to be made simpler and dynamically accurate when it is acquired. From Equation (2.26), the dynamic part is formed:

$$R = \frac{1}{OUR(t)} \cdot \frac{\mathrm{d}OUR(t)}{\mathrm{d}t}; \rightarrow \frac{1}{\mu + \beta/\alpha} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}t} + \mu = R. \tag{2.27}$$

Further on, the relationship between specific growth rate and stoichiometry is expressed from the main equation as follows:

$$T = \frac{1}{\mu + \beta/\alpha}. \tag{2.28}$$

The final step is the z-transform which is used to extract the discrete *OUR* measurement-based SGR estimate technique. Figure 2.8 presents the discrete algorithm of the SGR estimator.

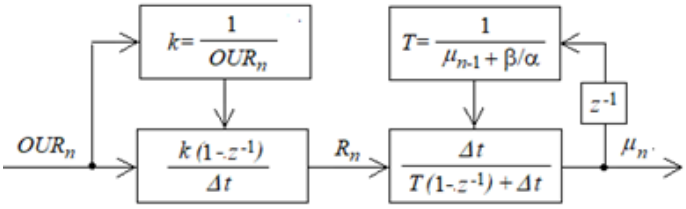


Figure 2.8. Block scheme of the SGR estimation algorithm (z-1 is the backward-shift operator, Δt is the sampling time, and n is the number time discretization point) [30]

The link between the dynamics of oxygen consumption and the time constant, which determines the value of SGR, is depicted in the SGR estimator's structure diagram. The application of the z-transform to the equation yields the final SGR estimator formula:

$$\mu_n = R_n \cdot \frac{\Delta t}{T + \Delta t} + \mu_{n-1} \cdot \frac{T}{T + \Delta t}. \tag{2.29}$$

The SGR estimator that is presently shown is adaptable and may be used to track a variety of cultivation processes. The stoichiometric parameter ratio β/α is the only one tuning parameter of the approach, which is unique to a certain strain of

microorganisms and may be obtained in reference books or determined from preliminary batch culture experiments [30].

2.4.3. SGR tests in computer simulation

By using computer simulation on the *MATLAB/Simulink* platform, the first stage in evaluating the performance of the SGR estimator was chosen. Different fed-batch cultivation procedures with various SGR time profiles were represented in the simulation by altering the feeding solution's speed. This is a description of a mathematical model of the *E. coli* culture technique [75, 76]. The biomass concentration (g/L) formula is as follows:

$$\frac{\mathrm{dX}}{\mathrm{d}t} = \mu(s) \cdot \mathrm{X} - F \cdot \frac{X}{V}, \tag{2.30}$$

here, X is the biomass concentration, g/L; V is the working volume, L; μ is a specific cells growth rate, 1/h; F is the substrate feeding rate, g/h; t is the cultivation process time, h. The glucose concentration model is as follows:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -q_s(s) \cdot \mathrm{X} + F\frac{s_\mathrm{f} - s}{V}, \tag{2.31}$$

here, q_s is the substrate consumption rate, g/(g*h); s_f is the glucose concentration in the feeding solution, g/L.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = F, \tag{2.32}$$

$$OUR = \alpha \cdot \mu(s) \cdot \mathrm{X} \cdot V + \beta \cdot \mathrm{X} \cdot V. \tag{2.33}$$

Equation (2.32) is the volume of media in a bioreactor dependent on the feeding solution. Equation (2.33) is the Luedeking-Piret model. The specific growth rate is expressed by the Monod model, when the growth rate is dependent on the glucose concentration [77, 78].

$$\mu(s) = \mu_{\max} \cdot \frac{s}{k_\mathrm{s} + s} \cdot \frac{k_\mathrm{i}}{k_\mathrm{i} + s}, \tag{2.34}$$

here, μ_{max} is the maximum possible specific cells growth rate of the cell culture, and k_s, k_i are the Monod coefficients. The final simulation equation is the specific glucose consumption rate:

$$q_s(s) = \frac{\mu(s)}{Y_{\mathrm{x/s}}} + m, \tag{2.35}$$

here, $Y_{x/s}$ is the yield coefficient (g/g); and m is the glucose requirement for biomass maintenance g/(g*h).

Table 2.7. Values of model coefficients [44]

Parameter	Value	Dimension	Parameter	Value	Dimension
k_i	85	g/L	α	0.82	g/g
k_s	0.7	g/L	β	0.01	g/(g*h)
m	0.02	g/(g*h)	μ_{max}	1.1	1/h
s_f	150	g/L	X (0)	0.5	g/L
$Y_{x/s}$	0.8	g/g	s (0)	5.0	g/L
V (0)	8.0	L			

With the given Equations (2.30–2.35) and the coefficients from Table 2.7, various combinations of the specific growth rate of the cultivation process were tested in simulation with different feeding profiles. Additionally, white noise was introduced to the oxygen consumption signal to test the model's stability and tolerance to disturbances:

$$OURm_n = OUR_n + \sigma \cdot OUR_n \cdot Rand. \tag{2.36}$$

2.4.4. Results

The recursive estimating algorithm's time discretization step was set in the simulation trials to t = 0.0025 h, and the ratio β/α (the tuning parameter) value was established as 0.01. Figure 2.9 (Experiment I and Experiment II) shows the outcomes of the simulation experiments carried out under different cultivation disturbances.

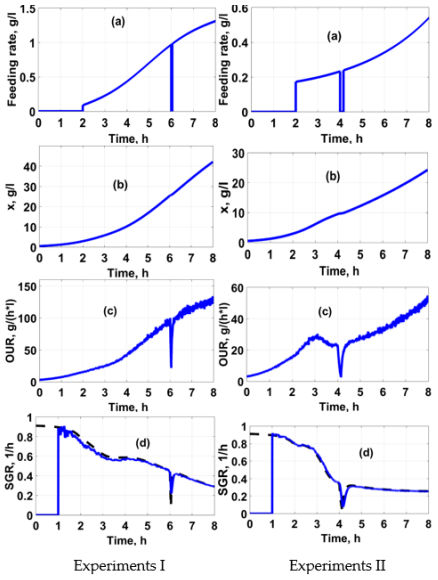


Figure 2.9. Simulation results of SGR estimator performance by tracking various SGR time trajectories (Experiments I, II): (a), (b), (c) feeding rate, biomass growth, and oxygen uptake rate curves, respectively; d) comparison of the simulated SGR versus estimated SGR curves (dotted and solid lines, respectively)

Furthermore, the approach of this study was tested with the real cultivation process data of three types of *E. coli* cell strains which were cultivated in different R&D laboratories and in different bioreactors. All those cell strains are described in the *Material and Methods* section. The MAE and RMSE approaches were utilized to determine the indications for the model's correctness. Furthermore, for all the three cell strains, the SGR estimation parameter β/α ratio remains the same (the ratio β/α = 0.04). The process data for the whole set of 20 cultivations was run via the SGR approach during the validation testing by using actual experiments. In a fed-batch culture, 17 processes were carried out with limited substrate feed, while 3 processes were carried out with unlimited dosage feeding. The results showed that the overall average RMSE of the SGR estimation was 0.074 1/h, and the overall average MAE of SGR was 0.044 1/h. Additionally, overall, the average MAPE of SGR was 9.77%. These findings demonstrate the suitability of this strategy for both restricted and unlimited fed-batch culture methods using different *E. coli* cell strains.

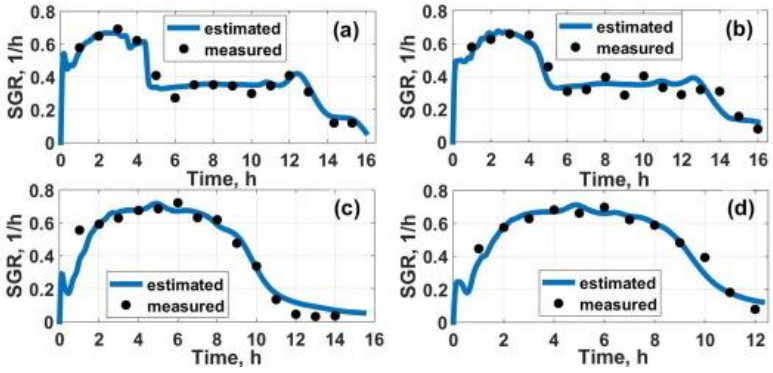


Figure 2.10. SGR estimation results with cultivation process data: (a–b) limited fed-batch cultivation processes; (c–d) unlimited fed-batch cultivation processes

2.5. Viable Cell Estimation of Mammalian Cells Using Off-Gas-Based Oxygen Uptake Rate and Aging-Specific Functional

The fourth published research article is *Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional* written by **Arnas Survyla**, Renaldas Urniezius, and Rimvydas Simutis. This research focuses on the estimation of the active biomass concentration – that is, on the estimation of viable cells from the oxygen uptake rate by using stoichiometric principles and the aging phenomena. This method works with a wider variety of cell types, including bacteria and mammalian cells. The primary innovation in this study is the application of the *aging* term to assess the state of long-term culture processes, where conventional approaches are ineffective.

2.5.1. Material and methods

This study's viable cell estimator for a mammalian cell culture was developed by using information from the growth of CHO-K1 (CHO-S, No. 11619-012, Karlsruhe, Germany). The cultivation procedures used in *Biostat B* bioreactors are

described in the study [79], and the details of the bioreactor system are shown in Table 2.8. Offline viable and total cell concentrations were measured by using an automated cell counter (*CASY TT*; *Roche Innovatis AG*, Mannheim, Germany). A quadrupole mass spectrometer (*Balzers QMA 200*; Balzers, Liechtenstein) was used to analyse exhaust gases.

Table 2.8. Cultivation process details of a mammalian cell

Condition	State	Condition	State
Bioreactor volume	2 L	Broth volume	1 L
Temperature	37 °C	pH	7.15
pO_2	20%	Airflow	0.1 L min^{-1}
Stirrer	60-400 RPM	Feeding start	At 75 h

Furthermore, the biomass was estimated by using information from bioprocesses involving the *Escherichia coli* strain. *E. coli* BL21 (DE3) pET21-IFN-alfa-5 bacteria were grown on a minimal-mineral medium (Table 2.9) [44]. The *BlueSens BlueinOne Ferm* was used for gas analysis.

Table 2.9. Cultivation process details of recombinant *E. coli*

Condition	State	Condition	State
Bioreactor volume	7 L	Broth volume	3.7 L
Temperature	37 °C	pH	6.8
pO_2	20%	Feeding start	At 5–7 h

2.5.2. Development of the viable cells estimation algorithm

The Luedeking-Piret-type model is a common exhaust gas based contender for the stoichiometric link between the total OUR and the development and maintenance of the biomass [16, 80]. In order to offer a general estimator for the number of viable cells, the Piret model is additionally altered. The time dependency of both kinetic parameters suggests a general inhomogeneous first-order differential equation [81].

$$X'(t) + \frac{\beta(t)}{\alpha(t)} \cdot X(t) = \frac{OUR(t)}{\alpha(t)}. \tag{2.37}$$

In Equation (2.37), X is the concentration of viable cells, t is the process duration time, α and β are variables establishing the corresponding stoichiometric relationship with the growth and maintenance of viable cells. Regarding article [42], a cumulative form of information is advisable to remove the disturbance effect on estimation. Hence, Equation (2.37) restructuring for viable cell concentration variable and apply integral is derived:

$$X(t) = \frac{X_0 + \int_0^t \frac{OUR(t_1)}{\alpha(t_1)} \, e^{\int_0^{t_1} \frac{\beta(t_2)}{\alpha(t_2)} dt_2} dt_1}{e^{\int_0^t \frac{\beta(t_3)}{\alpha(t_3)} \, dt_3}}. \tag{2.38}$$

In Equation (2.38), variable X_0 is the concentration of viable cells at the inoculation moment. According to study [42], in microbial bioprocesses, the

maintenance term is insignificant before the induction phase of biosynthesis. The age-related threshold of the viable cell concentration serves as the rational hypothesis to assume the beginning of the cell maintenance effect because there is no induction in the mammalian upstream development. The verge is described as follows in the interim expression: $k_{cX} \equiv \int_0^{t_{cX}} X(t)\, dt$, where t_{cX} is a time when the cumulative biomass concentration value reaches variable $k_{cX,}$ which is proposed in this study after analysis of the data. Hence, the final form of the number of the cell estimation combines Equation (2.38) and the maintenance term form based on the study [42]:

$$\begin{cases} X(t) = X_0 + \int_0^t \dfrac{OUR(t_1)}{\alpha(t_1)}\, dt_1; \quad if\ k_{cX} \geq \int_0^t X(t_1)\, dt_1 \\ X(t) = \dfrac{X_0 + \int_0^t \dfrac{OUR(t_1) + \beta(t) \cdot X_{cX}}{\alpha(t_1)}\, e^{\int_0^{t1} \frac{\beta(t_2)}{\alpha(t_2)}\, dt_2} dt_1}{e^{\int_0^t \frac{\beta(t_3)}{\alpha(t_3)}\, dt_3}}, \quad otherwise. \end{cases} \tag{2.39}$$

In Equation (2.39), variable X_{cX} is the cells concentration at the moment when the condition applies $k_{cX} = \int_0^{t_{cX}} X(t)\, dt$.

Based on the findings of studies [17, 82, 83], both kinetic parameters α and β of the Luedeking-Piret model are typically functions of time. However, stoichiometry values dependencies on time do not hold true when the target object of interest is mammalian cells, and cultivation processes are long-term ones in such a case. Then kinetic parameter dependencies are more akin to the average age of the cell population.

$$Age(t) = \frac{\int_0^t X(t_1)\, dt_1}{X(t)}. \tag{2.40}$$

The aging approach, in both fed-batch and continuous biosynthesis, is equally practical for non-invasive estimates. The age expression is more influenced by the condition than by the passage of time. Such a premise is pertinent for perfusion bioprocesses [84, 85], in which the biomass concentration (for microorganisms) or the viability rate (for mammals) may be age-invariant. Incorporating a parametric hypothesis for a fed-batch mammalian culture was the decision which we took. The model fitting classes used to enable a non-invasive online estimate of the kinetic coefficient *α(t)* at runtime were the following functionals:

$$\alpha(t) = \frac{\alpha_{max}}{1 - e^{-\frac{t}{Lag_{time}}}} \cdot \frac{Age(t)}{t}, \tag{2.41}$$

here, the theoretical aerobic oxidative capacity is represented by the maximum growth-based oxygen consumption yield (α_{max}) for cells, and the lag time (Lag_{time}) is connected to exponential decay [85] and designates the point at which cells transition from the lag phase to the exponential growth phase. The oxygen consumption yield for maintenance ($\beta(t)$) allows cells to remain alive, the expression which is dependent on the aging term as follows:

$$\beta(t) = \beta \cdot \frac{Age(t)}{Age(t) + k_{age}}. \quad (2.42)$$

In Equation (2.42), the aging-specific parameter k_{age} is the 'half-age-constant' when the maintenance coefficient β is treated as the maximal maintenance value.

As a result, Equation (2.41)–(2.42) demonstrates the direct dependence of the stoichiometry parameters on the aging component. These two expressions are incorporated into the viable cell estimate method created in Equation (2.39) to make the model completely applicable to both brief bacteria cultivations and extended fed-batch procedures.

2.5.3. Results

The active biomass concentration – the viable cells estimation approach – was tested with CHO mammalian cells and *E. coli* bacteria. The prediction accuracy is compared with the offline data by applying the *Mean Absolute Error* (MAE) and the *Root Mean Square Error* (RMSE) indicators Equations (2.10)–(2.12) [59]. Additionally, in order to find out the model's parameters to a CHO mammalian cell culture, cross-validation was employed [86]. The main goal of the approach that was selected was to obtain a result from the candidates that were supplied by averaging them according to a weight that relies on the item's importance. Specifically, by using the ensemble averaging equation, the weights of 10 candidate sets were based on RMSE:

$$\hat{y} = \sum_{i=1}^{n} w_i \, y_i(x), \quad (2.43)$$

$$w_i = \frac{\sum_{j=1}^{n} RMSE_j - RMSE_i}{\sum_{j=1}^{n} RMSE_j \cdot (n-1)}, \quad (2.44)$$

here, $\hat{y}$ is the final guess of the parameter, w_i is the weight of the parameter y_i, and n is the number of parameters' sets n = 10. The final CHO cell culture's parameter set is shown in Table 2.10.

Table 2.10. The final parameter set of the estimation model

Parameter	Value	Unit
Lag_{time}	20.489	h
α_{max}	0.727	g e^9cells^{-1}
β	0.034	g e^9 cells^{-1} h^{-1}
k_{cX}	29.99	e^9cells h L^{-1}
k_{age}	102.05	h

Hence, with the proposed parameter set for the CHO cells, the prediction approach yields accurate results. The average MAE and RMSE results were 0.139 and 0.158, respectively. Additionally, the average MAPE results were 5.15%. Figure 2.11 and the associated confidence bands show how well the model estimations performed. The confidence band of the procedure may be identified by categorizing the error values between the measured and computed points for 6 tests throughout a range of

viable cell concentrations. The confidence band $\alpha = 0.01$ in Figure 2.11 is shown with the purple shadow. The error statistics in this sector have a high (pessimistic) bound because it contains a systematic error of 0.1^9 cells L^- and a random error of 4%.

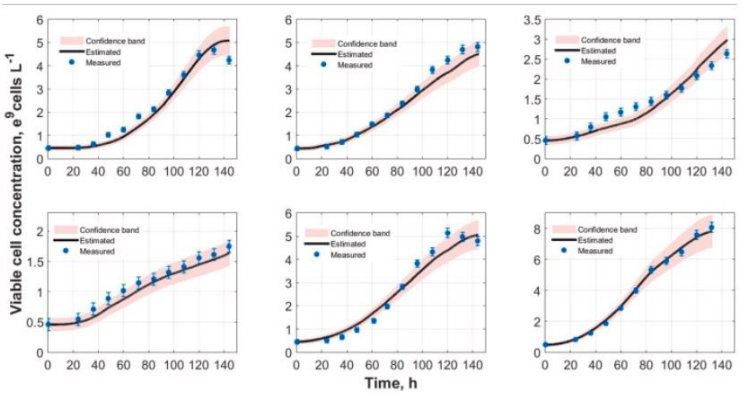


Figure 2.11. CHO viable cell estimation results from experiments Nos. 1–6. Vertical error bars indicate the total error. The purple shadow represents the prediction band

The same cross-validation technique was applied to the *E. coli* bacteria strain, which is also the test subject of this study. The data of a total of 12 experiments was analysed, and 10 sets of parameters were generated to establish the optimal values of the parameters. The final optimization results of the parameters are shown in Table 2.11.

Table 2.11. *E. coli* strain parameter set for the estimation approach

Parameter	Value	Unit
Lag_{time}	0	h
α_{max}	0.75	g g^{-1}
β	0.16	g g^{-1} h^{-1}
k_{cX}	17	g h L^{-1}
k_{age}	0	h

The results of the validation procedure were 1.78 g L^{-1} MAE, 2.53 g L^{-1} RMSE and 6.97% MAPE. An analogous process for microbial analysis is shown in Fig. 13, and the confidence band connection for bacterial analysis is the same as it is for mammalian analysis. The purple shadow indicates the confidence band $\alpha = 0.01$. The error bars consisted of a systematic error of 0.2 g L^{-1} and a random error of 4%. These errors reflect the bounds of experimentation-related errors and device characteristics.

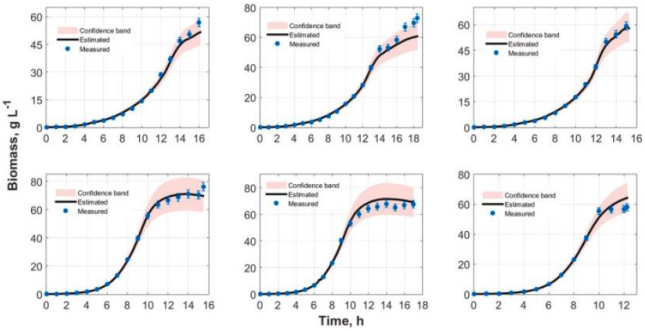


Figure 2.12. *E. coli* bacteria biomass estimation results of the six initial experiments. The vertical error bars indicate the total error. The purple shadow represents the prediction band

3. CONCLUSION

1. In this dissertation, a biomass estimation method based on stoichiometry and oxygen consumption has been proposed. The biomass concentration estimation algorithm is suitable for use with a wide spectrum of cell cultures: yeast, three different bacterial strains, and mammalian cells. These cultures are the main cells used in industry. Hence, the described algorithm has a great potential for marketing. It would solve the problem of sampling and increase the efficiency of human work. Concerning the method precision for yeast and bacteria, the average mean percentage error (MAPE) is 6.97%, while the average MAPE for mammalian cells is 5.15%.
2. This dissertation has proposed and applied a target protein modelling approach based on stoichiometry and oxygen consumption. The product synthesis model showed that the target product concentration depends on the specific growth rate at the time of induction and biomass increment. The aforementioned relationships allow guiding the process to the optimal conditions in the cultivation process for efficient potential protein production. The method's model fitting average MAE and MAPE accuracy is 0.099 g L^{-1} and 8.22%, respectively.
3. In this dissertation, a specific cells growth rate estimation method based on stoichiometry and oxygen consumption has been offered and put into application. The aforementioned approach has a single tuning parameter which varies based on the cell strain. This leads to its resilience and simplicity. This method combines well with control algorithms as it provides feedback information from the cultivation process. The accuracy of the method on MAPE is 9.77%, and it directly estimates SGR without using any additional observations.
4. It has been demonstrated in this dissertation that stoichiometry coefficients cannot be assumed as stationary in all contexts. The maintenance term's dependence on product synthesis is significant. In mammalian biosynthesis, the age of the culture participates in expressions describing the cells growth and maintenance. This is crucial to cultivation methods which are continuous or significantly prolonged. The proposed innovation allows the algorithms of this dissertation to be used in various types of growing processes: batch, fed-batch, or continuous. Furthermore, the algorithms are fit to be used in bioprocesses with inductor (ITPG).

4. FUTURE WORKS

The proposed aging term in biomass estimation showed that the age of the culture has an impact to the yields of oxygen consumption. In this dissertation, the aging term is significant in cultivation processes with mammalian cells because the growth was prolonged due to the low specific growth rate of the cell. The future work is to prolong the cultivation processes with *E. coli* bacteria and make the growth process in the continued mode, which implies that, at any moment, some of the media with biomass would be taken off from the bioreactor and replaced with a fresh media. In this case, the inhibition of an excessive amount of biomass would be avoided, and the age of the cell culture would continue to increase. Data of those cultivation processes would prove that the aging term for all the cell cultures is relevant.

5. SANTRAUKA

5.1. Įvadas

5.1.1. Tiriama problema

Bioreaktoriuje gliukozė yra gyvybiškai svarbus maistas įvairių tipų ląstelėms [87]. Gliukozė vartojama gaminant įvairius ląstelių produktus ir veikia kaip pagrindinis ląstelių metabolizmo ir anabolizmo energijos šaltinis [88]. Mažas gliukozės kiekis bioreaktoriuje lemia mažesnį augimo greitį [89]. Kita vertus, per didelis gliukozės kiekis sukelia kenksmingų šalutinių produktų susidarymą ir ląstelių neigiamus metabolizmo pokyčius, kurie gali turėti įtakos bendram ląstelių augimui ir produkto sintezei [90]. Tyrėjai turi kruopščiai subalansuoti gliukozės tiekimą pagal esamą ląstelių kultūros būseną bioreaktoriuje, kad optimizuotų ląstelių augimo greitį ir produkto gamybą [91]. Tam reikia stebėti tipinius kultivavimo proceso parametrus, tokius kaip gliukozės koncentracija bioreaktoriuje, santykinis ląstelių augimo greitis, biomasės koncentracija, produkto gamybos greitis [92]. Norint išlaikyti optimalų kultivavimo proceso efektyvumo lygį, pagal šiuos kintamuosius privaloma pritaikyti gliukozės tirpalo tiekimo greitį [92].

Tačiau be brangių aparatinės įrangos jutiklių neįmanoma išmatuoti esminių bioproceso parametrų realiuoju laiku. **Informacijos apie proceso būseną trūkumas yra pagrindinė biotechnologijų problema, proceso automatizavimas ir idealus valdymas neįmanomas be reikalingų grįžtamojo ryšio signalų.** Todėl šioje disertacijoje yra dėstomi būdai, kaip panaudoti netiesioginio įvertinimo algoritmus (programuojamieji jutikliai), kad būtų galima stebėti pagrindinius parametrus [93].

5.1.2. Tyrimo objektas

Trys pagrindiniai ląstelių kultivavimo proceso kintamieji, kurių negalima išmatuoti tiesiogiai, – specifinis ląstelių augimo greitis, biomasės koncentracija ir tikslinio produkto koncentracija.

5.1.3. Tyrimo tikslas ir iškelti uždaviniai

Tyrimo tikslas – padidinti kultivavimo proceso našumą pritaikant tikslius, patikimus įvertinimo modelius (programuojamus jutiklius), pagrįstus bioreaktoriaus išmetamosiomis dujomis, leidžiančius stebėti esminius augimo proceso parametrus, kurių negalima tiesiogiai įvertinti. Norint pasiekti šio tyrimo tikslus, buvo iškelti šie tikslai:

1. Sukurti ir ištirti biomasės koncentracijos įvertinimo algoritmą, kuris leistų sekti proceso būseną.
2. Sudaryti ir ištirti tikslinio produkto sintezavimo modelį ir pasiūlyti produkto sintezės išeigos priklausomybes.
3. Sukurti ir ištirti santykinio ląstelių augimo greičio įvertinimo metodą, skirtą pateikti informaciją valdymo sistemai.
4. Sudaryti ir ištirti įvertinimo algoritmą, skirtą gyvybingoms ląstelėms nustatyti,, žinduolių ląstelių kultūrai, kultivavimo būsenai sekti.

5.1.4. Mokslinis naujumas

Kiekvienoje šios disertacijos publikacijoje buvo pateikta reikšmingų mokslinių kintamųjų įvertinimo / prognozavimo naujovių, kurios pagrįstos fundamentaliomis žiniomis ir matematiniais metodais. Paprastumas, lengvas pritaikomumas, tikslumas ir nauja koncepcija, kuri turi įtakos algoritmui, yra kriterijai, įrodantys naujumą.

A1 straipsnio naujovė yra Luedeking-Piret modelio (1959 m.) [16] patobulinimas, pasiūlyta nauja stechiometrijos parametro, kuris aprašo deguonies suvartojimą ląstelių gyvybinėms funkcijoms palaikyti, išraiška. Remiantis pasiūlyta teorija, kai biomasės koncentracija nėra didelė, biomasės palaikymo narys proceso pradžioje yra pastebimai mažas, todėl jo galima nevertinti ir po indukcijos. Kai produkto sintezė pradedama arba kai biomasės kiekis viršija slenkstinę vertę, tada palaikymo narys turi būti vertinamas.

A2 publikacija yra **A1** darbo tęsinys, pasiūlytas naujumas yra biomasės palaikymo nario išskaidymas į dvi dalis: produkto sintezę ir biomasės palaikymo narį. Taip pat straipsnio kitas pasiūlytas naujumas susijęs su tikslinio produkto sintezės priklausomybe nuo santykinio ląstelių augimo greičio.

A3 straipsnyje pasiūlyto algoritmo unikalumas ir naujumas slypi jo atsikartojamume ir paprastume, nes jis turi tik vieną derinimo parametrą ir yra tinkamas naudoti viso auginimo proceso metu, įvairiomis sąlygomis ir esant įvairiems metabolizmo tipams. Siekiant pašalinti neapibrėžtumus ir perskaičiavimo klaidas, specifinis augimo greitis apskaičiuojamas tiesiogiai iš dujų duomenų (deguonies suvartojimo greičio), o ne pagal įvertintą biomasę.

A4 darbe pateikto metodo naujovė yra ląstelių senėjimo termino vartojimas, skirtas gyvybingoms ląstelėms įvertinti. Kultūros amžius turi įtakos auginimo proceso produktyvumui, taip pat sintezės dinamikai ir efektyvumui.

5.1.5. Tyrimo metodologija

Šio tyrimo pagrindinis tikslas yra įvertinti svarbius auginimo proceso kintamuosius, kurių negalima išmatuoti tiesiogiai. Norint užtikrinti algoritmų patikimumą ir paprastumą, Luedeking-Piret modelis [16] buvo pasirinktas kaip pagrindas. Kadangi biomasės koncentracija apibrėžia viso auginimo proceso būklę, todėl pirmasis tikslas buvo ištirti ir sukurti biomasės įvertinimo metodą (**A1**). Antra publikacija (**A2**) yra straipsnio (**A1**) tęsinys, joje siūlomas išplėstas biomasės koncentracijos algoritmas, papildytas baltymo sintezės kintamuoju. Taip pat tyrime pateikiama santykinio augimo greičio indukcijos metu įtaka produkto sintezės efektyvumui. Kadangi ląstelių specifinio augimo greičio kintamasis yra svarbus įėjimo signalas baltymų sintezės modeliui, trečias publikuotas straipsnis yra santykinio augimo greičio (SGR) įvertinimo algoritmas (**A3**). Paskutiniame straipsnyje (**A4**) aprašomas patobulintas biomasės koncentracijos algoritmas, galintis įvertinti gyvybingų ląstelių skaičių ir apimantis platesnį ląstelių kultūrų spektrą. Tyrime (**A4**) panaudotas ląstelių kultūros amžiaus kintamasis, kuris pateiktas kolegų parašytame straipsnyje [17].

5.1.6. Praktinė reikšmė

1. Šioje disertacijoje pateikti įvertinimo algoritmai yra reikalingi ir reikšmingi ne tik nacionaliniu, bet ir tarptautiniu mastu. Sukurti algoritmai gali būti naudingi įmonėms, užsiimančioms biotechnologine veikla, pavyzdžiui nacionalinėms įmonėms:
    a) *Northway Biotech*, UAB,
    b) *Celltechna*, UAB,
    c) *Thermo Fisher Scientific Baltics*, UAB,
    d) *Roquette Amilina*, AB.
2. Šiame darbe pateiktiems metodams galimybę vystyti sudarė Europos Sąjungos struktūrinių fondų finansuojamas projektas „Biotechnologinių procesų būsenos įvertinimo programinių jutiklių sistemos sukūrimas ir taikymas“ (Nr. 01.2.2-LMT-K-718), 2019–2023 m.

5.1.7. Tyrimo rezultatų aprobavimas

1. Pateikti metodai publikuoti 4 skirtinguose tarptautiniuose mokslo žurnaluose, turinčiuose cituojamumo rodiklį „Web of Science“ duomenų bazėje. Trys pasirinkti žurnalai yra Q1 lygio, o 1 žurnalas yra Q2 lygio.
2. Esminiai rezultatai pristatyti trijose tarptautinėse konferencijose.
3. Visi šiame tyrime pateikti įvertinimo algoritmai yra įdiegti ir naudojami Kauno technologijos universiteto bioprocesų modeliavimo ir valdymo laboratorijoje.
4. Inovatyvios medicinos centro (IMC) MTEP laboratorijoje įdiegtas ir pritaikytas biomasės koncentracijos įvertinimo algoritmas (**A1**).

5.1.8. Disertacijos ginamieji teiginiai

1. Pirmoje kultivavimo proceso pusėje (*lag* ir *log* fazės) ląstelių gyvybinių funkcijų palaikymo narys yra nereikšmingas ir turi minimalų poveikį įvertinimo procedūrų tikslumui.

1. Tikslinio produkto sintezės efektyvumas tiesiogiai priklauso nuo santykinio ląstelių augimo greičio indukcijos momentu.
2. Bioproceso būsenos įvertinimas remiantis deguonies suvartojimo greičiu yra priimtinas, jo atsikartojamumui neturi įtakos *E. coli* bakterijų metabolizmai.
3. Ląstelių kultūros vidutinis amžius yra reikšmingas kintamasis, kuris pagerina tikslinio produkto ir biomasės koncentracijos įvertinimų tikslumą.

5.2. Straipsnių apžvalga

5.2.1. Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate

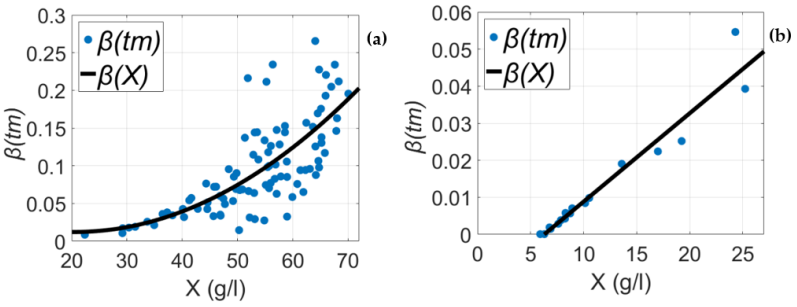
Vienas iš reikalingiausių kultivavimo proceso tiesiogiai nematuojamų parametrų yra biomasės koncentracijos įvertis. Todėl pirmojo išleisto straipsnio *Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate* tikslas – – suteikti galimybę nuolat stebėti bioproceso būklę ir atsisakyti mėginių ėmimo procedūros. Straipsnyje yra pateiktos dvi naujovės: biomasės palaikymo narys, kuris apibūdina, kiek deguonies suvartoja biomasė vien tik kvėpavimui (gyvybinių funkcijų palaikymui), nėra konstanta, kaip dažniausiai traktuojama kituose straipsniuose [16, 50]. Straipsnyje pasiūlyta biomasės palaikymo narį laikyti augančios vertės kintamuoju. Kita hipotezė yra tai, kad kultivavimo pradžioje palaikymo narys yra labai mažas, todėl galima jo nevertinti, ir tik susidarius tam tikram biomasės kiekiui, šis narys pradeda didėti. Šios hipotezės ypač aktualios kultivavimo procesuose, kuriuose yra naudojamas induktorius (isopropyl-D-1-thiogalactopyranoside/IPTG), nes šis preparatas suaktyvina tikslinio baltymo sintezę, todėl proceso dinamika staigiai kinta [57]. Didėjančiame biomasės palaikymo kintamajame yra įvertintas deguonies suvartojimas baltymų sintezei. Ši prielaida leidžia šiame straipsnyje pateikti nesudėtingą matematinį biomasės koncentracijos įvertinimo metodą, kuris sudarytas iš dviejų dalių: biomasės palaikymo įverčio ir ląstelių dalijimosi. Pasiūlyto algoritmo privalumai yra tikslumas ir nesudėtinga algoritmo forma, kuri tinkama naudoti biomasės koncentracijai įvertinti viso auginimo proceso metu.

Šio straipsnio algoritmo pagrindą sudaro Luedeking-Piret matematinė išraiška, kuri apibūdina ryšį tarp deguonies suvartojimo ir stechiometrijos parametrų [16]:

$$OUR(t) = \alpha \cdot X'(t) + \beta \cdot X(t), \tag{5.1}$$

čia α ir β yra stechiometrijos parametrai, *OUR* yra deguonies suvartojimo greitis, *X* yra biomasės koncentracija. Koeficientas α apibūdina deguonies suvartojimą ląstelių populiacijos augimui ir β yra deguonies suvartojimas gyvybinėms funkcijoms palaikyti. Naudojant klasikinę Piret išraišką su duomenimis, gautais iš eksperimentų, , kurių metu buvo atlikta indukcija, gauti rezultatai turi dideles paklaidas. Todėl pradėtas tyrimas siekiant išsiaiškinti dinamiką deguonies suvartojimo biomasės palaikymui ir jo priklausomybė nuo biomasės koncentracijos prieš ir po indukcijos. Šiam tyrimui buvo naudojami duomenys, surinkti iš kultivavimo procesų eksperimentų, kurių metu buvo naudojamos keturios skirtingos ląstelių kultūros: mielės ląstelės S. cerevisiae (no DY7221), E. coli BL21(DE3) pET9a-IdeS, E. coli BL21 (DE3) pET21-IFN-alfa-5, E. coli BL21(DE3) pLysS. Ląstelių įvairovė tyrime buvo pasirinkta dėl modelio rezultatų patikimumo ir universalumo. Deguonies suvartojimo biomasei palaikyti tyrimas buvo vykdomas iš bendro deguonies suvartojimo greičio eliminavus deguonies suvartojimą gyvybinėms funkcijoms palaikyti β: $OUR(t) = \alpha \cdot X'(t)$. Šiam tyrimui atlikti buvo pasitelktos biomasės koncentracijos vertės iš eksperimentinių kultivavimo proceso. Rezultatai

pademonstravo, kad deguonies suvartojimas biomasei palaikyti kultivavimo proceso pradžioje lygus nuliui ir esant kritiniam biomasės kiekiui, šis kintamasis pradeda didėti. Palaikymo dinamika priklauso nuo ląstelių kultivavimo proceso, kaip parodyta 5.1 paveiksle.



5.1. pav. Biomasės koncentracijos įtaka deguonies suvartojimui, skirtam biomasei palaikyti, (a) *E. coli,* (b) *S. cerevisiae*

Pagal gautus rezultatus, straipsnyje pasiūlyta galutinės biomasės koncentracijos įvertinimo modelis, kuriame gyvybinių funkcijų palaikymas yra vertinimas nuo tam tikros biomasės koncentracijos ir jo forma aprašoma antros eilės polinomu:

$$X_m = \frac{cOUR_m}{\alpha} + X_0, \tag{5.2}$$

$$X_m \cong \frac{cOUR_m - \sum_{l=i}^{m-1} \beta(X_l) \cdot X_l \cdot \Delta t_{l,l-1}}{\alpha} + X_0, \tag{5.3}$$

$$\beta(X) \equiv \beta\big(X(t)\big) = k_{\beta 2} \cdot X^2(t) + k_{\beta 1} \cdot X(t) + k_{\beta 0}. \tag{5.4}$$

Šiose formulėse X_0 yra biomasės koncentracija inokuliavimo momentu, *cOUR* yra deguonies suvartojimo greičio integralas: $cOUR = \int_0^t OUR(t)\, dt$. $k_{\beta 2}$, $k_{\beta 1}$, $k_{\beta 0}$ yra biomasės palaikymo koeficientai, pateikti išraiškoje (5.3), kurie priklauso nuo ląstelių kultūros. Biomasės palaikymo koeficientai gauti naudojant duomenis iš kultivavimo procesų ir pritaikant iškilaus optimizavimo metodą [42].

5.2.2. Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process

Toliau tęsiant tyrimą susijusį su deguonies suvartojimu biomasei palaikyti ir jo dinamikos priklausomybe, gauti rezultatai publikuoti straipsnyje *Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process.* Šiame darbe toliau tęsiamas tyrimas tarp deguonies suvartojimo ir stechiometrijos [16, 42]. Anksčiau aprašytame straipsnyje biomasės palaikymo narys apėmė ir produkto sintezės deguonies suvartojimą, kad būtų galima išlaikyti modelio paprastumą [42]. Tačiau, norint padidinti biomasės koncentracijos prognozavimo tikslumą, produkto įvertinimas realiu laiku yra neišvengiamas [46]. Šiame darbe pasiūlyta išplėsta Luedeking-Piret modelio išraiška [16]:

$$OUR(t) = \alpha \cdot X'(t) + \beta \cdot X(t) + \gamma \cdot P'(t), \quad (5.5)$$

čia γ yra produkto sintezės išeigos parametras, nusakantis, kiek deguonies reikia sintetinant baltymus, P yra produkto koncentracija. Ši išraiška galioja daugumai kultivavimo procesų, taip pat ir procesuose, kai yra naudojamas induktorius [56]. Šiame straipsnyje iš esmės yra analizuojami ląstelių kultivavimo procesai, kuriuose yra naudojamas IPTG preparatas, ir jo įtaka produkto sintezei. Tyrimo objektas šiame darbe yra rekombinantinės *E. coli* BL21 (DE3) pET28a bakterijos [57]. Iki induktoriaus suleidimo į bioreaktorių bakterijos negamina produkto, tik dauginasi. Todėl iki indukcijos deguonies suvartojimas produkto sintezei yra laikomas lygus nuliui ($P'(t) = 0$) [94]. Tyrimo eksperimentuose induktorius buvo suleidžiamas skirtingais momentais. Tai leido aptikti dėsningumą, kad induktoriaus suleidimo momentu biomasės koncentracija bioreaktoriuje turi įtakos produkto sintezei $\gamma(X)$ [66]:

$$\gamma(X) = k_\gamma \cdot (X(t) - X_{ind}), \quad (5.6)$$

čia k_γ yra produkto skyrimosi išeigos koeficientas, kuris yra konstanta, X_{ind} biomasės koncentracija indukcijos metu. Sujungę straipsnio biomasės palaikymo naują išraišką [42] ir formulėje (5.6) pateiktą produkto sintezės išraišką, gauname visą deguonies suvartojimo greičio išraišką:

$$\begin{aligned}\int_{t_0}^{t} \mathrm{OUR}(t^*)dt^* = \alpha \cdot \int_{t_0}^{t} \mathrm{X}'(t^*)dt^* + k_\gamma \cdot \int_{t_0}^{t} (\mathrm{X}(t^*) - \mathrm{X_{ind}}) \cdot \mathrm{P}'^{(t^*)}dt^* \\ + \int_{t_0}^{t} \left(k_{\beta 2} \cdot \mathrm{X}^2(t^*) + k_{\beta 1} \cdot \mathrm{X}(t^*) + k_{\beta 0}\right) \cdot \mathrm{X}(t^*)dt^*.\end{aligned} \quad (5.7)$$

E. coli bakterijos eksperimentų duomenys buvo paimti iš pramoninių bioproceso vystymo laboratorijų, kuriose induktorius buvo suleidžiamas kultivavimo procesų logaritminėje fazėje [48] (esant apie 30 g/L biomasės koncentracijai), kultūrai nepasiekus stacionariosios fazės [91]. Atliekant tyrimus su pateikta visa deguonies suvartojimo formule (5.7), pastebėta, kad biomasės palaikymo narys yra minimalus, nes, esant nedidelei biomasės koncentracijai suleidus induktorių, deguonis iš esmės yra suvartojimas produkto sintezei ir minimaliam ląstelių dalijimuisi:

$$\int_{t_0}^{t} \mathrm{OUR}(t^*)dt^* = \alpha \cdot \int_{t_0}^{t} \mathrm{X}'^{(t^*)}dt^* + k_\gamma \cdot \int_{t_0}^{t} (\mathrm{X}(t^*) - \mathrm{X_{ind}}) \cdot \mathrm{P}'^{(t^*)}dt^*. \quad (5.8)$$

Rezultatai parodė, kad deguonies suvartojimas gyvybingoms funkcijoms palaikyti yra labai mažas, kad pagrindiniai kultivavimo proceso veiksniai ląstelių dalyba ir baltymo sintezė nustelbia. Galutinis šiame straipsnyje pasiūlytas biomasės koncentracijos įvertinimo modelis išlieka nesudėtingos formos:

$$X_m = \frac{cOUR_m + \alpha \cdot X_0 + k_\gamma \cdot P_m \cdot X_{ind} + k_\gamma \cdot \sum_{l=1}^{m}(X_l - X_{l-1}) \cdot P_l}{\alpha + P_m \cdot k_\gamma}. \quad (5.9)$$

Šiame straipsnyje pasiūlytas ir kitas naujumas – produkto koncentracijos įvertinimo modelis. Kadangi biomasės koncentracijos įverčio algoritmui yra reikalinga produkto koncentracijos vertė, todėl vienu metu buvo vystomi du

prognozavimo modeliai. Produkto koncentracijos įvertinimo algoritmo idėja paimta iš darbo [68], kai produkto sintezė priklauso nuo santykinio ląstelių augimo greičio:

$$\frac{\mathrm{d}P_x}{\mathrm{dX}} = q_{px}(\mu, P_x), \tag{5.10}$$

čia P_x yra baltymų koncentracija, padalinta iš biomasės koncentracijos $P_x(t) = P(t)/X(t)$, q_{px} yra santykinis produkto kaupimosi greitis (U/g/h), μ yra santykinis ląstelių augimo greitis (1/h) [68]. Atlikus tyrimus su eksperimentų duomenimis, pastebėta, kad produkto sintezė yra slopinama produkto koncentracijos bioreaktoriuje. Tai yra biomasės koncentracija po indukcijos turi ribą, kiek produkto galima pagaminti [69]. Todėl produkto sintezės greitis mažėja, didėjant baltymų koncentracijai:

$$\frac{dP_x}{dx} = q_{px}(\mu, P_x) = P_{max}(\mu, X) - k_t \cdot P_X(t), \tag{5.11}$$

čia kintamasis k_t yra nuo laiko nepriklausomas koeficientas, kuris nulemia savaiminio slopinimo elgseną [70], P_{max} yra maksimali santykinė produkto koncentracija. Šis kintamasis tiesiogiai priklauso nuo μ santykinio ląstelių augimo greičio indukcijos metu. Vėliau, po indukcijos, P_{max}, maksimali tikslinio produkto koncentracija, nustatoma pagal santykinį biomasės augimo greitį ir biomasės koncentraciją:

$$\begin{gathered} P_{max}(\mu, X) = \mu(t) \cdot \left(k_{m0} + k_{m1} \cdot (X(t) - X_{ind})\right),\ t \geq t_{induction}, \\ P_{max}(\mu, X) = \mu(t) \cdot k_{m0}, \qquad t < t_{induction}, \end{gathered} \tag{5.12}$$

čia k_{m0} ir k_{m1} yra šio tyrimo pasiūlyti empiriniai parametrai, k_{m0} koeficientas, susijęs su augimo greičiu ir baltymų sinteze, o k_{m1} susijęs su biomasės koncentracija indukcijos momentu [65], $t_{induction}$ yra indukcijos momentas. Šiame straipsnyje visuose pagrindiniuose įvertinimo modeliuose yra naudojama kumuliatyvinė informacija dėl jos savybės sumažinti triukšmo įtaką įverčiams. Todėl ir produkto prognozavimo algoritmo ((5.11) formulė) galutinė lygtis yra išreikšta per integralą:

$$\mathrm{P_m} = \frac{\left(\sum_{\mathrm{j=i}}^{\mathrm{m}} \mathrm{P_{max,j}} \cdot \Delta \mathrm{t_{j,j-1}} - \mathrm{k_t} \cdot \sum_{\mathrm{j=1}}^{\mathrm{m-1}} \mathrm{P_{x,j}} \cdot \Delta \mathrm{t_{j,j-1}}\right) \cdot \mathrm{X_m}}{1 + \Delta \mathrm{t_{m,m-1}} \cdot \mathrm{k_t}}. \tag{5.13}$$

Straipsnyje taip pat pristatomas parametrų identifikavimo procesas, naudojant išgaubto optimizavimo metodą. Taip pat pasiūlytas naujas kriterijus, skirtas rasti parametrų reikšmėms, kurios galiotų viso kultivavimo proceso metu:

$$\sigma^2_{X,m} \sim \frac{X_m^2}{1 - K_{exp} + X_m^2 \cdot K_{exp}}, \tag{5.14}$$

čia K_{exp} yra derinimo koeficientas, kuris reikalingas pakoreguoti neapibrėžtumui ir sujungti du optimizavimo kriterijus: mažiausių kvadratų metodo ir vidutinės procentinės paklaidos. Kai K_{exp} kintamojo vertė lygi 1($0 \leq K_{exp} < 1$), tai kriterijaus išraiška yra lygi mažiausių kvadratų metodui. O kintamojo reikšmei prilygus 0, kriterijus tampa lygus vidutiniam procentiniam paklaidos metodui.

5.2.3. An oxygen-uptake-rate-based estimator of the biomass-specific growth rate in microbial cultivation processes

Pirmiau pristatyti du prognozavimo metodai straipsniuose pateikė labai svarbių kultivavimo proceso kintamųjų skaičiavimo būdus [42, 43]. Turint šiuos pagrindinius kintamuosius: biomasės koncentraciją ir produkto koncentraciją, tolesnis žingsnis automatizuojant ląstelių auginimus bioreaktoriuje yra automatinio valdymo sistema. Šios sistemos svarbi dedamoji yra grįžtamasis ryšys iš valdymo objekto. Šis signalas turi būti tikslus ir atspindintis proceso dinamiką. Biomasės koncentracija, kaip grįžtamasis ryšys, gautas iš anksčiau pateiktų programinių jutiklių [42, 43], nėra tinkamas. Priežastis yra ta, kad šio kintamojo dinamika yra lėta ir jo, kaip grįžtamojo ryšio signalo, tinkamumas mažėja, ypač esant didelėms koncentracijoms [46]. Neseniai išleistose publikacijose yra pastebėta, kad santykinis ląstelių augimo greitis yra patikimesnis signalas grįžtamajam ryšiui valdiklyje [88]. Straipsnyje *An oxygen-uptake-rate-based estimator of the biomass-specific growth rate in microbial cultivation processes* yra pasiūlytas algoritmas, skirtas santykiniam ląstelių augimo greičiui apskaičiuoti tiesiogiai iš deguonies suvartojimo greičio. Pagrindinis naujumas yra santykinio ląstelių augimo greičio įvertinimo modelis, kuris tiesiogiai iš dujų, naudojant tik vieną derinimo parametrą, pateikia priimtiną rezultatą.

Šame darbe modelis buvo vystomas remiantis dviem klasikiniais modeliais, Luedeking-Piret ir biomasės augimo greičiu [16, 72]:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu \cdot X(t). \tag{5.15}$$

Siekiama gauti santykinio augimo greičio įvertinimo išraišką, kurioje nebūtų biomasės koncentracijos dedamosios. Todėl sujungiant Piret ir biomasės augimo greičio formules, biomasės koncentracija išreikšta per deguonies suvartojimą ir stechiometrijos parametrus:

$$\frac{1}{OUR(t)} \cdot \frac{\mathrm{d}OUR(t)}{\mathrm{d}t} = \frac{1}{\mu + \beta/\alpha} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}t} + \mu. \tag{5.16}$$

Kitas žingsnis yra supaprastinti modelį taip, kad šis būtų patikimas, nesudėtingas ir dinamiškai adekvatus proceso elgsenai. Šiam tikslui (5.16) išraiška yra skiriama į dvi dalis: dinaminę dalį ir fundamentaliąją dalį, susiejant deguonies suvartojimo greitį su ląstelių augimo greičiu. Dinaminė dalis yra įvardinta kaip *R* kintamasis:

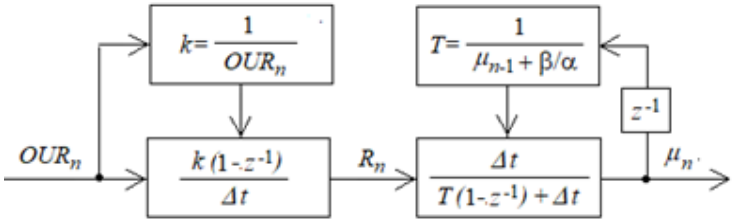
$$R = \frac{1}{OUR(t)} \cdot \frac{\mathrm{d}OUR(t)}{\mathrm{d}t}; \rightarrow \frac{1}{\mu + \beta/\alpha} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}t} + \mu = R. \tag{5.17}$$

Fundamentalioji dalis yra įvardinama kaip *T* kintamasis:

$$T = \frac{1}{\mu + \beta/\alpha}. \tag{5.18}$$

Paskutinis žingsnis yra z transformacijos pritaikymas. Šio metodo tikslas - pritaikyti algoritmą diskretiniams deguonies suvartojimo greičio matavimams (*OUR*) tam, kad sėkmingai įvertintume santykinį biomasės koncentracijos augimo greitį. 5.2

paveiksle pavaizduotas visas diskretinis santykinio ląstelių augimo greičio įverčio algoritmas.



5.2. pav. SGR įvertinimo algoritmo blokinė schema (z^{-1} – poslinkio atgal operatorius, Δt – laiko diskretizavimo žingsnis, n – diskretizacijos taškų skaičius)

Ryšys tarp deguonies suvartojimo dinamikos ir laiko konstantos, lemiančios santykinio augimo greičio reikšmę, pavaizduotas struktūrinėje diagramoje. Pritaikius z transformaciją lygčiai, gaunama galutinė modelio formulė:

$$\mu_n = R_n \cdot \frac{\Delta t}{T + \Delta t} + \mu_{n-1} \cdot \frac{T}{T + \Delta t}. \tag{5.19}$$

Galutinė algoritmo išraiška yra supaprastinta, lengvai integruojama į bet kokią valdymo sistemą ir gali būti naudojama įvairių ląstelių kultūrų auginimo procesams sekti. Stechiometrijos koeficientų santykis β/α yra tik vienas metodo derinimo parametras, kuris būdingas tik tam tikrai mikroorganizmų padermei ir gali būti rastas žinynuose arba nustatytas atlikus išankstinius kartotinius kultūros kultivavimo eksperimentus [42].

Kartu su santykinio ląstelių augimo greičio įvertinimo modeliu pristatytas kompiuterinis proceso elgsenos modeliavimas, kurio metu buvo testuojamas šio straipsnio pagrindinis netiesioginis būsenos įvertinimo algoritmas. Šio modeliavimo tikslas yra grubiai atkurti bioreaktoriuje vykstančius procesus ir ištestuoti santykinio augimo greičio įvertinimo algoritmo veikimą kritinėmis situacijomis, kai šio greičio nuostatas staigiai operatoriaus pakeičiamas arba pakinta dėl išorinių sąlygų. Kompiuterinio modeliavimo pagrindą sudaro trys pagrindinės diferencialinės lygtys: biomasės augimo, gliukozės suvartojimo ir bioreaktoriaus darbinio tūrio kitimas [75, 76].

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu(s) \cdot X - F \cdot \frac{X}{V}, \tag{5.20}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -q_s(s) \cdot X + F\frac{s_f - s}{V}, \tag{5.21}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = F. \tag{5.22}$$

Šiose diferencialinėse lygtyse F kintamasis yra gliukozės substrato tiekimo greitis g/h, q_s gliukozės suvartojimo greitis g/(g h), V bioreaktoriaus tūris L, s gliukozės koncentracija bioreaktoriuje g/L, S_f gliukozės koncentracija pamaitinimo tirpale g/L. Šios trys pagrindinės diferencialinės lygtys aprašo bioreaktoriuje vykstančius pagrindinius reiškinius, nesigilinant į pašalinius metabolinius veiksnius,

kurie netrukdo santykinio ląstelių augimo įvertinimo modelio adekvatumo patikrinimui. Kadangi šio straipsnio modelis yra paremtas deguonies suvartojimu, pateiktos papildomos trys lygtys, kurios kartu atkartoja ląstelių augimo greičio elgseną, kuri priklauso nuo gliukozės koncentracijos bioreaktoriuje, gliukozės suvartojimo ir klasikinės Piret išraiškos (formulė (5.1)) [78, 79]:

$$\mu(s) = \mu_{\max} \cdot \frac{s}{k_s + s} \cdot \frac{k_i}{k_i + s}, \tag{5.23}$$

$$q_s(s) = \frac{\mu(s)}{Y_{x/s}} + m\,. \tag{5.24}$$

Paskutinėse dviejose lygtyse, μ_{max} yra maksimalus galimas ląstelių kultūros augimo greitis, k_s, k_i yra Monod išraiškos koeficientai [78], m yra gliukozės suvartojimo parametras biomasei palaikyti, $Y_{x/s}$ yra biomasės koncentracijos išeigos koeficientas. Pateiktas ląstelių kultivavimo proceso skaitmeninis dvynys yra labai patogus naujų matematinių modelių testavimo įrankis. Skaitmeninio dvynio privalumas yra galimybė testuoti kuriamus modelius kritinėse situacijose, neatliekant sudėtingų eksperimentų.

5.2.4. Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional

Publikavus du straipsnius apie biomasės koncentraciją ir pradėjus juos naudoti, rezultatai atitiko spaudoje pateiktus tikslumus, kai objektas yra mikroorganizmai, kurių santykinis augimo greitis yra didesnis nei 0,1 h^{-1}, ir kultivavimo procesas vyksta iki kelių parų. Tačiau pateikti modeliai yra netinkami naudoti su žinduolių ląstelėmis, kai šios kultūros auginimas vyksta mėnesiais [79]. Todėl, iškilus poreikiui, pradėtas tiriamasis darbas, kuris aprašytas ketvirtame straipsnyje *Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional*. Šis tyrimas skirtas aktyviai biomasei – gyvybingoms ląstelėms įvertinti pagal deguonies įsisavinimo greitį, naudojant stechiometrinius principus ir ląstelių kultūros senėjimo fenomeną. Šio straipsnio pagrindinis naujumas – vidutinis ląstelių populiacijos amžius, kurio taikymas skirtas ilgalaikių kultūros procesų būklei įvertinti, pavyzdžiui, žinduolių ląstelėse, kur įprasti metodai yra netinkami [80, 81]. Amžiaus kintamojo naudojimas leido straipsnio [42] modelį padaryti dar universalesnį, tinkamą naudoti plačiam ląstelių tipų diapazonui, įskaitant ir žinduolių ląsteles.

Remiantis tyrimų išvadomis [17, 82, 83], abu Luedekingo-Piret modelio kinetikos parametrai α ir β paprastai yra laiko funkcijos. Tačiau stechiometrinių verčių priklausomybės nuo laiko negalioja, kai objektas yra žinduolių ląstelės ir auginimo procesai užsitęsia. Tada kinetinių parametrų priklausomybės yra labiau panašios į vidutinį ląstelių populiacijos amžių.

$$Age(t) = \frac{\int_0^t X(t_1)\, dt_1}{X(t)}. \tag{5.25}$$

Amžiaus išraiškai didesnę įtaką daro proceso būklė, o ne laiko eiga [85]. Atlikus tyrimą su žinduolių ląstelių auginimo proceso duomenimis, nuspręsta įtraukti parametrinę hipotezę žinduolių kultūrai, kai kinetinis koeficientas *α(t)* kinta kultivavimo proceso metu pagal dėsningumą:

$$\alpha(t) = \frac{\alpha_{max}}{1 - e^{-\frac{t}{Lag_{time}}}} \cdot \frac{Age(t)}{t}, \tag{5.26}$$

čia α_{max} kintamasis nusako maksimalią oksidatyvinę galią ląstelių dauginimuisi, Lag_{time} kintamasis yra susijęs su ląstelių *lag* faze, t.y. per kiek laiko ląstelės užtrunka pereiti iš *lag* fazės į logaritminę fazę [91]. Deguonies suvartojimas aktyvios biomasės – gyvybingų ląstelių gyvybinių funkcijoms palaikyti (*β(t)*) leidžia ląstelėms išlikti gyvybingoms, o šio kintamojo priklausomybė nuo senėjimo nario pateikta žemiau:

$$\beta(t) = \beta \cdot \frac{Age(t)}{Age(t) + k_{age}}. \tag{5.27}$$

(5.27) lygtyje senėjimui būdingas parametras k_{age} yra „pusės amžiaus konstanta“, o gyvybinių funkcijų palaikymo koeficientas β laikomas maksimalia verte.

Šio straipsnio įvertinimo modelio pagrindą sudaro [42] straipsnio biomasės įvertinimo išraiška, kai deguonies suvartojimas aktyviai biomasei – gyvybingoms ląstelėms palaikyti pradedamas vertinti, kai aktyvi biomasė – gyvybingų ląstelių kiekis – pasiekia kritinę vertę:

$$\begin{cases} X(t) = X_0 + \int_0^t \frac{OUR(t_1)}{\alpha(t_1)}\, dt_1; \quad kai\ k_{cX} \geq \int_0^t X(t_1)\, dt_1\ , \\ X(t) = \frac{X_0 + \int_0^t \frac{OUR(t_1) + \beta(t) \cdot X_{cX}}{\alpha(t_1)}\ e^{\int_0^{t_1} \frac{\beta(t_2)}{\alpha(t_2)} dt_2} dt_1}{e^{\int_0^t \frac{\beta(t_3)}{\alpha(t_3)}\, dt_3}}, \quad kitaip. \end{cases} \tag{5.28}$$

(5.28) formulėje k_{cX} yra parametras, kuris apibrėžia laiko momentą, nuo kurio išraiškoje (5.25) pradeda dalyvauti gyvybinių funkcijų palaikymo narys *β*.

5.3. Rezultatai

Visuose šios disertacijos straipsniuose lyginimui yra naudojami kriterijai: vidutinė absoliuti paklaida (MAE), vidutinė absoliutinė procentinė paklaida (MAPE) ir vidutinė kvadratinė paklaida (RMSE), kad būtų galima gautus rezultatus lyginti ir įvertinti.

Pirmuosiuose dviejuose straipsniuose **A1 (**Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate) ir **A2 (**Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process**)** aprašyti du biomasės koncentracijos įvertinimo metodai, kurių pagrindą sudaro deguonies suvartojimas ir stechiometrija. **A1** straipsnyje pateiktas modelis, kai deguonies suvartojimas ląstelių gyvybei palaikyti kultivavimo proceso pradžioje yra nevertinamas, ir tik biomasės kiekiui viršijant kritinę biomasės vertę, šis kintamasis pradeda galioti, ir jo

priklausomybė nuo biomasės yra laikoma tiesiškai priklausoma. Antrajame straipsnyje **A2** modelyje pateikta, kad kultivavimo procesuose su indukcija (IPTG) deguonies suvartojimo biomasei palaikyti galima nevertinti, nes IPTG produktas yra suleidžiamas logaritminėje fazėje, kai deguonies suvartojimas biomasei palaikyti yra lygus nuliui. Įrodyta, kad vietoj šio kintamojo galima naudoti baltymų sintezės greičio kintamąjį ir išeigos koeficientą deguonies suvartojimui. Abiejuose darbuose pasitelkta *E. coli* bakterija kaip eksperimentinis objektas. 5.1 lent. pateikti šios kultūros stechiometrijos parametrai, kurie buvo optimizuoti pagal dviejų straipsnių modelius, taip pat pateikiant šių dviejų biomasės koncentracijos įvertinimo modelių eksperimentų rezultatų vidurkius pagal MAE ir MAPE kriterijus. Abiejų biomasės koncentracijos modelių rezultatai grafiškai pavaizduoti Figure 2.4-2.5 ir Figure 2.6 (rezultatų skiltyje ir toliau vartojamos nuorodos į angliškoje versijoje pavaizduotas iliustracijas „Figure“).

5.1. lentelė. Biomasės koncentracijos modelių rezultatų palyginimas

Model	α	$k_{\beta 0}$	$k_{\beta 1}$	$k_{\beta 2}$	k_{γ}	MAE	MAPE
A1	0.996	0.07	0.00084	0	—	1.1	7.28%
A2	0.997	0	0	0	2.705	0.68	7.09%

Taip pat **A2** straipsnyje aprašytas produkto koncentracijos įvertinimo modelis. Pademonstruota idėja, kad maksimali produkto koncentracija priklauso nuo santykinio ląstelių augimo greičio indukcijos momentu ir biomasės koncentracijos pokyčio po indukcijos. Modelio įrodymas buvo atliktas su *E. Coli* BL21 (DE3) pET28a bakterijos duomenimis. Pateikti optimalūs proteino koncentracijos įvertinimo modelio parametrai su *E. coli* BL21 bakterija: k_{m0} = 0,2346, k_{m1} = -0,0172, k_t = 0,0687. 5.2 lent. pateiktas baltymo koncentracijos prognozavimo algoritmo visų eksperimentų rezultatų vidurkis pagal RMSE kriterijų. Šis rezultatas 5.2 lent. taip pat palygintas su tuo metu buvusiais naujausiais straipsniais, kurių autoriai tyrė netiesioginį produkto koncentracijos įvertinimą. Produkto koncentracijos įvertinimo algoritmo rezultatai grafiškai iliustruoti Figure 2.7.

5.2. lentelė. Modelio rezultatų palyginimas su kitų tyrėjų projektais, kuriuose pagrindinis dėmesys skirtas produkto prognozavimui

Straipsniai	RMSE (g)		
	Visa biomasė	Visas tirpus produktas	Visas netirpus produktas
Tradicinis modelis iš Gnoth et al. [62]	10,81	1,78	0,87
Hibridinis modelis iš Gnoth et al. [62]	4,71	1,28	0,62
A2 straipsnio modelis	4,577	—	0,656

Trečiame šios disertacijos straipsnyje **A3** (*An oxygen-uptake-rate-based estimator of the biomass-specific growth rate in microbial cultivation processes*) pateiktas santykinio ląstelių augimo greičio modelis, kurio pagrindą sudaro deguonies suvartojimo greitis (*OUR*) ir vienintelis derinimo parametras β/α, kuris priklauso nuo

ląstelių kultūros. Publikacijos modelio veiksnumas ištirtas naudojantis trimis skirtingomis bakterijų kultūromis: *E. coli* BL21(DE3) pET9a-IdeS, *E. coli* BL21(DE3) pET21-IFN-alfa-5, *E. coli* BL21(DE3) pLysS. Atlikus tyrimus, pateikta optimali derinimo parametro vertė β/α = 0,04, kuri galioja visoms ląstelių kultūroms, tirtoms šiame straipsnyje. Iš viso 20 kultivavimo proceso duomenų buvo testuojami naudojant SGR įvertinimo metodą. Rezultatai parodė, kad bendras vidutinis SGR įvertinimo RMSE buvo 0,074 1/h, bendras vidutinis SGR MAE buvo 0,044 1/h ir bendras vidutinis MAPE buvo 9,77 %. Keli kultivavimo proceso eksperimentai ir metodo rezultatai grafiškai pavaizduoti Figure 2.9–2.10.

Ketvirtame šios disertacijos straipsnyje **A4** (*Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional*) įrodytas gyvybingų ląstelių – aktyvios biomasės koncentracijos įvertinimo modelio veiksnumas su įvairaus tipo ląstelėmis (žinduolių ląstelės, bakterijos). Pateiktas modelis ištestuotas su dviejų tipų ląstelių kultūromis: CHO-K1 žinduolių ląstelės ir *E. coli* BL21 (DE3) pET21-IFN-alfa-5 bakterija. Pademonstruota ląstelių populiacijos vidutinio amžiaus įtaka stechiometrijos parametrams β,α pagal lygtis (2.40) ir (2.41). Kartu su gyvybingų ląstelių – aktyvios biomasės – koncentracijos įvertinimo modelio išraiška pateikti ir tiriamų kultūrų optimalūs parametrai, pavaizduoti 5.3 lent.

5.3. lentelė. CHO žinduolių ląstelių ir *E. coli* bakterijų parametrų rinkiniai prognozavimo algoritmui

CHO žinduolių ląstelė			***E. coli* bakterija**		
Parametras	**Vertė**	**Vienetai**	**Parametras**	**Vertė**	**Vienetai**
Lag_{time}	20,489	H	Lag_{time}	0	h
α_{max}	0,727	g e^9cells^{-1}	α_{max}	0,75	g g^{-1}
β	0,034	g e^9 cells^{-1} h^{-1}	β	0,16	g g^{-1} h^{-1}
k_{cX}	29,99	e^9cells h L^{-1}	k_{cX}	17	g h L^{-1}
k_{age}	102,05	H	k_{age}	0	h

Iš viso 10 CHO žinduolių ląstelių eksperimentų duomenys buvo išanalizuoti ir ištestuoti su šio straipsnio modeliu. Su pateiktu CHO ląstelių parametrų rinkiniu (5.3 lent.) gyvybingų ląstelių netiesioginis įvertinimo metodas pateikė priimtinus rezultatus. Vidutiniai MAE, RMSE ir MAPE rezultatai buvo atitinkamai 0,139 e^9cells L^{-1}, 0,158 e^9cells L^{-1} ir 5,15 %. Figure 2.11 grafiškai atvaizduoja pirmų šešių testų rezultatus.

Taip pat šio straipsnio prognozavimo algoritmas buvo ištestuotas su 12 skirtingais *E. coli* kultivavimo proceso duomenimis, optimalus parametrų rinkinys pateiktas 5.3 lent. Atlikus testus su modeliu gauti rezultatai: 1,78 g L^{-1} MAE, 2,53 g L^{-1} RMSE ir 6,97 % MAPE. Pirmų šešių testų rezultatai atvaizduoti Figure 2.12.

5.4. Išvados

1. Šioje disertacijoje siūlomas biomasės įvertinimo metodas, pagrįstas stechiometrija ir deguonies suvartojimu. Biomasės koncentracijos nustatymo algoritmas tinkamas naudoti su plataus spektro ląstelių kultūromis: mielėmis, trimis skirtingomis bakterijų padermėmis ir žinduolių ląstelėmis. Šios kultūros

yra pagrindinės pramonėje naudojamos ląstelės. Todėl aprašytas algoritmas turi didelį rinkodaros potencialą, kuris išspręstų mėginių ėmimo problemą ir padidintų žmogaus darbo efektyvumą. Metodo gautas tikslumas su mielėms ir bakterijoms yra 6,97%, vidutinė procentinė paklaida (MAPE), o su žinduolinėmis ląstelėmis gauta vidutinė procentinė paklaida (MAPE) yra 5,15%.

2. Šioje disertacijoje siūlomas ir taikomas tikslinių baltymų modeliavimo metodas, pagrįstas stechiometrija ir deguonies suvartojimu. Produkto sintezės modelis parodė, kad tikslinė produkto koncentracija priklauso nuo santykinio augimo greičio indukcijos metu ir nuo biomasės prieaugio po indukcijos. Šios įrodytos priklausomybės gali būti panaudotos kaip optimizavimo kriterijai, siekiant gauti reikiamas sąlygas kultivavimo procese, norint užtikrinti efektyvų baltymų sintezavimą. Metodo modelio pritaikymo vidutinis MAE ir MAPE tikslumas yra atitinkamai 0,099 g L^{-1} ir 8,22 %.
3. Šioje disertacijoje siūlomas ir pritaikytas santykinis ląstelių augimo greičio įvertinimo metodas, pagrįstas stechiometrija ir deguonies suvartojimu. Aprašytas metodas turi vieną derinimo parametrą, kuris skiriasi priklausomai nuo ląstelės kultūros, o tai lemia jo stabilumą ir paprastumą. Šis metodas puikiai derinamas su valdymo algoritmais, nes suteikia grįžtamąją informaciją apie auginimo procesą. Santykinio augimo greičio įvertinimo tikslumas yra vidutiniškai (MAPE) 9,77 %.
4. Šioje disertacijoje buvo parodyta, kad stechiometrijos koeficientai negali būti laikomi stacionariais įvairiomis sąlygomis. Ląstelių gyvybinių funkcijų palaikymo nario priklausomybė nuo produkto sintezės yra reikšminga. Žinduolių kultivavimo procesuose ląstelių kultūros amžius dalyvauja išraiškose, apibūdinančiose ląstelių augimą ir gyvybinių funkcijų palaikymą. Ląstelių kultūros amžius yra labai svarbus kintamasis kultivavimo procesams, kurie yra nepertraukiami arba ištęsti. Siūloma naujovė leidžia šios disertacijos algoritmus panaudoti įvairaus tipo kultivavimo procesuose: periodiniuose, periodiniuose su pamaitinimu, tęstiniuose. Be to, algoritmai gali būti naudojami bioprocesuose su induktoriumi (ITPG).

6. REFERENCES

1. C. Liao, S. Xiao, and X. Wang, “Bench-to-bedside: Translational development landscape of biotechnology in healthcare,” *Health Sciences Review*, vol. 7, p. 100097, 2023.
2. P. D. Costa *et al.*, “Biomedical and health informatics teaching in Portugal: Current status,” *Heliyon*, vol. 9, no. 3, p. 14163, 2023.
3. E. Woźniak and A. Tyczewska, “Bioeconomy during the COVID-19 and perspectives for the post-pandemic world: Example from EU,” *EFB Bioeconomy Journal*, vol. 1, p. 100013, 2021.
4. Y. Hasija, “Drug informatics,” in *All About Bioinformatics*, Elsevier, 2023, pp. 179–201.
5. R. Renneberg, “Myocardial Infarction, Cancer, and Stem Cells,” in *Biotechnology for Beginners*, Elsevier, pp. 325–360, 2023.
6. A. Survyla, B. Kemesis, and R. Urniežius, “Entropy Measure for Planning, Prediction and Online Estimation in Biotechnological Brocesses,” in *Proceedings of Entropy 2021: The Scientific Tool of the 21st Century*, p. 9839, 2021.
7. J. V. Reddy, K. Raudenbush, E. T. Papoutsakis, and M. Ierapetritou, “Cell-culture process optimization via model-based predictions of metabolism and protein glycosylation,” *Biotechnology Advances*, vol. 67, p. 108179, 2023.
8. J. Jones *et al.*, “Improved control strategies for the environment within cell culture bioreactors,” *Food and Bioproducts Processing*, vol. 138, pp. 209–220, 2023.
9. M. Takahashi, T. Kato, and H. Aoyagi, “Development of a feed-forward control system for medium in shake-flask culture,” *Biochemical Engineering Journal*, vol. 196, p. 108939, 2023.
10. R. Urniezius, V. Galvanauskas, A. Survyla, R. Simutis, and D. Levisauskas, “From Physics to Bioengineering: Microbial Cultivation Process Design and Feeding Rate Control Based on Relative Entropy Using Nuisance Time,” *Entropy*, vol. 20, no. 10, p. 779, 2018.
11. M. A. Jouned, J. Kager, C. Herwig, and T. Barz, “Event driven modeling for the accurate identification of metabolic switches in fed-batch culture of S. cerevisiae,” *Biochemical Engineering Journal*, vol. 180, p. 108345, 2022.
12. M. Matukaitis, D. Masaitis, R. Urniežius, L. Zlatkus, and V. Vaitkus, “Non-Invasive Estimation of Acetates Using Off-Gas Information for Fed-Batch E. coli Bioprocess,” in *ECP 2022*, MDPI, p. 5, 2022.
13. J. Kastenhofer, V. Rajamanickam, J. Libiseller-Egger, and O. Spadiut, “Monitoring and control of E. coli cell integrity,” *Journal of Biotechnology*, vol. 329, pp. 1–12, 2021.

14. V. Galvanauskas, R. Simutis, D. Levisauskas, R. Urniezius, T. Tekorius, and V. Vaitkus, “Design and Adaptation of Robust Feeding Profiles for Recombinant E.coli Fed-batch Cultivation Processes,” in *2021 29th Mediterranean Conference on Control and Automation (MED)*, PUGLIA, Italy: IEEE, pp. 532–537, 2021.

15. P. Unrean, “Bioprocess modelling for the design and optimization of lignocellulosic biomass fermentation,” *Bioresour. Bioprocess.*, vol. 3, no. 1, p. 1, 2016.

16. R. Luedeking and E. L. Piret, “A kinetic study of the lactic acid fermentation. Batch process at controlled pH,” *Biotechnol. Bioeng.*, vol. 1, no. 4, pp. 393–412, 1959.

17. R. Urniezius, B. Kemesis, and R. Simutis, “Bridging Offline Functional Model Carrying Aging-Specific Growth Rate Information and Recombinant Protein Expression: Entropic Extension of Akaike Information Criterion,” *Entropy*, vol. 23, no. 8, p. 1057, 2021.

18. Galvanauskas, Simutis, Levišauskas, and Urniežius, “Practical Solutions for Specific Growth Rate Control Systems in Industrial Bioreactors,” *Processes*, vol. 7, no. 10, p. 693, 2019.

19. Y. Luo, V. Kurian, and B. A. Ogunnaike, “Bioprocess systems analysis, modeling, estimation, and control,” *Current Opinion in Chemical Engineering*, vol. 33, p. 100705, 2021.

20. V. Galvanauskas, R. Simutis, D. Levisauskas, and R. Urniezius, “Combined Scheme for Basic Control Systems in Industrial Bioreactors,” in *15th European Workshop on Advanced Control and Diagnosis (ACD 2019)*, E. Zattoni, S. Simani, and G. Conte, Eds., in Lecture Notes in Control and Information Sciences – Proceedings, Cham: Springer International Publishing, pp. 1055–1069, 2022.

21. Y. Brignoli, B. Freeland, D. Cunningham, and M. Dabros, “Control of Specific Growth Rate in Fed-Batch Bioprocesses: Novel Controller Design for Improved Noise Management,” *Processes*, vol. 8, no. 6, p. 679, 2020.

22. P. Hrnciri, “Software Sensors for Monitoring of Biopolymer Production,” in *2021 23rd International Conference on Process Control (PC)*, Strbske Pleso, Slovakia: IEEE, pp. 308–312, 2021.

23. M. Abadli, L. Dewasme, D. Dumur, S. Tebbani, and A. V. Wouwer, “Generic model control of an Escherichia coli fed-batch culture,” in *2019 23rd International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, Romania: IEEE, pp. 212–217, 2019.

24. S. Chakraborty, S. K. Govindarajan, and S. N. Gummadi, “Influence of crucial reservoir properties and microbial kinetic parameters on enhanced oil recovery by microbial flooding under nonisothermal conditions: Mathematical modelling

and numerical simulation," *Journal of Petroleum Science and Engineering*, vol. 195, p. 107831, 2020.

25. S. Vaz, "Applied biotechnology for biomass," in *Renewable Carbon*, Elsevier, pp. 117–127, 2022.
26. C. L. Gargalo *et al.*, "Towards smart biomanufacturing: a perspective on recent developments in industrial measurement and monitoring technologies for bio-based production processes," *Journal of Industrial Microbiology and Biotechnology*, vol. 47, no. 11, pp. 947–964, 2020.
27. Y. K. Lin, H. Y. Leong, T. C. Ling, D.-Q. Lin, and S.-J. Yao, "Raman spectroscopy as process analytical tool in downstream processing of biotechnology," *Chinese Journal of Chemical Engineering*, vol. 30, pp. 204–211, 2021.
28. S. Goldrick *et al.*, "High-Throughput Raman Spectroscopy Combined with Innovate Data Analysis Workflow to Enhance Biopharmaceutical Process Development," *Processes*, vol. 8, no. 9, p. 1179, Sep. 2020.
29. S. Pinhal, D. Ropers, J. Geiselmann, and H. de Jong, "Acetate Metabolism and the Inhibition of Bacterial Growth by Acetate," *J Bacteriol*, vol. 201, no. 13, 2019.
30. D. F. Müller, D. Wibbing, C. Herwig, and J. Kager, "Simultaneous real-time estimation of maximum substrate uptake capacity and yield coefficient in induced microbial cultures," *Computers & Chemical Engineering*, vol. 173, p. 108203, 2023.
31. F. A. Ortega-Quintana, M. A. Trujillo-Roldán, H. Botero-Castro, and H. Alvarez, "Modeling the interaction between the central carbon metabolism of Escherichia coli and bioreactor culture media," *Biochemical Engineering Journal*, vol. 163, p. 107753, 2020.
32. A. Tuveri, F. Pérez-García, P. A. Lira-Parada, L. Imsland, and N. Bar, "Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess monitoring," *Journal of Process Control*, vol. 106, pp. 195–207, 2021.
33. P. Noll and M. Henkel, "History and Evolution of Modeling in Biotechnology: Modeling & Simulation, Application and Hardware Performance," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 3309–3323, 2020.
34. A. Beiroti, M. R. Aghasadeghi, S. N. Hosseini, and D. Norouzian, "Application of recurrent neural network for online prediction of cell density of recombinant *Pichia pastoris* producing HBsAg," *Preparative Biochemistry and Biotechnology*, vol. 49, no. 4, pp. 352–359, 2019.
35. M. R. Mowbray, C. Wu, A. W. Rogers, E. A. D. Rio-Chanona, and D. Zhang, "A reinforcement learning-based hybrid modeling framework for bioprocess

kinetics identification," *Biotech & Bioengineering*, vol. 120, no. 1, pp. 154–168, 2023.

36. B. Mahanty, "Hybrid modeling in bioprocess dynamics: Structural variabilities, implementation strategies, and practical challenges," *Biotech & Bioengineering*, vol. 120, no. 8, pp. 2072–2091, 2023.
37. H.-L. Lau, F. W. F. Wong, R. N. Z. R. A. Rahman, M. S. Mohamed, A. B. Ariff, and S.-L. Hii, "Optimization of fermentation medium components by response surface methodology (RSM) and artificial neural network hybrid with genetic algorithm (ANN-GA) for lipase production by Burkholderia cenocepacia ST8 using used automotive engine oil as substrate," *Biocatalysis and Agricultural Biotechnology*, vol. 50, p. 102696, 2023.
38. B. Bayer, B. Sissolak, M. Duerkop, M. von Stosch, and G. Striedner, "The shortcomings of accurate rate estimations in cultivation processes and a solution for precise and robust process modeling," *Bioprocess Biosyst Eng*, vol. 43, no. 2, pp. 169–178, 2020.
39. P. Lhamo and B. Mahanty, "Structural variability, implementational irregularities in mathematical modelling of polyhydroxyalkanoates (PHAs) production—A state-of-the-art review," *Biotech & Bioengineering*, vol. 119, no. 11, pp. 3079–3095, 2022.
40. R. Moscoviz and J. Jimenez, "Improving anaerobic digestion mass balance calculations through stoichiometry and usual substrate characterization," *Bioresource Technology*, vol. 337, p. 125402, 2021.
41. A. Hatzikioseyian and M. Tsezos, "Modelling of microbial metabolism stoichiometry: Application in bioleaching processes," *Hydrometallurgy*, vol. 83, no. 1–4, pp. 29–34, 2006.
42. R. Urniezius, A. Survyla, D. Paulauskas, V. A. Bumelis, and V. Galvanauskas, "Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate," *Microb Cell Fact*, vol. 18, no. 1, p. 190, 2019.
43. R. Urniezius and A. Survyla, "Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process," *Entropy*, vol. 21, no. 12, p. 1221, 2019.
44. A. Survyla, D. Levisauskas, R. Urniezius, and R. Simutis, "An oxygen-uptake-rate-based estimator of the specific growth rate in Escherichia coli BL21 strains cultivation processes," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 5856–5863, 2021.
45. A. Survyla, R. Urniezius, and R. Simutis, "Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional," *Talanta*, vol. 254, p. 124121, 2023.

46. X. Li, “Bioengineering of FGFs and New Drug Developments,” in *Fibroblast Growth Factors*, Elsevier, pp. 477–558, 2018.

47. B. Kuang, D. Hoang, Z. Wang, and S. Yoon, “Cell Metabolic Diagnosis and Control in CHO Fed-batch Process,” *IFAC-PapersOnLine*, vol. 55, no. 7, pp. 37–44, 2022.

48. S. Schaepe, A. Kuprijanov, R. Simutis, and A. Lübbert, “Avoiding overfeeding in high cell density fed-batch cultures of E. coli during the production of heterologous proteins,” *Journal of Biotechnology*, vol. 192, pp. 146–153, 2014.

49. J. Schubert, R. Simutis, M. Dors, I. Havlik, and A. Lübbert, “Bioprocess optimization and control: Application of hybrid modelling,” *Journal of Biotechnology*, vol. 35, no. 1, pp. 51–68, 1994.

50. S. B. Petkov and R. A. Davis, “On-line biomass estimation using a modified oxygen utilization rate,” *Bioprocess Engineering*, vol. 15, no. 1, pp. 43–45, 1996.

51. J. M. Barrigón, R. Ramon, I. Rocha, F. Valero, E. C. Ferreira, and J. L. Montesinos, “State and specific growth estimation in heterologous protein production by *Pichia pastoris*,” *AIChE Journal*, vol. 58, no. 10, pp. 2966–2979, 2012.

52. K. Kovárová-Kovar and T. Egli, “Growth Kinetics of Suspended Microbial Cells: From Single-Substrate-Controlled Growth to Mixed-Substrate Kinetics,” *Microbiol Mol Biol Rev*, vol. 62, no. 3, pp. 646–666, 1998.

53. M. Dabros, M. M. Schuler, and I. W. Marison, “Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass,” *Bioprocess Biosyst Eng*, vol. 33, no. 9, pp. 1109–1118, 2010.

54. S. Gnoth, M. Jenzsch, R. Simutis, and A. Lübbert, “Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control,” *Journal of Biotechnology*, vol. 132, no. 2, pp. 180–186, 2007.

55. P. Linko and Y. Zhu, “Neural network programming in bioprocess variable estimation and state prediction,” *Journal of Biotechnology*, vol. 21, no. 3, pp. 253–269, 1991.

56. S. Schaepe, A. Kuprijanov, C. Sieblist, M. Jenzsch, R. Simutis, and A. Lübbert, “Current Advances in Tools Improving Bioreactor Performance,” *CBIOT*, vol. 3, no. 2, pp. 133–144, 2013.

57. F. Garcia-Ochoa, E. Gomez, V. E. Santos, and J. C. Merchuk, “Oxygen uptake rate in microbial processes: An overview,” *Biochemical Engineering Journal*, vol. 49, no. 3, pp. 289–307, 2010.

58. A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, and Q. Li, “Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*,” *Protein Science*, vol. 18, no. 5, pp. 936–948, 2009.

59. C. Willmott and K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” *Clim. Res.*, vol. 30, pp. 79–82, 2005.

60. A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, “Mean Absolute Percentage Error for regression models,” *Neurocomputing*, vol. 192, pp. 38–48, 2016.

61. V. Galvanauskas, N. Volk, R. Simutis, and A. Lübbert, “DESIGN OF RECOMBINANT PROTEIN PRODUCTION PROCESSES,” *Chemical Engineering Communications*, vol. 191, no. 5, pp. 732–748, 2004.

62. S. Gnoth, R. Simutis, and A. Lübbert, “Selective expression of the soluble product fraction in Escherichia coli cultures employed in recombinant protein production processes,” *Appl Microbiol Biotechnol*, vol. 87, no. 6, pp. 2047–2058, 2010.

63. M. Jenzsch, R. Simutis, and A. Luebbert, “Generic model control of the specific growth rate in recombinant Escherichia coli cultivations,” *Journal of Biotechnology*, vol. 122, no. 4, pp. 483–493, Apr. 2006.

64. A. Zymnis, S. Boyd, and D. Gorinevsky, “Mixed linear system estimation and identification,” *Signal Processing*, vol. 90, no. 3, pp. 966–971, 2010.

65. V. Babaeipour, S. A. Shojaosadati, and N. Maghsoudi, “Maximizing Production of Human Interferon-γ in HCDC of Recombinant E. coli,” *Iran J Pharm Res*, vol. 12, no. 3, pp. 563–572, 2013.

66. M. Jenzsch, S. Gnoth, M. Kleinschmidt, R. Simutis, and A. Lübbert, “Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production,” *Journal of Biotechnology*, vol. 128, no. 4, pp. 858–867, 2007.

67. E. W. Swokowski, *Calculus with analytic geometry*, 2d ed. Boston: Prindle, Weber & Schmidt, 1979.

68. D. Levisauskas, V. Galvanauskas, S. Henrich, K. Wilhelm, N. Volk, and A. Lübbert, “Model-based optimization of viral capsid protein production in fed-batch culture of recombinant Escherichia coli,” *Bioprocess Biosyst Eng*, vol. 25, no. 4, pp. 255–262, 2003.

69. F. Miao and D. S. Kompala, “Overexpression of cloned genes using recombinant *Escherichia coli* regulated by a T7 promoter: I. Batch cultures and kinetic modeling,” *Biotech & Bioengineering*, vol. 40, no. 7, pp. 787–796, Oct. 1992.

70. 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.

71. S. Gnoth, M. Jenzsch, R. Simutis, and A. Lübbert, “Product formation kinetics in genetically modified E. coli bacteria: inclusion body formation,” *Bioprocess Biosyst Eng*, vol. 31, no. 1, pp. 41–46, 2008.

72. R. Luedeking and E. L. Piret, “Transient and steady states in continuous fermentaion. Theory and experiment,” *Biotechnol. Bioeng.*, vol. 1, no. 4, pp. 431–459, 1959.

73. M. Caramihai and I. Severi, “Bioprocess Modeling and Control,” in *Biomass Now - Sustainable Growth and Use*, M. D. Matovic, Ed., InTech, 2013.

74. M. Pappenreiter, B. Sissolak, W. Sommeregger, and G. Striedner, “Oxygen Uptake Rate Soft-Sensing via Dynamic kLa Computation: Cell Volume and Metabolic Transition Prediction in Mammalian Bioprocesses,” *Front. Bioeng. Biotechnol.*, vol. 7, p. 195, Aug. 2019.

75. D. Levisauskas, “Inferential control of the specific growth rate in fed-batch cultivation processes,” *Biotechnology Letters*, vol. 23, no. 15, pp. 1189–1195, 2001.

76. K. Carr-Brion, Ed., *Measurement and control in bioprocessing*. in Elsevier applied biotechnology series. London ; New York: Elsevier Applied Science, 1991.

77. K. Kovárová-Kovar and T. Egli, “Growth Kinetics of Suspended Microbial Cells: From Single-Substrate-Controlled Growth to Mixed-Substrate Kinetics,” *Microbiol Mol Biol Rev*, vol. 62, no. 3, pp. 646–666, 1998.

78. J. Monod, “The Growth of Bacterial Cultures,” *Annu. Rev. Microbiol.*, vol. 3, no. 1, pp. 371–394, 1949.

79. M. Aehle, A. Kuprijanov, S. Schaepe, R. Simutis, and A. Lübbert, “Increasing batch-to-batch reproducibility of CHO cultures by robust open-loop control,” *Cytotechnology*, vol. 63, no. 1, pp. 41–47, 2011.

80. R. Simutis and A. Lübbert, “Bioreactor control improves bioprocess performance,” *Biotechnology Journal*, vol. 10, no. 8, pp. 1115–1130, 2015.

81. *Differential Equations with Applications and Historical Notes, Third Edition*. Chapman and Hall/CRC, 2016.

82. B. Kemesis, R. Urniezius, T. Kondratas, L. Jankauskaite, D. Masaitis, and P. Babilius, “Bridging Functional Model of Arterial Oxygen with Information of Venous Blood Gas: Validating Bioprocess Soft Sensor on Human Respiration,” in *Intelligent and Safe Computer Systems in Control and Diagnostics*, vol. 545, Z. Kowalczuk, Ed., in Lecture Notes in Networks and Systems, vol. 545. , Cham: Springer International Publishing, pp. 42–51, 2023.

83. Survyla Arnas, Urniezius Renaldas, Kemesis Benas, Zlatkus Lukas, Masaitis Deividas, and Galvanauskas Vytautas, “Modeling the Specific Glucose Consumption Rate for the Recombinant E.coli Bioprocesses Based on Aging-

specific Growth Rate," *Chemical Engineering Transactions*, vol. 93, pp. 265–270, 2022.

84. M. Sbarciog, D. Coutinho, and A. Vande Wouwer, "A simple output-feedback strategy for the control of perfused mammalian cell cultures," *Control Engineering Practice*, vol. 32, pp. 123–135, 2014.

85. V. Galvanauskas, R. Simutis, S. C. Nath, and M. Kino-oka, "Kinetic modeling of human induced pluripotent stem cell expansion in suspension culture," *Regenerative Therapy*, vol. 12, pp. 88–93, 2019.

86. S. Haykin, *Neural networks: a comprehensive foundation*, 2nd ed. Delhi: Pearson Education, 1999.

87. P. Ducommun, I. Bolzonella, M. Rhiel, P. Pugeaud, U. von Stockar, and I. W. Marison, "On-line determination of animal cell concentration," *Biotechnol. Bioeng.*, vol. 72, no. 5, pp. 515–522, 2001.

88. M. M. Schuler and I. W. Marison, "Real-time monitoring and control of microbial bioprocesses with focus on the specific growth rate: current state and perspectives," *Appl Microbiol Biotechnol*, vol. 94, no. 6, pp. 1469–1482, 2012.

89. S. Areeya, A. Tawai, and M. Sriariyanun, "Model-Based Control for Fed-Batch Enzymatic Hydrolysis Reactor of Lignocellulosic Biomass," in *2020 6th International Conference on Control, Automation and Robotics (ICCAR)*, Singapore: IEEE, pp. 536–540, 2020.

90. V. Galvanauskas, R. Simutis, D. Levisauskas, M. Butkus, and R. Urniezius, "Development and Investigation of Efficient Substrate Feeding and Dissolved Oxygen Control Algorithms for Design of Recombinant E. coli Cultivation Process," in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, Munich, Germany: IEEE, pp. 657–660, 2018.

91. V. Rajinikanth and K. Latha, "Modeling, Analysis, and Intelligent Controller Tuning for a Bioreactor: A Simulation Study," *ISRN Chemical Engineering*, vol. 2012, pp. 1–15, 2012.

92. S. J. Yoo, J. H. Kim, and J. M. Lee, "Soft sensor design with state estimator for lipid estimation of microalgal photobioreactor system," in *2014 14th International Conference on Control, Automation and Systems (ICCAS 2014)*, Gyeonggi-do, South Korea: IEEE, pp. 1299–1304, 2014.

93. G. C. Goodwin, "Predicting the performance of soft sensors as a route to low cost automation," *Annual Reviews in Control*, vol. 24, pp. 55–66, 2000.

94. A. Seeger, B. Schneppe, J. E. G. McCarthy, W.-D. Deckwer, and U. Rinas, "Comparison of temperature- and isopropyl-β- d -thiogalacto-pyranoside-induced synthesis of basic fibroblast growth factor in high-cell-density cultures of recombinant Escherichia coli," *Enzyme and Microbial Technology*, vol. 17, no. 10, pp. 947–953, 1995.

7. CURRICULUM VITAE

Personal information
Arnas Survyla
Date of birth: 10 October 1994
Email address: arnas.survyla@ktu.lt

Institution:
Kaunas University of Technology
Faculty of Electrical and Electronics
Department of Automation

Education:

2019 – 2024	PhD studies of Informatics Engineering at Kaunas University of Technology, Faculty of Electrical and Electronics Engineering, Department of Automation, Kaunas, Lithuania.
2017 – 2019	Master's degree of Engineering sciences at Kaunas University of Technology, Faculty of Electrical and Electronics Engineering, Department of Automation, Kaunas, Lithuania.
2013 – 2017	Bachelor's degree of Control Systems at Kaunas University of Technology, Faculty of Electrical and Electronics Engineering, Department of Automation, Kaunas, Lithuania.

Research interests:
Biotechnology,
Process optimization and control,
Development of soft-sensors, mathematical models,
Software engineering.

Work experience:

2021 – 2023	Engineer, software developer at *Cumulatis*
2020 – 2023	Junior researcher at the national LMT project "Development and application of soft sensor system for state estimation of biotechnological processes"
2021 – 2022	Junior researcher at the national MITA project "Spiral freezer with intelligence function and development of new refrigeration recipes"
2020 – 2021	Engineer, software developer at the national project "Prototype of a non-invasive monitoring device for a COVID-19 patient"
2020 – 2021	Junior researcher, software developer at the national MITA project "Multilevel Artificial Intelligent Based Prototype Inverter of Wind-Solar Hybrid System"
2018 – 2020	Junior researcher, analyst at the national LVPA project "Intelligent control system platform for recombinant protein synthesis processes"

2018 – 2020	Laboratory technician at the national LMT project "Development and Investigation of Adaptive Control Algorithms and Systems for Biotechnological Processes"
2017 – 2018	Junior biotechnologist, automation operator at *Biotechpharma*, IMC.

8. LIST OF ARTICLES AND SCIENTIFIC CONFERENCES

Scopus and *Web of Science* database journals with the *Impact Factor*:

1. **A. Survyla**, R. Urniezius, and R. Simutis, "Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional," *Talanta*, vol. 254, p. 124121, 2023. Author's contribution: 0.334.
2. **A. Survyla**, D. Levisauskas, R. Urniezius, and R. Simutis, "An oxygen-uptake-rate-based estimator of the specific growth rate in Escherichia coli BL21 strains cultivation processes," *Computational and Structural Biotechnology Journal*, vol. 19, p. 5856–5863, 2021. Author's contribution: 0.250.
3. R. Urniezius and **A. Survyla**, "Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process," *Entropy*, vol. 21, No. 12, p. 1221, 2019. Author's contribution: 0.500.
4. R. Urniezius, **A. Survyla**, D. Paulauskas, V. A. Bumelis, and V. Galvanauskas, "Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate," *Microb Cell Fact*, vol. 18, No. 1, p. 190, 2019. Author's contribution: 0.200.

International conferences:

1. **A. Survyla,** R. Urniezius, V. Vaitkus, D. Levisauskas, L. Jankauskaite, D. Lukminaite, G. Laucaityte, "Noninvasive Continuous Tracking of Partial Pressure of Oxygen in Arterial Blood: Adapting Microorganisms Bioprocess Soft Sensor Technology for Holistic Analysis of Human Respiratory System," in *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Karlsruhe, Germany: IEEE, p. 1–5, 2021. Author's contribution: 0.148.
2. **Survyla Arnas**, Urniezius Renaldas, Kemesis Benas, Zlatkus Lukas, Masaitis Deividas, and Galvanauskas Vytautas, "Modeling the Specific Glucose Consumption Rate for the Recombinant E. coli Bioprocesses Based on Aging-specific Growth Rate," *Chemical Engineering Transactions*, vol. 93, p. 265–270, 2022. Author's contribution: 0.170.
3. **A. Survyla**, B. Kemesis, and R. Urniežius, "Entropy Measure for Planning, Prediction and Online Estimation in Biotechnological Brocesses," in *Proceedings of Entropy 2021: The Scientific Tool of the 21st Century*, Sciforum.net: MDPI, 2021, p. 9839. Author's contribution: 0.333.

Patents registered in the State Patent Office of the Republic of Lithuania and patent applications registered in other foreign patent offices:

1. Galvanauskas, Vytautas (author of invention); Simutis, Rimvydas (author of invention); Levišauskas, Donatas (author of invention); Urniežius, Renaldas (author of invention); Vaitkus, Vygandas (author of invention); Tekorius, Tomas

(author of invention); Butkus, Mantas (author of invention); **Survyla, Arnas** (author of invention). Maitinimo profilių, skirtų valdyti pusiau-periodinius rekombinantinių E. coli kultivavimo procesus, formavimo ir adaptavimo būdas / inventors: V. Galvanauskas, R. Simutis, D. Levišauskas, R. Urniežius, V. Vaitkus, T. Tekorius, M. Butkus, **A. Survyla**; owner: Kaunas University of Technology. LT 6861 B. 10 Nov 2021. Author's contribution: 0.125.

9. COPIES OF PUBLISHED PAPERS

Urniezius *et al. Microb Cell Fact* (2019) 18:190
https://doi.org/10.1186/s12934-019-1241-7

Microbial Cell Factories

RESEARCH Open Access

Generic estimator of biomass concentration for *Escherichia coli* and *Saccharomyces cerevisiae* fed-batch cultures based on cumulative oxygen consumption rate

Renaldas Urniezius[1*], Arnas Survyla[1], Dziugas Paulauskas[2], Vladas Algirdas Bumelis[2] and Vytautas Galvanauskas[1]

Abstract

Background: The focus of this study is online estimation of biomass concentration in fed-batch cultures. It describes a bioengineering software solution, which is explored for *Escherichia coli* and *Saccharomyces cerevisiae* fed-batch cultures. The experimental investigation of both cultures presents experimental validation results since the start of the bioprocess, i.e. since the injection of inoculant solution into bioreactor. In total, four strains were analyzed, and 21 experiments were performed under varying bioprocess conditions, out of which 7 experiments were carried out with dosed substrate feeding. Development of the microorganisms' culture invariant generic estimator of biomass concentration was the main goal of this research.

Results: The results show that stoichiometric parameters provide acceptable knowledge on the state of biomass concentrations during the whole cultivation process, including the exponential growth phase of both *E. coli* and *S. cerevisiae* cultures. The cell culture stoichiometric parameters are estimated by a procedure based on the Luedeking/Piret-model and maximization of entropy. The main input signal of the approach is cumulative oxygen uptake rate at fed-batch cultivation processes. The developed noninvasive biomass estimation procedure was intentionally made to not depend on the selection of corresponding bioprocess/bioreactor parameters.

Conclusions: The precision errors, since the bioprocess start, when inoculant was injected to a bioreactor, confirmed that the approach is relevant for online biomass state estimation. This included the lag and exponential growth phases for both *E. coli* and *S. cerevisiae*. The suggested estimation procedure is identical for both cultures. This approach improves the precision achieved by other authors without compromising the simplicity of the implementation. Moreover, the suggested approach is a candidate method to be the microorganisms' culture invariant approach. It does not depend on any numeric initial optimization conditions, it does not require any of bioreactor parameters. No numeric stability issues of convergence occurred during multiple performance tests. All this makes this approach a potential candidate for industrial tasks with adaptive feeding control or automatic inoculations when substrate feeding profile and bioreactor parameters are not provided.

Keywords: Biomass estimator, Stoichiometry, Relative entropy, Microbial cultivation

*Correspondence: renaldas.urniezius@ktu.lt
[1] Department of Automation, Kaunas University of Technology, 51367 Kaunas, Lithuania
Full list of author information is available at the end of the article

Background

Biotechnology industry development over the last years made quality assurance more stringent for pharmacy production [1]. As a tool to resolve process data distortion and prevent operator from accidently making mistakes, bioengineering solutions help to automate tasks, which results in rise of cultivation process performance and quality. To strengthen product quality, to more efficiently acquire coefficient values, to improve safety and flexibility of adaptive feedback control, the soft/noninvasive sensors [2] become a rational choice for development of sustainable engineering solutions. Implementation of feedback control system requires a feedback signal from soft sensors or estimators that provide parameters [3], which are unavailable to be directly measured online [4]. The control algorithm and the feedback signal consider the product and the main characteristics of bioprocess parameters—the biomass concentration and the specific growth rate [5, 6].

This study delves into biomass estimator development based on stoichiometric parameters and Luedeking–Piret model. The cell's yields and stoichiometry both form a generic information, which is an acceptable candidate to be included in estimators when the microorganisms culture does not change from experiment to experiment. Depending on stoichiometry, the estimator of biomass concentration can be used to automatically inject the inoculant solution at a predefined level of the optical density in bioreactor medium. At this point, cumulative oxygen uptake rate signal from an off-gas analyzer is informative to determine the biomass concentration.

The biomass estimator described in this study includes optimization algorithm, which returns the stoichiometric parameters of the controlled culture. The algorithm refers to several optimization criteria and is based on a gray box model originating from Luedeking–Piret model. Then offline maximization of entropy leads to satisfactory parameters values for estimation procedure, which is then applied to *Escherichia coli* bacteria and *S. cerevisiae* yeast cultures. In other words, the stoichiometry optimization algorithm must be performed once for each strain to determine the necessary coefficients. These coefficients can be later used in the subsequent experiments to estimate biomass concentration online, unless the strain does not change. Such offline analysis can be considered as an estimator tuning algorithm for a specific microorganisms' culture.

The "Materials and methods" section describes the materials, strains and the bioreactor system operating conditions. The "Comparative analysis of biomass estimators" section reviews literature references of the off-gas analysis approaches and introduces the motivation for this study. The "General mathematical model of stoichiometric parameters estimation" section layouts the derivation of the bioengineering approach for both the offline (stoichiometry) analysis and the online (biomass concentration) analysis stages. It also resolves a general formulation of the oxygen consumption for biomass maintenance coefficient, which is relevant for both *E. coli* and *S. cerevisiae* cultures. The "Experimental validation" section provides experimental proof of the developed stoichiometry coefficients offline identification and the biomass concentration online estimation algorithms. The "Conclusions" section discusses the results and concludes the final statements of this study.

Materials and methods

Cell strain's

Four types of strain cultivation were analyzed in this work to verify biomass estimation. *S. cerevisiae* (no DY7221) strain was used as representative of yeasts cells. The recombinant strains *E. coli* BL21(DE3) pET9a-IdeS, *E. coli* BL21 (DE3) pET21-IFN-alfa-5 (cloning of fused gene into bacterial systems with strong bacteriophage T7 promoter, pET21a+plasmid) [7] and *E.coli* BL21(DE3) pLysS [8] were used in bacterial cultivations.

Medium and culture conditions

In order to check biomass estimator's reliability and accuracy, data were collected from different cell strains which have been cultivated in multiple different R&D laboratories, including the laboratory of bioprocessing modeling and management in Kaunas University of Technology. *Saccharomyces cerevisiae* (no DY7221) strain was cultivated in the standard nutrient medium (YPD) [9, 10], which contained 1% yeast extract, 2% Bacto peptone, and 0.1% glucose. The feed solution contained 600 g/kg glucose which increased the solution density to 1.21 g/l.

The medium temperature was maintained at 30 °C and it was monitored by using temperature sensor "Pt100", and pH was kept constant at 4.9 by addition of NaOH(aq) [11]. Dissolved oxygen tension DOT in the bioreactor was measured by oxygen electrode Mettler Toledo and controlled by shifting stirrer speed from 230 to 600 rpm. The DOT set point was chosen as 30% of air saturation. The air flow was kept around 4 l/min and measured by a mass air flow sensor. The off-gas from bioreactor was measured online by BlueSens gas analyzer (BCpreFerm, BlueSens, Herten, Germany), which has O_2, CO_2 and pressure sensors. The culture broth mass was measured online with balanced reactor vessel which contained load cell weight sensor. The initial substrate concentration in the bioreactor was equal to zero, $S=0$ g/kg. Hence, after inoculation the substrate solution feeding was started. The cultivation process was performed in 5 l bioreactor.

The cell strain of *E. coli* BL21 (DE3) pET21-IFN-alfa-5 was cultivated in 7 l bioreactor. Cultivation medium was based on minimal mineral medium, which was made of 46.55 g potassium dihydrogen phosphate, 14 g ammonium phosphate dibasic, 5.6 g citric acid monohydrate, 3 ml of concentrate antifoam, 35 g magnesium sulphate heptahydrate, 105 g D (+) glucose monohydrate. The initial volume of medium was 3.7 kg. At the cultivation process the environment parameters were kept constant. The temperature setpoint was 37 °C, DOT set at 20% of air saturation and pH kept at pH 6.8 by addition of NaOH(aq). The stirrer rpm range was from 800 to 1200 rpm, the air flow rage was from 1.75 to 3.75 l/min. In order to increase oxygen transfer rate during cultivation process, pure oxygen flow was provided to bioreactor at range from 0 to 7.5 l/min. The off-gas from bioreactor was measured online by BlueSens.

The other cell strain of *E. coli* BL21 (DE3) pET9a-IdeS was cultivated in 15 l bioreactor. Cultivation medium based as minimal mineral medium. At the cultivation process the environment parameters: temperature set point was 37 °C, DOT set at 30% of air saturation and pH kept at pH 6.98 by addition of NaOH(aq). The stirrer rpm range was from 300 to 750 rpm, the air flow range was from 0.3 to 15 l/min. During the cultivation process pure oxygen flow was provided to bioreactor at range from 0 to 7.5 l/min. The off-gas from bioreactor was measured online by BlueSens.

For diversity of validation, the fourth cell strain was *E. coli* (BL21(DE3) pLysS) [8]. The cultivation medium used as minimal mineral medium composed with $(NH_4)_2SO_4$, 2.46 g/l; NH_4Cl, 0.5 g/l; $NaH_2PO_4 \times H2O$, 3.6 g/l; Na_2SO_4, 2 g/l; K_2HPO_4, 14.6 g/l; $(NH_4)_2$-citrate, 1 g/l; 1 M $MgSO_4$ solution, 5 ml/l; trace elements solution, 2 ml/l; and no glucose. Initial masses of all cultures were 5 kg. The glucose solution and initial substrate concentration at the bioreactor used same as at cultivation with yeasts, pH kept constant at pH 7 and temperature was regulated to 30 °C. Dissolved oxygen tension DOT was measured by an amperometric oxygen electrode (Mettler–Toledo) and the DOT set point was 30% of the saturation. The size of bioreactor was 15 l working volume (Biostat C, Sartorius Stedim Biotech) and the stirrer speed varied from 100 to 1400 rpm.

Comparative analysis of biomass estimators

In order to adaptively control and monitor chemical or biotechnological process, it is mandatory to implement a data collection system that provides desired variables at real time with acceptable precision and performance. This requires corresponding equipment, which may be unaffordable, not implementable in system or the required instrument doesn't exist. Hence, the better alternative is to use soft or noninvasive sensors, which collect measurable variables and estimate unmeasurable parameters [2, 12]. Especially in biotechnology processes, there are complex relationships between process and variables, so the best way to infer online unmeasurable parameters is to use corresponding estimators [4].

Over time, the studies of both bioprocesses and industrial production perspectives have shown that a biomass estimator requires data, which is closely related to biomass growth rate and biomass concentration. It can be indirectly measured online, with well-established and validated devices and soft sensors [4, 13], which are still in development. Oxygen uptake rate (OUR) and carbon dioxide production rate (CPR) are directly related to biomass growth rate and biomass concentration [14, 15]. Oxygen uptake rate (OUR) and carbon dioxide production rate (CPR) data for estimator must be computed from online signals that are reliable and measured directly in bioreactor system. These signals are the concentration of O_2 and CO_2 in the off-gas [16]. The proposed noninvasive biomass concentration estimation procedure was intentionally made to not depend on the selection of bioprocess/bioreactor parameters. The approach is valid for aerobic cultures as long as it is possible to obtain the off-gas measurements of sufficient quality.

The main model, dedicated to biomass concentration estimation in this work, is a Luedeking–Piret model derived from the stoichiometric equations for oxygen consumption. It represents relationship between biomass X growth/maintenance and oxygen uptake rate in bioreactor [14, 15]:

$$OUR(t) = \alpha \cdot X'(t) + \beta \cdot X(t) \quad (1)$$

Stoichiometric coefficients α and β represent cell's metabolisms of oxygen consumption and correspond to the yield coefficients of these biochemical conversions. In Eq. (1) coefficient α means specific cell's oxygen consumption yield ($\alpha \equiv Y_{o_2/X}$) for growth and β is a model parameter termed as oxygen consumption for maintenance ($\beta \equiv m_{o_2/X}$) [17–20]. The generic structure of the Eq. (1) that describes the process does not include any strain specific information and there are no any initial conditions assumed for the values of both α and β.

Simutis and Lübbert (2006) improved a hybrid model estimator [21]. The main improvement of a dynamical mathematical model was a modification of mass balance equation to the new one, which was based on the oxygen uptake rate OUR, the carbon dioxide rate CPR and the base consumption rate BCR [22]. In order to further improve hybrid model's capacity, Kalman filter (EKF) was introduced to biomass estimations [23]. The new improved hybrid model produced better results and

accuracy, but general drawbacks remained, estimator's complexity, a lot of data required for artificial neural network training and biomass estimation offline with a large execution duration [22–24]. In 2010, Simutis and Lübbert improved biomass estimator with cumulative variables that made model more conventional. The estimator procedure was transformed to a simpler system.

When comparing stoichiometry biomass estimators' mathematical models to the hybrid model estimator approaches, the latter contains more main state variables: biomass (X), oxygen uptake rate (OUR), specific biomass growth rate (μ), broth weight (w), carbon dioxide production rate (CPR), base consumption and other model coefficients. Additionally, additional equations and a fuzzy expert system are required. The latter gives an input to the combination of a dynamical mathematical model (DMM) represented by a set of nonlinear ordinary differential equations with an artificial neural network (ANN) [24]. The main advantage of the stoichiometry biomass estimator, compared to hybrid model, is its simplicity and accuracy. As hybrid model consists of several modeling systems, a common problem of estimation arrives from artificial neural network (ANN) training [21, 23, 24]. Meanwhile, stoichiometry biomass estimator was based only on OUR and stoichiometric parameters α and β, which both were kept static for a particular cell strain. This led to ability to calculate biomass online [14, 22–24, 28]. A general comparison of different biomass estimators is presented in Fig. 1. This work's biomass estimation approach is depicted by Fig. 1d. The estimation methods, which are based on gas consumption stoichiometry, are shown in Fig. 1e, f. The main differences consist of the approach picked, its complexity and the number of input signals and prerequisite parameters or initial conditions required. The main purpose of this paper is to show that biomass estimation can be treated from the fundamental point of view based on the stoichiometry Eq. (1). The idea comes from entropic and Bayesian inference approaches involving integral optimizations [29, 30]. The focus lays on the implementation, which can be not only used in scientific R&D laboratories, but also on the industrial plants level.

This paper presents a generic biomass estimation routine that is suitable for determination of biomass state in high diversity of bioreactors (Fig. 2) with potentially wide variety of industrial microorganisms. Prior to biomass determination, it is necessary to identify cell strain's stoichiometry parameters α and β, which both describe oxygen consumption by a microbial culture. This is accomplished by offline analysis Fig. 3 (stage A).

Afterwards, industrial scale cultivation processes reuse information about strain information for corresponding biomass concentration estimation in online analysis (stage B), as shown in Fig. 3. In order to achieve better accuracy at strain stoichiometry analysis during upstream development, it is recommended to identify α and β parameters at the laboratory scale bioreactors, Fig. 3 (stage A). This way, strain stoichiometry analysis, based on "ground-truth" of stage A, is economically beneficial, and data from cultivation process consists of less disturbances in more flexible control environment.

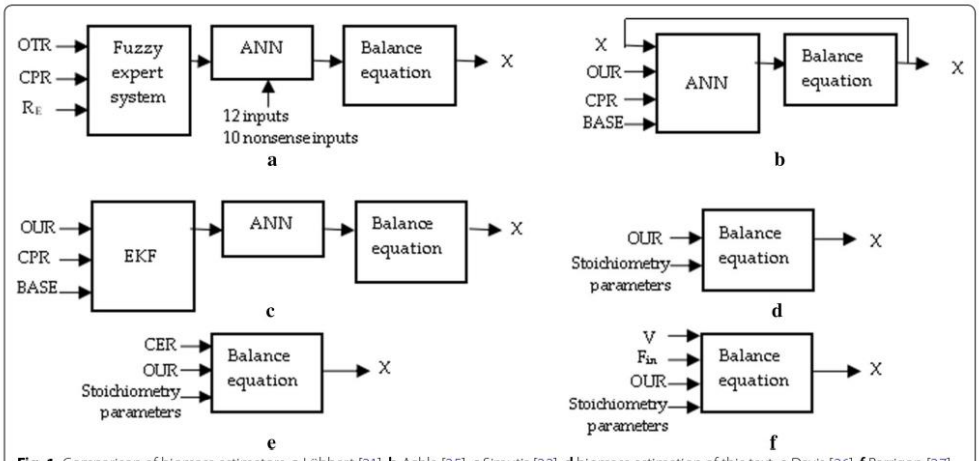


Fig. 1 Comparison of biomass estimators: **a** Lübbert [21], **b** Achle [25], **c** Simutis [22], **d** biomass estimation of this text, **e** Davis [26], **f** Barrigon [27]

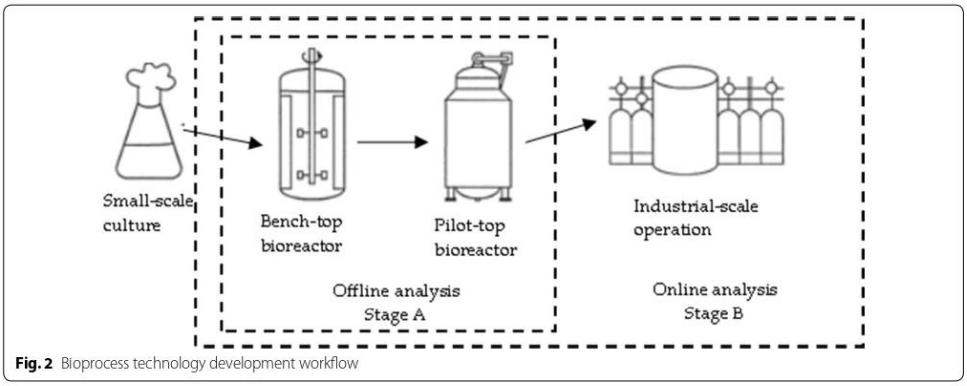


Fig. 2 Bioprocess technology development workflow

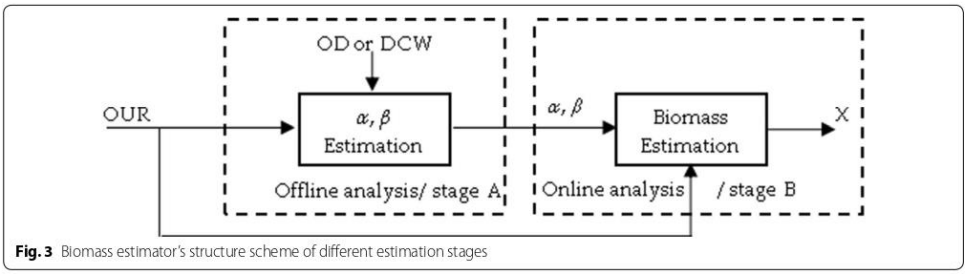


Fig. 3 Biomass estimator's structure scheme of different estimation stages

General mathematical model of stoichiometric parameters estimation

During the cultivation process, the real-time data collected from the devices has interference and disturbances, which may cause distortion of parameters and estimated values [14]. Simutis and Lübbert [4] stated "the reason for cumulating the original signals is to improve the signal-to-noise ratio (SNR) and thus increasing the information content about the process. Additionally, as the biomass and its metabolic products are accumulated during the cultivation, these masses are better correlated with the cumulative signals of OUR and CPR". The main method of the current text is also based on the integral approach, which can be considered as a filter eliminating noise [22]. Hence, the Luedeking–Piret model Eq. (1) outcomes are being protected from disturbances by integrating it:

$$\int_{t_0}^{t} OUR(t^*)\,\mathrm{d}t^* = \alpha \cdot \int_{t_0}^{t} X'(t^*)\,\mathrm{d}t^* + \beta \cdot \int_{t_0}^{t} X(t^*)\,\mathrm{d}t^*. \tag{2}$$

According to data from bioprocesses and previous experience, the stoichiometric parameter β is assumedly not a process constant. During the cultivation, parameter β—oxygen maintenance coefficient for biomass, increases due to biomass concentration growth. The phenomenon of increasing value of parameter β can be explained by the fact that the consumption of oxygen for biomass maintenance also includes the generation of the product and other factors. Such situation occurs at the end of the exponential phase of a microbial cultivation (for recombinant protein synthesis) when the induction (e.g., with isopropyl-D-1-thiogalactopyranoside/IPTG) is performed and the synthesis of the product increases noticeably. As a result, oxygen consumption for biomass maintenance also increases [31, 32]. The parameter β consists of two additive terms

$$\beta = \frac{1}{Y_{XO}} + \frac{1}{Y_{PO}}; \tag{3}$$

where Y_{XO} is oxygen consumption for cells breathing and Y_{PO} is oxygen consumption for product formation.

Consequently, biomass has linear/polynomial relationship to parameter β which is directly dependent on biomass concentration.

The observational data used for proposed biomass estimation was obtained from the processes that involve recombinant protein expression. As it can be seen from the Eq. (3), the parameter β accounts for both, biomass and product, yields. This parameter may exhibit different behavior depending on the process phase and the strain/product involved. However, comprehensive comparison of various strains with respect to the impact, that particular product has on the biomass estimator performance, or to explore the effect on metabolic noise debugging in strain engineering, goes beyond the scope of this study.

To remove the assumption that the stoichiometric parameter β is a function of a biomass, this parameter is expressed as a function of time in the mathematical model. Hence, Eq. (3) is rewritten to linear regression of time:

$$\beta = k_1 * t + k_2; \tag{4}$$

where k_1 and k_2 are linearly dependent mathematical coefficients. When bioprocess is at lag phase or early phase of exponential growth (when biomass concentration is relatively low), the β parameter is extremely small and negligible. Only after induction or specific value of biomass concentration, oxygen consumption for maintenance becomes appreciable. Hence, during a time prior to fact when the Eq. (4) comes into effect, the parameter β should be set to zero in the estimation procedure. At that moment the biomass concentration reaches a value from which the consumption of oxygen for biomass maintenance becomes significant:

$$0 = k_1 * t_i + k_2 \xrightarrow{yields} \tag{5}$$

$$k_2 = -k_1 * \mathrm{t_i}. \tag{6}$$

Then parameter β becomes

$$\beta = k_1 * t - k_1 * t_i; \tag{7}$$

where t_i is the duration from cultivation process start to the time when amount of biomass reaches value resulting in appreciable oxygen maintenance, or when induction is performed and product formation noticeably increases, or when stoichiometry parameter β is no longer zero [9, 31, 32]. In order to have full mathematical model formula, main balance Eq. (2) has parameter β replaced in the linear regression Eq. (7):

$$\int_{t_0}^{t} OUR(t^*)\mathrm{d}t^* = \alpha \cdot \int_{t_0}^{t} X'(t^*)\mathrm{d}t^* + \int_{t_0}^{t} k_1 \cdot (t^* - t_i) \cdot X(t^*)\mathrm{d}t^*. \tag{8}$$

Offline analysis of stoichiometry parameters (stage A)

Prior to the estimation of the biomass, specific cell strain's stoichiometric parameters must be identified during offline analysis. There are few compulsory inputs to approach this task.

- Model fitting procedure requires offline observations: dry cell weight (DCW) or optical density OD value (in o.u.) multiplied by a coefficient of biomass concentration (approximately 0.4 g/l/o.u.) [33];
- Process duration time since cells' inoculation to bioreactor, in hours;
- Oxygen uptake rate (OUR) data since the inoculation;

For model fitting a chosen mathematical expression is equated to gray box model since the collected experimental data is combined with fundamental knowledge about bioprocess [34]. Considering that the bioprocess consists of two main parts, prior to induction and after it, the parameters fitting procedure is based on two independent gray box models. The first one covers the first two cultivation process phases: the lag and exponential. During these phases the amount of biomass is low and materials, resources concentrate to biomass growth [35]. Hence, oxygen requirement for biomass maintenance is minimum and stoichiometric parameter β is negligible:

$$\int_{t_0}^{t_i} OUR(t^*)\mathrm{d}t^* = \alpha \cdot \int_{t_0}^{t_i} X'(t^*)\mathrm{d}t^*, \tag{9}$$

In the Eq. (9) the variable t_i is the time of the induction or the time when biomass reaches a quantity where oxygen usage for maintenance is appreciable. The second cultivation stage represents the biomass growth deceleration and increasing product formation. In this cultivation phase, additional term comes into effect, oxygen consumption for maintenance and product formation, known as stoichiometric parameter β. To properly describe second gray box model, the induction time or time when biomass concentration reaches specific amount must be identified. Throughout this period the maintenance term becomes significant and can't be negligible. After applying maintenance parameter to a model, the second gray box model's expression is generalized to

$$\int_{t_0}^{t} OUR(t^*)\mathrm{d}t^* = \alpha \cdot \int_{t_0}^{t} X'(t^*)\mathrm{d}t^* + \int_{t_0}^{t} k_1 \cdot (t^* - t_i) \cdot X(t^*)\mathrm{d}t^*. \tag{10}$$

In summary, the Eqs. (10) and (11) both yield the conditional definition of cumulative oxygen uptake rate function:

$$\begin{cases} cOUR(t \le t_i) \equiv \int_{t_0}^{t} OUR(t^*)\mathrm{d}t^* = \alpha \cdot (X(t) - X_0); & t \le t_i; \\ cOUR(t > t_i) \equiv \int_{t_0}^{t} OUR(t^*)\mathrm{d}t^* \approx \alpha \cdot (X(t) - X_0) + \sum_{l=i}^{m} k_1 \cdot (t_l - t_i) \cdot X(t_l) \cdot \Delta t_{l,l-1}; & t > t_i. \end{cases} \tag{11}$$

In Eq. (11) the last sum of products is the expression of left Riemann sum [36], i.e. $\int_{t_0}^{t} k_1 \cdot (t^* - t_i) \cdot X(t^*)\mathrm{d}t^* \approx \sum_{l=i}^{m} k_1 \cdot (t_l - t_i) \cdot X(t_l) \cdot \Delta t_{l,l-1}$, when time's t sample is indexed by m. Discrete DCW values define variable $X_l \equiv X(t_l)$, where $l \in [1, n_m]$, n_m is the total number (e.g. hourly) of offline sampling intervals with index m and $X_0 \equiv X(t_0)$ is an initial biomass concentration after inoculation into bioreactor.

Procedure for offline analysis of stoichiometry parameters

The prediction value of the cumulative OUR model [37] for Eq. (11) is

$$cOUR_m \equiv \begin{cases} cOUR(t_m \le t_i) = \alpha \cdot (X_m - X_0); \\ cOUR(t_m > t_i) = \alpha \cdot (X_m - X_0) + \sum_{l=i}^{m} k_1 \cdot (t_l - t_i) \cdot X_l \cdot \Delta t_{l,l-1}; \end{cases} \tag{12}$$

Then the posterior distribution for m-th offline sample is

$$P_{posterior}(cOUR_m) \sim N\left(cOUR_m, \sigma^2_{cOUR}\right), \tag{13}$$

where every sampled prediction m has constant variance σ^2_{cOUR}.

Prior distribution also has the form of Gaussian distribution [38]

$$P_{likelihood}(cOUR_m) \sim N\left(cOUR^*_m, \sigma^2_{cOUR,m}\right), \tag{14}$$

where $cOUR^*_m$ is the m-th observation value of the cumulative OUR and its unique variance is $\sigma^2_{cOUR,m}$.

In previous work [37] the uncertainty of prior distribution was assumed to be equal to the square of observed value, i.e. $\sigma^2_{cOUR,m}$ was assumed to be proportional to $cOUR^{*2}_m$. However, this assumption is not quite rational from practical considerations based on this work experience when deriving a generic estimator for both *E. coli* and yeast cultures. It appears that the assumption of $\sigma^2_{cOUR,m} \sim cOUR^{*2}_m$ is just a special case, which has even more general form. Interestingly this form matches the form of Monod formulation [39] applied to uncertainty, i.e.

$$\sigma^2_{cOUR,m} = \sigma^2_{max} \frac{X^2_m}{K_{X^2} + X^2_m}, \tag{15}$$

where scenario with $K_{X^2} = 0$ resembles least squares approach, i.e. all samples' relative weights become equal, and $K_{X^2} \to \infty$ means that $\sigma^2_{cOUR,m} \sim cOUR^{*2}_m$ as in previous work [37]. In other words, empirical coefficient K_{X^2} is a "weight" coefficient between the two additive terms of optimization criterion. The first term is the least squares criterion and the other is "squared MAPE" criterion as in [37]. Another note about Monod Eqs. (15) and (12) is that the relationship of $\sigma^2_{cOUR,m} \sim \sigma^2_{X,m}$ is valid, i.e. the uncertainty of cumulative OUR is proportional to the uncertainty of biomass variable.

To rationally prepare Eq. (15) for simplified numeric operations avoiding infinities when estimating values, an intrinsic variable K_{exp} expression replaces $K_{X^2} \to \frac{1-K_{exp}}{K_{exp}}$ and transforms Eq. (15) to

$$\sigma^2_{cOUR,m} = \sigma^2_{max} \frac{X^2_m}{\frac{1-K_{exp}}{K_{exp}} + X^2_m} \xrightarrow{yields} \sigma^2_{max} \frac{X^2_m \cdot K_{exp}}{1 - K_{exp} + X^2_m \cdot K_{exp}}, \tag{16}$$

The fact, that σ^2_{max} and K_{exp} both are positive scalar values and do not depend on the index m of a sampling interval, allows to simplify Eq. (16) to

$$\sigma^2_{cOUR,m} \sim \frac{X^2_m}{1 - K_{exp} + X^2_m \cdot K_{exp}}. \tag{17}$$

Equation (17) exposes the physical meaning of K_{exp}. The scenario with $K_{exp} = 0$ recovers $\sigma^2_{cOUR,m} \sim X^2_m$ as in [37]. The scenario with $K_{exp} = 1$ recreates the least squares method as in [38, 40]. Both scenarios show that K_{exp} is an exponential weight, which constructs a hybrid criterion for both least squares and the MAPE squared. Later in the text, the experimental validation will show that there exists a rational empirical value of K_{exp}, which enables estimation of the biomass concentration, with an acceptable precision, for both yeast and *E. coli* cultures since the beginning of the cultivation right after the culture was inoculated to a bioreactor.

After gray box model is identified and hybrid criterion derived, the next step is to use optimization approach to find the stoichiometry parameters. The main equation solving for unknown parameters comes from the maximization of entropy [37, 39] based on Eqs. (13), (14) and (17)

$$S = -\sum_{m=1}^{m\le i} \frac{(cOUR_m - \alpha\cdot(X_m - X_0))^2}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}} - \sum_{m=i+1}^{n_m} \frac{\left(cOUR_m - \alpha\cdot(X_m - X_0) - \sum_{l=1}^{m} k_1\cdot(t_l - t_i)\cdot X(t_l)\cdot\Delta t_{l,l-1}\right)^2}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}}. \quad (18)$$

Hence, at the optimization method, which is shown at the Eq. (18), the whole S expression is maximized, and unknown stoichiometry parameters are found by solving partial derivative of Eq. (18) with respect to α and k_1

$$\begin{cases} \frac{\partial S}{\partial \alpha} = 0; \\ \frac{\partial S}{\partial k_1} = 0. \end{cases} \quad (19)$$

Equation (19) yields the linear system of two equations

$$\begin{cases} \alpha\cdot B + k_1\cdot C = A; \\ \alpha\cdot E + k_1\cdot F = D; \end{cases} \quad (20)$$

where Eq. (20) parameters are:

$$A = \sum_{m=1}^{n_m} \frac{(X_m - X_0)\cdot cOUR_m}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}}; \quad (21)$$

$$B = \sum_{m=1}^{n_m} \frac{(X_m - X_0)^2}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}}; \quad (22)$$

$$C = \sum_{m=i+1}^{n_m} \frac{(X_m - X_0)\cdot\sum_{l=1}^{m}(t_l - t_i)\cdot X(t_l)\cdot\Delta t_{l,l-1}}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}}; \quad (23)$$

$$D = \sum_{m=i+1}^{n_m} \frac{cOUR_m\cdot\sum_{l=1}^{m}(t_l - t_i)\cdot X_l\cdot\Delta t_{l,l-1}}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}}; \quad (24)$$

$$E = \sum_{m=i+1}^{n_m} \frac{(X_m - X_0)\cdot\sum_{l=1}^{m}(t_l - t_i)\cdot X_l\cdot\Delta t_{l,l-1}}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}}; \quad (25)$$

$$F = \sum_{m=i+1}^{n_m} \frac{\left(\sum_{l=1}^{m}(t_l - t_i)\cdot X_l\cdot\Delta t_{l,l-1}\right)^2}{\frac{X_m^2}{K_{exp}\cdot X_m^2+(1-K_{exp})}}. \quad (26)$$

Equations (20)–(26) finalizes the offline estimation of stoichiometry parameters, which are then later used for online estimation of biomass concentration. However, the variable t_i has no direct meaning with yeast cultures, so it must be dealt with separately. First, the specific time when the maintenance coefficient becomes appreciable is analyzed in the next subsection.

Identification of yeasts' specific time for maintenance

Variable t_i at Eq. (12) is the time of induction or the time when biomass concentration reaches a specific amount when oxygen maintenance for cells becomes non negligible. In the case of cultivation processes of *E. coli*, the induction time is known, i.e. it can be defined by the time moment when IPTG solution is injected into bioreactor. In the cultivation process of *S. cerevisiae* yeasts the IPTG solution was not used. Hence, the variable t_i defines the time when biomass concentration reaches a specific value when maintenance coefficient becomes noticeable. The search for t_i utilizes the convex optimization method and maximization of entropy [37, 41]. The optimization procedure is depicted in Fig. 4.

The knowledge of the specific time t_i enables the biomass concentration estimation. However, the specific time t_i is not known in advance prior to online experiment with yeast cells, because it has just a theoretical meaning in this case. Therefore, a generic relationship between the maintenance coefficient value and the biomass concentration will be inferred in the next subsection. Such a generic form of maintenance coefficient will enforce online estimation without dependence on the type of the microbial culture. Moreover, the value of the specific time t_i becomes irrelevant for the online estimation procedure.

Identification of maintenance coefficient parts

After optimization of stoichiometry parameters, which had determined unknown parameters of the mathematical method, the next step is to validate those identified parameters with experimental data. Prior to comparison of theoretical and experimental data, the mathematical model, as in Eq. (7), must be reconstructed so that β is no longer a function of time and still satisfies the actual behavior of biotechnological process. The stoichiometric parameter β directly depends on biomass concentration

$$\beta(X) \equiv \beta(X(t)) = k_{\beta 2}\cdot X^2(t) + k_{\beta 1}\cdot X(t) + k_{\beta 0}; \quad (27)$$

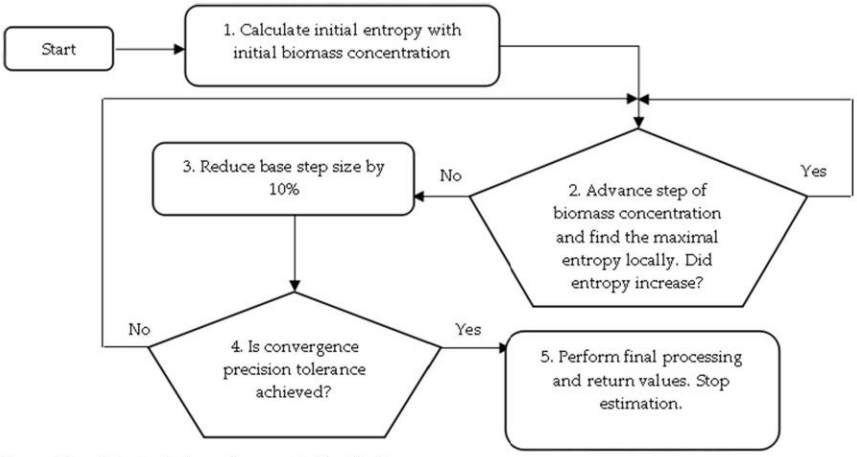


Fig. 4 The workflow of structural scheme for parameter identification

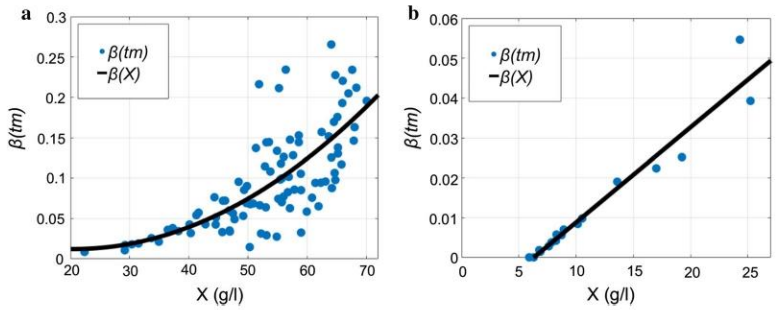


Fig. 5 Dependence of oxygen consumption for maintenance on biomass concentration, **a** *E. coli*, **b** *Saccharomyces cerevisiae*

The expression of parameter $\beta(X)$ represents a parabola regression of biomass in the case of the *E. coli* strain Fig. 5a. Meanwhile, *S. cerevisiae* oxygen consumption for maintenance is dependent linearly on biomass concentration, thus $k_{\beta s2} = 0$,

$$\beta_{Saccharomyces}(X) \equiv \beta_{Saccharomyces}(X(t)) = k_{\beta s1} \cdot X(t) + k_{\beta s0}; \quad (28)$$

In Eqs. (27) and (28) regression coefficients connect maintenance coefficient β to biomass variable. In both culture cases, stage A helps to obtain β values from linear regression based on Eq. (7) output

$$\beta(X_m) \cong \beta(t_m) = k_1 * (t_m - t_i). \quad (29)$$

The assumed relationship of $\beta(X)$ considering biomass concentration is presented in Fig. 5.

According to data from cultivation processes of *E. coli* in Fig. 5, the stoichiometric parameter of cell maintenance can be assumed as directly dependent on biomass in parabolic manner. At the cultivation processes of *E. coli*, the induction of IPTG, which initiates product synthesis, may cause nonlinear dependence of oxygen consumption on biomass maintenance. Based on Eqs. (27) and (28), it is possible to calculate strain's specific biomass concentration ($X_{specific}$) when oxygen consumption for maintenance is no longer negligible. This is done by setting Eqs. (27) and (28) to zero and solving them for the specific biomass concentration $X_{specific}$

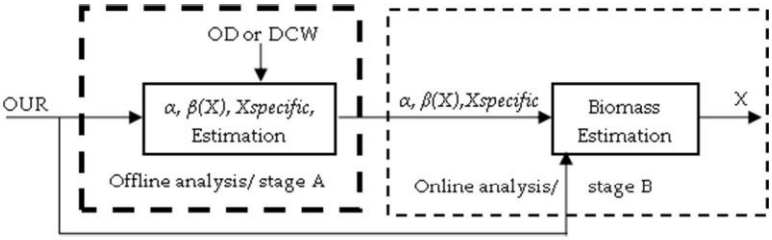


Fig. 6 Biomass estimation workflow

$$\beta\left(X_{specific}\right) \equiv \beta(X(t)) = 0; \tag{30}$$

The workflow of both stoichiometry and biomass estimations improves structure, as in Fig. 3, to the shape of the one in Fig. 6.

The solution of Eq. (30) identifies the specific biomass concentration $X_{specific}$ and finalizes the offline estimation of stoichiometry coefficients for a strain. After the stoichiometry coefficients are found in stage A, a generic procedure for online biomass estimation can be performed independently on the knowledge of bioreactor parameters. In conclusion, β, as in Eq. (27), transforms Eq. (1) into

$$\begin{cases} OUR(t) = \alpha \cdot X'(t) + k_{\beta 2} \cdot X^3(t) + k_{\beta 1} \cdot X^2(t) + k_{\beta 0} \cdot X(t), X(t) > X_{specific}; \\ OUR(t) = \alpha \cdot X'(t), X(t) \leq X_{specific}. \end{cases} \tag{31}$$

In spite of the fact that Eq. (31) form is the third order function, it is still the same equation as Eq. (1). However, it was inferred by the estimation procedure and the observation data in Fig. 5. Variable β manipulation compensates the effect of biomass concentration X on β and makes all Eq. (31) coefficients linearly dependent and constant throughout the course of the experiment. Eventually, this serves as a prerequisite to the simplified generic procedure for estimation of biomass concentration, coming in the next subsection.

Online estimation of biomass concentration (stage B)

In this paper, estimation of biomass concentration is based on stoichiometric parameters and cumulative oxygen uptake rate *cOUR*. When stoichiometric parameters are discovered in stoichiometry estimation, stage A, or it was given, only one input from bioreactor system, cumulative oxygen uptake rate, is necessary to estimate the biomass state. This procedure is depicted by stage B (online analysis) in Fig. 6. The block of "biomass estimation", Fig. 6, consists of two main scenarios which both return biomass concentration at a time instance with index m.

Prior to the specific biomass $X_{specific}$ level is reached, i.e. when oxygen consumption for maintenance is very low or negligible, biomass state estimator equation is

$$X_m = \frac{cOUR_m}{\alpha} + X_0. \tag{32}$$

After biomass concentration exceeds $X_{specific}$ during the second scenario, i.e. oxygen consumption becomes noticeable, the stoichiometric parameter β comes into effect as a function of biomass concentration. Equation (12) helps to derive the approximate estimator for biomass state, as follows

$$X_m \cong \frac{cOUR_m - \sum_{l=i}^{m-1} \beta(X_l) \cdot X_l \cdot \Delta t_{l,l-1}}{\alpha} + X_0. \tag{33}$$

The variable X_0, as in Eqs. (32) and (33), is an initial biomass concentration at the time of inoculation into bioreactor. Its value can be either a dry biomass measurement value or optical density OD value (in o.u.) multiplied by a coefficient of biomass concentration (approximately 0.4 g/l/o.u.).

This subsection initializes the online biomass estimation procedure (Fig. 7), which can be used in biotechnological industrial practices. The suggested approach does not require the bioreactor-dependent parameters, it serves as a good candidate to be applied to more microbial strains and the experimental validation, in the coming section, will show that such an approach can be used for biomass estimation since the time moment of inoculation into bioreactor.

Experimental validation

Validation performance indicators

Both mean absolute error (MAE) and mean absolute percentage error (MAPE) were used as indicators to evaluate

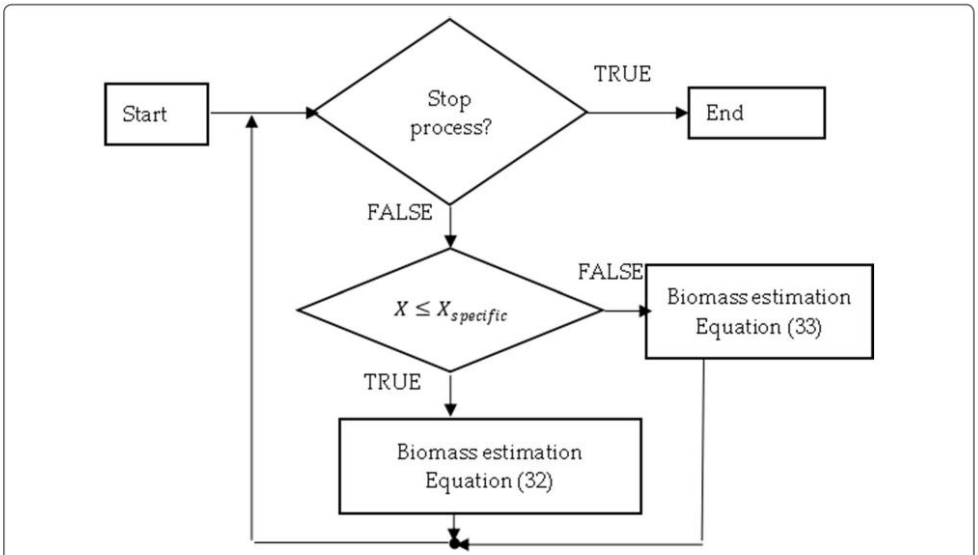


Fig. 7 Workflow of online biomass estimation (stage B)

the estimation results. MAE and MAPE methods both evaluate the errors between estimated and observed biomass values of a cultivation process. MAE approach is defined as follows [42]:

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}, \tag{34}$$

where n is the number of data counts, $\hat{y}_i$ is estimation result, which is compared to y_i, the observed value from the cultivation process. Mean absolute error represents average vertical distance between both values. MAPE method can be expressed as follows [43]:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|. \tag{35}$$

The mean absolute percentage error is a statistical measure representing the accuracy of a forecast system, in percentage. Root mean square error represents the square root of residuals of the differences between predicted values and observed values. RMSE method's formula are as follows [42]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}. \tag{36}$$

Table 1 Stoichiometric parameters of cell strains

Escherichia coli	*Saccharomyces cerevisiae*
$\alpha = 1.01$	$\alpha = 1.35$
Confidence Interval ∓ 0.0186	Confidence Interval ∓ 0.149
$k_{\beta e2} = 7.2 \cdot 10^{-5}$	$k_{\beta s2} = 0$
$k_{\beta e1} = -2.9625 \cdot 10^{-3}$	$k_{\beta s1} = 2.3851 \cdot 10^{-3}$
$k_{\beta e0} = 4.27047\text{d} \cdot 10^{-2}$	$k_{\beta s0} = -1.5014 \cdot 10^{-2}$
$X_{specific} = 20.6\text{g/l}$	$X_{specific} = 6.29\text{g/l}$
$K_{exp} = 0.4$	

Comparative analysis of experimental results

Experimental biomass measurements and data of cumulative oxygen uptake rate *cOUR* from fed-batch experiments of *E. coli* and *S. cerevisiae* were taken from [8], experiments led by authors of this text and industrial R&D laboratories. There were three cultivations of *E. coli* cells in 15 l bioreactor with limited substrate feed [8] and two R&D laboratory cultivations of *S. cerevisiae* yeasts in 5 l bioreactor with limited substrate feed. Additionally, there was one cultivation of *E. coli* in 12 l bioreactor with limited substrate feed and there were 15 cultivations in 5 l bioreactor, out of which 7 cultivations were with dosed substrate feeding. As the first step, all cultivation data was analyzed in the stoichiometric parameters'

estimation (stage A). The estimation procedure ignored both metabolism pathways, occurring during dosed substrate feed cultivations, and increasing product formation due to IPTG injections. The results of offline analysis of stoichiometric parameters are present in Table 1.

The tuning coefficient K_{exp} was identified empirically and its value of 0.4 showed acceptable outcome for the performed experiments. However, *S. cerevisiae* stoichiometric results come from just two cultivation experiments. Therefore, the results might still be improved when more experimental data becomes available in the future.

In industrial processes, strain's stoichiometric parameters are given, unless they were estimated using offline analysis, stage A. Then biomass concentration is calculated iteratively using both Eqs. (32) and (33) from *cOUR* signal (online analysis, stage B). This work's biomass estimation method used different cultivation experiments, with different cell strains, bioreactor volumes, type of substrates feeding solution, different IPTG induction time moment and their corresponding OD levels at IPTG injection, different substrate feeding limitations and different time of starting the substrate feed. Estimation results are shown in Table 2.

Seven experiments (#5–#11) were performed with dosed substrate feeding. Meanwhile the rest of experiments had limited feeding with various combinations of control strategies described in [37]: multiple different substrate limited feedings prior to induction and after it.

The overall average MAE of biomass estimation since inoculation is 1.1 g/l and overall average MAE of biomass estimation since feed start is 1.41 g/l. The overall average MAPE of biomass estimation since inoculation is 7.28% and overall average MAPE of biomass estimation since feed start is 6.29%. Overall average RMSE value of *S. cerevisiae* cultivations is 0.5 g/l. RMSE value of *E. coli* cultivations with limited substrate feeding is 1.26 g/l and for cultivations with dosed substrate feeding is 2.44 g/l. RMSE value of *E. coli* cultivations before stationary phase, when DCW reaches ~40 g/l (to compare with results in [22]) with limited substrate feeding, is 1.07 g/l and for cultivations with dosed substrate feeding is 1.2 g/l. These results show that this approach improves the precision achieved in [22] without compromising the simplicity of the implementation. Offline analysis (stage A) execution lasted 2–15 ms and online analysis (stage B) calculations took 13–30 ms on a single core CPU in bioprocess engineering software tool dedicated for the purposes of this work. No initial conditions for numeric

Table 2 Analysis of experiments for biomass estimation

No.	Strain	Bioreactor volume, l	MAE since inoculation g/l	MAE since feed start, g/l	MAPE since inoculation, %	MAPE since feed start, %
1	*Escherichia coli*	15	1.04	1.04	5.7	5.7
2			0.96	0.96	4.11	4.11
3			1.38	1.38	5.61	5.61
4		12	1.45	1.6	5.3	4.91
5		7	1.65	2.3	7.26	5.34
6			2.5	3.3	6.9	7
7			1.54	2.13	8.67	6.28
8			1.29	1.94	7.57	6.12
9			1.52	2.54	7.12	7.96
10			1.55	2.09	9.35	6
11			1.87	2.7	10.75	6.6
12			1.16	1.55	6.88	6.61
13			0.97	1.22	9.03	6.89
14			0.78	1.01	9.88	6.26
15			1	1.22	8.92	6.99
16			0.76	1	7.14	7.85
17			0.74	0.9	8.51	5.69
18			1.18	1.54	7.83	7.55
19			0.8	1.07	5.83	6.30
20	Yeasts	5	0.29	0.29	6.69	6.69
21			0.66	0.66	7.28	7.28

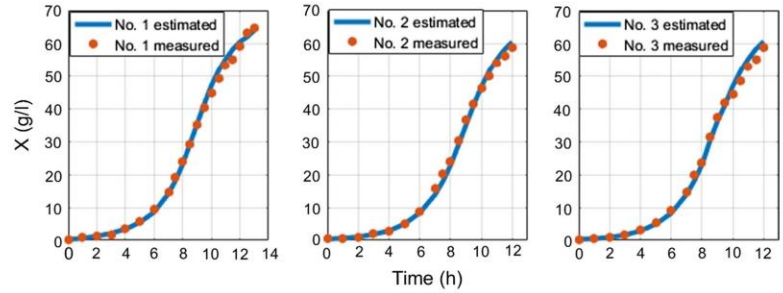


Fig. 8 Biomass concentration estimation result with recombinant *E.coli* BL21(DE3) pLysS strain at 15 L bioreactor

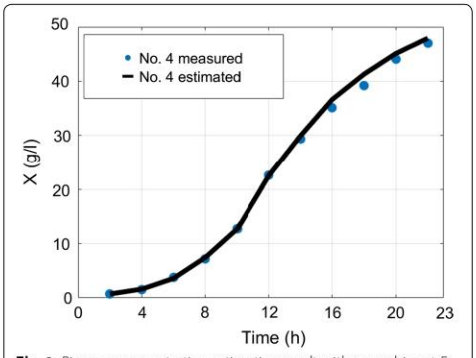


Fig. 9 Biomass concentration estimation result with recombinant *E. coli* BL21(DE3) pET9a-IdeS strain at 12 l bioreactor

optimization procedure were used. The speed of online estimation can be explained by the fact that the prediction value of biomass concentration estimate is calculated once during the whole estimation procedure. There is no updating performed for the predicted value of biomass. In the future, this optimization condition might be released though. The substrate feed was started from the beginning of cultivation process right after inoculation moment in the experiments #1–#3 and #20–#21, while for the rest of cultivations had their substrate feed started after 5–6 h since inoculations. The errors between off-line and on-line data mainly originate from offline measurements. Especially in #5–#19, because historically the accuracy of offline measurements was not of high priority during these experiments. Therefore, in the future the true ground truth of biomass concentration might testify that the approach suggested in this work has even higher overall precision than the one stated in above. All biomass state estimation results are shown at the Figs. 8, 9, 10, 11, 12.

Conclusions

The suggested biomass estimation's numeric approach using cumulative oxygen uptake rate signal showed no dependability on selection of the initial variable values for optimization procedures. This study assumed, by Pareto principle, that the proposed method is only dependent on stoichiometry parameters of the strain, i.e. the developed noninvasive biomass estimation procedure was made to not depend on both the manipulation with a specific growth rate variable and the selection of corresponding bioreactor parameters. The precision errors, since the bioprocess start, when inoculant was injected to a bioreactor, confirmed that the approach is relevant for online biomass state estimation. This included the lag and exponential growth phases for both *E. coli* and *S. cerevisiae*. The experimental investigation of *E. coli* and *S. cerevisiae* cultures showed that the estimation procedure is identical for both cultures. The overall average MAE of biomass estimation since inoculation is 1.1 g/l and the overall average MAPE of biomass

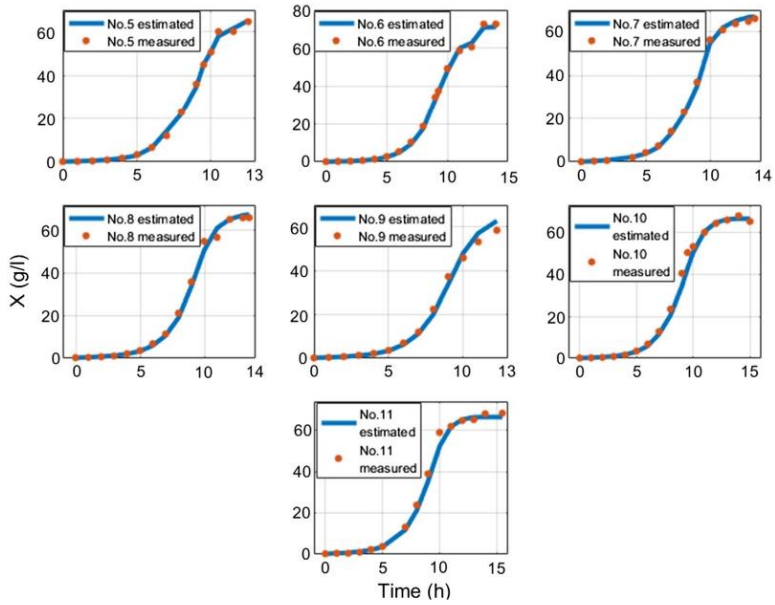


Fig. 10 Biomass concentration estimation result with recombinant *E. coli* BL21 (DE3) pET21-IFN-alfa-5 strain at 7 L bioreactor with dosed substrate feeding

estimation since inoculation is 7.28%. RMSE value of *E. coli* cultivations before stationary phase, when DCW reaches ~40 g/l (to compare with results of other authors) with limited substrate feeding, is 1.07 g/l and for cultivations with dosed substrate feeding is 1.2 g/l. These results show that this approach improves the precision achieved by other authors without compromising the simplicity of the implementation. Moreover, the suggested approach is a candidate method to be the microorganisms' culture invariant approach, it does not depend on any numeric initial optimization conditions, and it does not require any of bioreactor parameters. No numeric stability issues of convergence occurred during multiple performance tests. All this makes this approach a potential candidate for industrial tasks with adaptive feeding control or automatic inoculations when substrate feeding profile and bioreactor parameters are not provided.

Neither numeric artifacts nor abrupt worst-case scenarios were experienced during both offline and online analysis of 21 experiments, out of which 7 ones were carried out with dosed substrate feeding. The experiments executed in 5 l, 7 l, 12 l and 15 l bioreactor volumes. Feed start, inoculation, bioreactor medium, feeding limitation and other conditions varied with no manual control or adjustment. This encourages the use of such estimator in adaptive feedback control systems. Both online and offline estimations were tested on a single core CPU processing and each procedure took no more than 30 ms when overall 1-min interval data was sampled

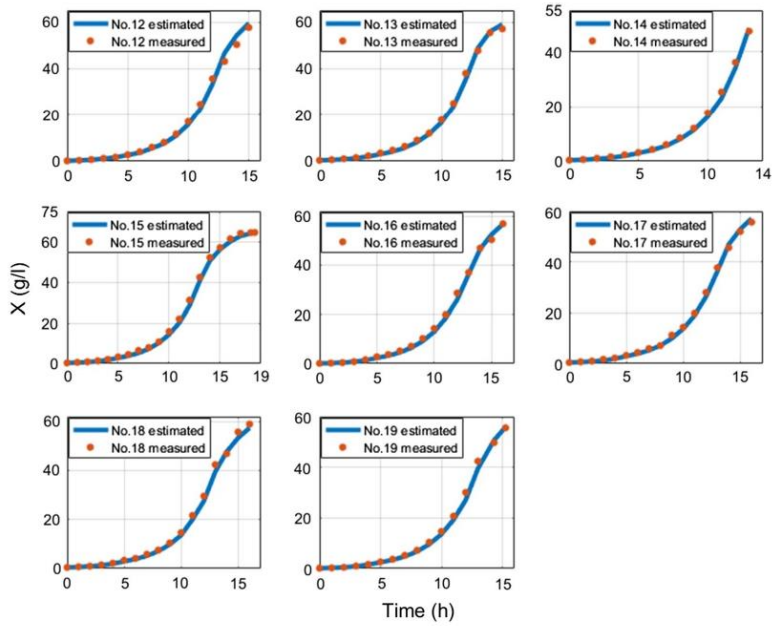


Fig. 11 Biomass concentration estimation result with recombinant *E. coli* BL21 (DE3) pET21-IFN-alfa-5 strain at 7 L bioreactor with limited substrate feeding

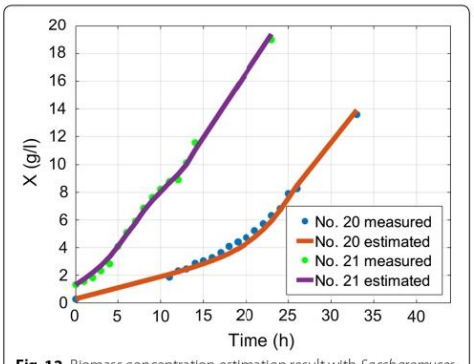


Fig. 12 Biomass concentration estimation result with *Saccharomyces cerevisiae* (no DY7221) strain at 5 L bioreactor

from cumulative oxygen uptake signal, which makes the approach of practical use too. Finally, this estimator does require a usage of regular industrial gas analysis equipment such as BlueSens etc.

Acknowledgements
We are grateful to professor Rimvydas Simutis (Kaunas University of Technology) for kindly providing the motivation and support that inspired and encouraged this publication.

Authors' contributions
RU and AS contributed to the preparation of the manuscript. All authors have read and approved the final manuscript. Conceptualization: RU; methodology: RU; software: RU, AS; validation: RU, AS, DP; formal analysis: RU, AS; investigation: RU, AS, DP; writing—original draft preparation: AS, RU; writing—review and editing: RU, VG; supervision: VB, RU; project administration: RU, VB, VG; funding acquisition: VB, RU, VG. All authors read and approved the final manucript.

Funding
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.

Availability of data and materials
Some datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Automation, Kaunas University of Technology, 51367 Kaunas, Lithuania. [2] Biopharmaceutical Division of Centre for Innovative Medicine, 08406 Vilnius, Lithuania.

Received: 31 July 2019 Accepted: 23 October 2019
Published online: 05 November 2019

References

1. OPS Process Analytical Technology—(PAT) Initiative. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pat-framework-innovative-pharmaceutical-development-manufacturing-and-quality-assurance. Accessed 31 Oct 2019.
2. Goodwin GC. Predicting the performance of soft sensors as a route to low cost automation. Annu Rev Control. 2000;24:55–66. https://doi.org/10.1016/S1367-5788(00)90013-0.
3. Larroche C, Sanromán MÁ, Du G, Pandey A, editors. Current developments in biotechnology and bioengineering: bioprocesses, bioreactors and controls. Amsterdam: Elsevier; 2016.
4. Schaepe S, Kuprijanov A, Sieblist C, Jenzsch M, Simutis R, Lübbert A. Current advances in tools improving bioreactor performance. CBIOT. 2013;3:133–44. https://doi.org/10.2174/22115501026661312172352 46.
5. Galvanauskas V, Volk N, Simutis R, Lübbert A. Design of recombinant protein production processes. Chem Eng Commun. 2004;191:732–48. https://doi.org/10.1080/00986440490276056.
6. Simutis R, Lübbert A. Bioreactor control improves bioprocess performance. Biotechnol J. 2015;10:1115–30. https://doi.org/10.1002/biot.201500016.
7. Bumelis VA. European Patent No. EP2532734A1; 2012. https://patents.google.com/patent/EP2532734A1. Accessed 31 Oct 2019.
8. Schaepe S, Kuprijanov A, Simutis R, Lübbert A. Avoiding overfeeding in high cell density fed-batch cultures of *E. coli* during the production of heterologous proteins. J Biotechnol. 2014;192:146–53. https://doi.org/10.1016/j.jbiotec.2014.09.002.
9. Rosenfeld E, Beauvoit B, Blondin B, Salmon J-M. Oxygen consumption by anaerobic *Saccharomyces cerevisiae* under enological conditions: effect on fermentation kinetics. Appl Environ Microbiol. 2003;69:113–21. https://doi.org/10.1128/AEM.69.1.113-121.2003.
10. van Dijken JP, Weusthuis RA, Pronk JT. Kinetics of growth and sugar consumption in yeasts. Antonie Van Leeuwenhoek. 1993;63:343–52. https://doi.org/10.1007/BF00871229.
11. Gnoth S, Kuprijanov A, Simutis R, Lübbert A. Simple adaptive pH control in bioreactors using gain-scheduling methods. Appl Microbiol Biotechnol. 2010;85:955–64. https://doi.org/10.1007/s00253-009-2114-5.
12. Mansano R, Godoy E, Porto A. The benefits of soft sensor and multi-rate control for the implementation of wireless networked control systems. Sensors. 2014;14:24441–61. https://doi.org/10.3390/s141224441.
13. Galvanauskas V, Simutis R, Levisauskas D, Repšyte J, Lübbert A. Comparison of state estimation techniques for biotechnological processes. In: 8th international conference on electrical and control technologies, ECT 2013; p. 70–5.
14. Linko P, Zhu Y. Neural network programming in bioprocess variable estimation and state prediction. J Biotechnol. 1991;21:253–69. https://doi.org/10.1016/0168-1656(91)90046-X.
15. Luedeking R, Piret EL. A kinetic study of the lactic acid fermentation. Batch process at controlled pH. Biotechnol Bioeng. 1959;1:393–412. https://doi.org/10.1002/jbmte.390010406.
16. Simutis R, Galvanauskas V, Levisauskas D, Repsyte J, Vaitkus V. comparative study of intelligent soft-sensors for bioprocess state estimation. JOLST. 2013. https://doi.org/10.12720/jolst.1.3.163-167.
17. Unrean P. Bioprocess modelling for the design and optimization of lignocellulosic biomass fermentation. Bioresour Bioprocess. 2016;3:1. https://doi.org/10.1186/s40643-015-0079-z.
18. Caramihai M, Severi I. Bioprocess modeling and control. In: Matovic MD, editor. Biomass now—sustainable growth and use. Rijeka: InTech; 2013. https://doi.org/10.5772/55362.
19. Gnoth S, Jenzsch M, Simutis R, Lübbert A. Process Analytical Technology (PAT): batch-to-batch reproducibility of fermentation processes by robust process operational design and control. J Biotechnol. 2007;132:180–6. https://doi.org/10.1016/j.jbiotec.2007.03.020.
20. Wechselberger P, Sagmeister P, Herwig C. Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. Bioprocess Biosyst Eng. 2013;36:1205–18. https://doi.org/10.1007/s00449-012-0848-4.
21. Schubert J, Simutis R, Dors M, Havlik I, Lübbert A. Bioprocess optimization and control: application of hybrid modelling. J Biotechnol. 1994;35:51–68. https://doi.org/10.1016/0168-1656(94)90189-9.
22. Jenzsch M, Simutis R, Eisbrenner G, Stückrath I, Lübbert A. Estimation of biomass concentrations in fermentation processes for recombinant protein production. Bioprocess Biosyst Eng. 2006;29:19–27. https://doi.org/10.1007/s00449-006-0051-6.
23. Gnoth S, Jenzsch M, Simutis R, Lübbert A. Control of cultivation processes for recombinant protein production: a review. Bioprocess Biosyst Eng. 2008;31:21–39. https://doi.org/10.1007/s00449-007-0163-7.
24. Galvanauskas V, Simutis R, Lübbert A. Hybrid process models for process optimisation, monitoring and control. Bioprocess Biosyst Eng. 2004;26:393–400. https://doi.org/10.1007/s00449-004-0385-x.
25. Aehle M, Simutis R, Lübbert A. Comparison of viable cell concentration estimation methods for a mammalian cell cultivation process. Cytotechnology. 2010;62:413–22. https://doi.org/10.1007/s10616-010-9291-z.
26. Petkov SB, Davis RA. On-line biomass estimation using a modified oxygen utilization rate. Bioprocess Eng. 1996;15:43–5. https://doi.org/10.1007/BF00435527.
27. Barrigón JM, Ramon R, Rocha I, Valero F, Ferreira EC, Montesinos JL. State and specific growth estimation in heterologous protein production by *Pichia pastoris*. AIChE J. 2012;58:2966–79. https://doi.org/10.1002/aic.12810.
28. Karim MN, Rivera SL. Artificial neural networks in bioprocess state estimation. Modern biochemical engineering. Berlin: Springer; 1992. p. 1–33. https://doi.org/10.1007/bfb0000703.
29. Caticha A. Entropic priors. In: AIP conference proceedings. jackson hole, Wyoming (USA): AIP; 2004. p. 371–80. https://doi.org/10.1063/1.1751380.
30. Gencaga D, Knuth K, Rossow W. A recipe for the estimation of information flow in a dynamical system. Entropy. 2015;17:438–70. https://doi.org/10.3390/e17010438.
31. Garcia-Ochoa F, Gomez E, Santos VE, Merchuk JC. Oxygen uptake rate in microbial processes: an overview. Biochem Eng J. 2010;49:289–307. https://doi.org/10.1016/j.bej.2010.01.011.
32. Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. Protein Sci. 2009;18:936–48. https://doi.org/10.1002/pro.102.
33. Shiloach J, Fass R. Growing *E. coli* to high cell density—a historical perspective on method development. Biotechnol Adv. 2005;23:345–57. https://doi.org/10.1016/j.biotechadv.2005.04.004.
34. Bohlin T. Practical grey-box process identification: theory and applications. London: Springer; 2006.
35. Schuler MM, Marison IW. Real-time monitoring and control of microbial bioprocesses with focus on the specific growth rate: current state and perspectives. Appl Microbiol Biotechnol. 2012;94:1469–82. https://doi.org/10.1007/s00253-012-4095-z.
36. Swokowski EW. Calculus with analytic geometry. 2d ed. Boston: Prindle, Weber & Schmidt; 1979.
37. Urniezius R, Galvanauskas V, Survyla A, Simutis R, Levisauskas D. From physics to bioengineering: microbial cultivation process design and feeding rate control based on relative entropy using nuisance time. Entropy. 2018;20:779. https://doi.org/10.3390/e20100779.
38. Giffin A, Urniezius R. The Kalman filter revisited using maximum relative entropy. Entropy. 2014;16:1047–69. https://doi.org/10.3390/e16021047.
39. Monod J. The growth of bacterial cultures. Annu Rev Microbiol. 1949;3:371–94. https://doi.org/10.1146/annurev.mi.03.100149.002103.

40. Giffin A, Urniezius R. Simultaneous state and parameter estimation using maximum relative entropy with nonhomogenous differential equation constraints. Entropy. 2014;16:4974–91. https://doi.org/10.3390/e16094974.
41. Urniezius, R. Convex programming for semi-globally optimal resource allocation; 2016. p. 040002. https://doi.org/10.1063/1.4959056.
42. Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res. 2005;30:79–82. https://doi.org/10.3354/cr030079.
43. de Myttenaere A, Golden B, Le Grand B, Rossi F. Mean absolute percentage error for regression models. Neurocomputing. 2016;192:38–48. https://doi.org/10.1016/j.neucom.2015.12.114.

Publisher's Note

entropy

MDPI

Article

Identification of Functional Bioprocess Model for Recombinant *E. Coli* Cultivation Process

Renaldas Urniezius * and Arnas Survyla

Department of Automation, Kaunas University of Technology, Kaunas LT-51367, Lithuania; arnas.survyla@ktu.lt
* Correspondence: renaldas.urniezius@ktu.lt

Received: 10 November 2019; Accepted: 12 December 2019; Published: 14 December 2019

Abstract: The purpose of this study is to introduce an improved Luedeking–Piret model that represents a structurally simple biomass concentration approach. The developed routine provides acceptable accuracy when fitting experimental data that incorporate the target protein concentration of *Escherichia coli* culture BL21 (DE3) pET28a in fed-batch processes. This paper presents system identification, biomass, and product parameter fitting routines, starting from their roots of origin to the entropy-related development, characterized by robustness and simplicity. A single tuning coefficient allows for the selection of an optimization criterion that serves equally well for higher and lower biomass concentrations. The idea of the paper is to demonstrate that the use of fundamental knowledge can make the general model more common for technological use compared to a sophisticated artificial neural network. Experimental validation of the proposed model involved data analysis of six cultivation experiments compared to 19 experiments used for model fitting and parameter estimation.

Keywords: gray box; relative entropy; microbial cultivation; numerical convex optimization; parameter estimation; stoichiometry

1. Introduction

Biotechnology plants seek to increase the productivity and controllability of cell cultivations. In order to achieve those two quality conditions, they need a trustworthy data collection system that provides mandatory variables in real time to smoothly control processes and achieve the required productivity. A system like this would require compatible equipment that might restrict access due to financial concerns, because it is not compatible with the chosen system or may lack functionality. However, it is worth replacing sophisticated equipment with soft sensors that estimate the desired non-observable parameters from the measured data collected [1,2].

In previous works [3], the biomass estimation model relied on stoichiometry, where biomass maintenance eventually proved to be the third-order polynomial term. The biomass maintenance term consists of oxygen consumption, not only for maintenance, but also for product synthesis [4]. This study suggests fundamental knowledge based on the Luedeking–Piret model [5], in which the infrastructure of the maintenance term consists of both actual biomass maintenance and target protein production. In this case, the proposed model is clearer and can achieve greater accuracy. The main aim of this paper is to estimate the Luedeking–Piret model parameters in the offline mode using product information. Simultaneously, this paper provides an alternative way to fit a protein production model and analyze the parameters of the model based on offline data. To date, few studies and publications have estimated the state of target proteins using a soft-sensor approach. The majority of published works estimating target protein productivity and biomass concentration use an artificial neural network (ANN) approach. The novelty of this study is that it involves the fundamental knowledge of incorporating target protein synthesis into a product and biomass concentration model. This results in a rational parametric model that can serve as an alternative approach to ANN. Parameters of the proposed model of estimation

have practical significance; therefore, approach-related artifacts are less expected and their elimination is manually controllable during the development of industrial processes.

Section 2: Materials and Methods describe the materials, strains, and bioreactor system operating conditions. Section 3: Basis of Biomass and Product Model Fitting reviews the idea and basis of this study. Section 4: System Identification and Parameter Estimation presents the derivation of a known method for fitting to target protein and biomass concentration models. It also lays out a general formula for oxygen consumption according to the stoichiometric coefficients of biomass, which is relevant to the specific culture of Escherichia coli. Section 5: Experimental Validation provides results from experimental data supporting the validity of the improved Luedeking–Piret model and an offline estimation of target protein and biomass concentrations. Section 6: Conclusions discusses the results and concludes the final statements of this study.

2. Materials and Methods

2.1. Cell Strains

In this work, *E. coli* BL21 (DE3) pET28a (Novagen) served as the test object in all experiments [4] in order to validate biomass and protein model fitting. The product of *E. coli* BL21 (DE3) appeared in two forms: Active soluble and insoluble forms, which were formed as inclusion bodies. In this study, the target product was insoluble protein, inclusion bodies. The protein's expression was under the control of the T7 promoter after induction with 1 mM isopropyl-D-1-thiogalactopyranoside (IPTG).

2.2. Medium and Culture Conditions

Experimental data [6–8] served as the basis for analysis in this study. Genetically modified *E. coli* BL21 (DE3) pET28a strain was cultivated in a B. Braun 10 L bioreactor. Due to confidentiality restrictions, the authors of Reference [6] claimed that the organism expressed commercial protein, and no specific details are available on the target recombinant protein. The initial medium volume at inoculation was 5 L. The cultivation medium contained mineral salt medium, consisted of Na_2SO_4, 2.0 g/L; $(NH_4)_2SO_4$, 2.46 g/L; NH_4Cl, 0.5 g/L; K_2HPO_4, 14.6 g/L; $NaH_2PO_4 \times H_2O$, 3.6 g/L; $(NH_4)_2$–H–citrate 1.0 g/L; $MgSO_4 \times 7H_2O$, 1.2 g/L; trace element solution, 2 mL/L; thiamine, 0.1 g/L; and kanamycin, 0.1 g/L [6]. Cultivation experiments took place in fed-batch mode with zero glucose concentration in the bioreactor at the inoculation time. Pumping of the feed solution containing glucose and mineral salts in the same composition as the starting medium started after inoculation in the bioreactor [7]. During all experiments, after inoculation, the initial biomass inside the bioreactor was 0.25 g/L of dry cell weight (DCW). At the beginning of cultivation, the feed rate of the substrate was set very low, approximately 11–15 g/h, and used substrate solution with as low as 300 g/kg glucose concentration to avoid overdose, which resulted in substrate inhibition or a different metabolic pathway. At ~4 g/L biomass concentration in medium, feed solution of 600 g/kg replaced the one with 300 g/kg glucose concentration [8]. The temperature set point in the bioreactor was set at 35 °C. The induction time was 10 h since inoculation. Tracking of off-gassing from the bioreactor was done online, and a paramagnetic oxygen sensor (Maihak Oxor 610) operated for O_2 concentration observation. An Ingold DO probe (Mettler Toledo) measured dissolved oxygen tension (DOT) values. The DOT set point was set to 25% of saturation [9]. pH was measured with an Ingold pH probe (Mettler Toledo) and kept at 7.0 using a PID controller [10]. After the action of cell disruption, separation of the soluble fraction, and solubilization of inclusion bodies, SDS-PAGE electrophoresis helped to determine the amount of the target protein. The method for measuring protein concentration (g/L) consisted of several preparation steps. Initially, 200 g of wet biomass was dissolved in 1 mL of solution and mixed for 30 min. After that SDS-PAGE, (sodium dodecyl sulfate polyacrylamide gel) electrophoresis was performed on 200 µL of the suspension sample to measure the amount of total protein concentration. The remainder of the suspension was mixed with SDS (sodium dodecyl sulfate) buffer to dissolve all proteins and centrifuged for 15 min at 4 °C with 20,000 G force. After centrifugation, SDS-PAGE electrophoresis was used to

determine the soluble protein concentration in a 200 µL sample. The remaining supernatant discarded and replaced with 1 mL of water, then mixed and centrifuged. After decanting the supernatant, 1 mL of solubilization buffer (8M urea; 50 mM, pH 8.0 Tris base) was added and mixed for approximately 12 h. The final step after mixing was the measuring of insoluble protein (inclusion bodies) concentration with SDS-PAGE electrophoresis.

3. Basis of Biomass and Product Model Fitting

A previous study [3] showed that the development of a biomass concentration estimator required data that was linked to the biomass growth rate. Oxygen uptake rate (OUR) was the main characteristic variable that provided information about the biosynthesis phenomenon [7,11,12]. To enforce soft sensors [13,14], OUR must have been an online measurement coming from devices that registered not only mass airflow, but also O_2 concentration in the off-gas [15]. This study proposes biomass concentration and protein model fitting based on a mass balance equation. For fed batch cultivations, such a model originates from the Luedeking–Piret model. The mass balance model represents the relationship between oxygen uptake rate (OUR) and biomass growth characteristics [5]:

$$\mathrm{OUR}(t) = \alpha \cdot X'(t) + \beta \cdot X(t) \tag{1}$$

In Equation (1), X is dry biomass concentration (g/L), t is duration time since inoculation (h) and stoichiometric coefficients α and β represent cell metabolism of oxygen consumption, where α describes the cell's oxygen consumption yield for biomass growth ($\alpha \equiv Y_{o_2/X}, [\,g(O_2)/g(X)]$) and β describes the cell's oxygen consumption for maintenance ($\beta \equiv m_{O_2/X}, [g(O_2)/g(X)/h]$) [16–18]. Çalik [19], studying the effects of pH on benzaldehyde lyase production by *Escherichia coli*, and Kocabaş [20], studying L-tryptophan production, clarified that oxygen consumption consisted of three parts: Cell growth, maintenance, and byproduct formation. In order to enable model fitting of protein and biomass concentration, this study suggests modifying the Luedeking–Piret model in Equation (1) by adding parameter γ, which represents the oxygen consumption yield for protein synthesis rate $P'(t)$ ($\gamma \equiv Y_{o_2/p}$) [4,21]:

$$\mathrm{OUR}(t) = \alpha \cdot X'(t) + \beta \cdot X(t) + \gamma \cdot P'(t), \tag{2}$$

where the last term represents the oxygen update rate for product formation.

4. System Identification and Parameter Estimation

4.1. Stoichiometric Parameter Estimation

In a previous study [3], there was an assumption that stoichiometric parameter β, the oxygen maintenance term, was not a process constant, and one explanation was that it embraced the target protein P production:

$$\mathrm{OUR}(t) = \alpha \cdot X'(t) + \beta(X) \cdot X(t), \tag{3}$$

where the β function had the form

$$\beta(X) \equiv \beta(X(t)) = k_{\beta 2} \cdot X^2(t) + k_{\beta 1} \cdot X(t) + k_{\beta 0}. \tag{4}$$

Equation (3) gives acceptable results, but is highly uncertainty for the β term [3], which can be seen in Figure 1, where $\beta(X)$ is the maintenance coefficient as a function of biomass concentration and β(tm) is the maintenance coefficient observed at discrete time t_m and is associated with biomass X at time t_m. Graph data for Figure 1 originated from Reference [3].

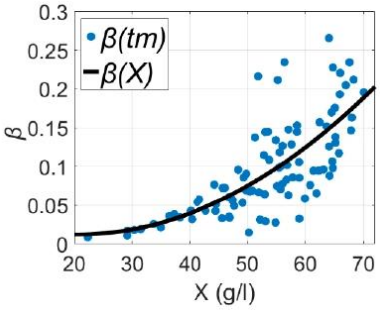


Figure 1. Dependence of oxygen consumption for maintenance on biomass concentration of *E. coli* estimated as a function of biomass and observed at discrete time t_m, taken from Reference [3].

In order to refine the model to a simpler and more versatile one, an additional parameter $\gamma(X)$ extends the parsimonious model [22] to the shape of Equation (5):

$$\mathrm{OUR(t)} = \alpha \cdot \mathrm{X}'(\mathrm{t}) + \beta(\mathrm{X}) \cdot \mathrm{X(t)} + \gamma(\mathrm{X}) \cdot \mathrm{P}'(\mathrm{t}). \tag{5}$$

This represents the main novelty of this study, protein production yield γ [4,21], which is assumed to be a function of biomass concentration X in a gray box model of Equation (3) [6,22]. The motivation of Equation (5) is that, through a convex programming procedure, the parameters with higher statistical significance overcome those with lower significance by leaving their entries populated with zero values. Babaeipour et al. [23] showed that protein productivity depends on IPTG and biomass concentrations at the induction time. In previous research [6], experiments had the same 1 mM amount of isopropyl-D-1-thiogalactopyranoside (IPTG). However, the biomass concentration at the induction time in each cultivation process was different. We found that it had a significant impact on the biomass model fitting. Our analysis showed that the product formation parameter $\gamma(X)$ is a function of biomass concentration at induction time [24]:

$$\gamma(\mathrm{X}) = \mathrm{k}_\gamma \cdot (\mathrm{X(t)} - \mathrm{X}_{\mathrm{ind}}), \tag{6}$$

where X_{ind} is biomass concentration at induction time and k_γ is the product synthesis yield, which is assumed to be constant. In summary, the full gray box model of the estimator has the form:

$$\begin{aligned} \mathrm{OUR(t)} = \alpha \cdot \mathrm{X}'(\mathrm{t}) + \mathrm{k}_\gamma \cdot (\mathrm{X(t)} - \mathrm{X}_{\mathrm{ind}}) \cdot \mathrm{P}'(\mathrm{t}) + \\ \left(\mathrm{k}_{\beta 2} \cdot \mathrm{X}^2(\mathrm{t}) + \mathrm{k}_{\beta 1} \cdot \mathrm{X(t)} + \mathrm{k}_{\beta 0}\right) \cdot \mathrm{X(t)}. \end{aligned} \tag{7}$$

In electrical systems, disturbances and interferences are inevitable, and the model's parameters and estimated values are distorted [11]. Urniezius et al. [3] and Schaepe et al. [13] showed that cumulative signals had less disturbance and an improved signal-to-noise ratio (SNR). In order for the original signal to be cumulative, this study employs an integral approach, which is a good noise filter [25]. After integration, the improved Luedeking–Piret model in Equation (7) becomes more resistant to state variable noise [26]:

$$\begin{aligned} \int_{t_0}^{t} \mathrm{OUR}(t^*) dt^* = \alpha \cdot \int_{t_0}^{t} \mathrm{X}'(t^*) dt^* + \mathrm{k}_\gamma \cdot \int_{t_0}^{t} (\mathrm{X}(t^*) - \mathrm{X}_{\mathrm{ind}}) \cdot \mathrm{P}'(t^*) dt^* + \\ \int_{t_0}^{t} \left(\mathrm{k}_{\beta 2} \cdot \mathrm{X}^2(t^*) + \mathrm{k}_{\beta 1} \cdot \mathrm{X}(t^*) + \mathrm{k}_{\beta 0}\right) \cdot \mathrm{X}(t^*) dt^*. \end{aligned} \tag{8}$$

After model analysis and calculations, the obtained results show that the stoichiometric parameter $\beta(X)$, the oxygen maintenance term for biomass concentration, is extremely low in comparison to other

stoichiometric parameters during the whole cultivation process. The convex estimation of coefficients $k_{\beta 2}$, $k_{\beta 1}$, $k_{\beta 0}$, manifested in this study, shows that all of those coefficients obtain zero values in this parsimonious model. The phenomenon where the biomass maintenance factor is absent from the growth process can be explained by the fact that the biomass concentration at the induction moment is relatively low (around 30 g/L) and the biomass maintenance term is negligible in this specific situation. A previous study [3] presented biomass maintenance model fitting procedures; therefore, Equation (8) considers only two terms of oxygen consumption:

$$\int_{t_0}^{t} \mathrm{OUR}(t^*)dt^* = \alpha \cdot \int_{t_0}^{t} X'(t^*)dt^* + k_\gamma \cdot \int_{t_0}^{t} (X(t^*) - X_{\mathrm{ind}}) \cdot P'(t^*)dt^*. \tag{9}$$

Integration with parts [27] of the last term in Equation (9) enables model fitting of biomass concentration. This helps to remove the protein production rate containing considerable uncertainty:

$$\begin{gathered} \int_{t_0}^{t} \mathrm{OUR}(t^*)dt^* = \alpha \cdot (X(t) - X(t_0)) + \\ k_\gamma \cdot \left(P(t) \cdot (X(t) - X_{\mathrm{ind}}) - \int_{t_0}^{t} (X(t^*) - X_{\mathrm{ind}})' \cdot P(t^*)dt^* \right), \end{gathered} \tag{10}$$

The differential of current biomass concentration minus biomass concentration at induction time simplifies to $(X(t^*) - X_{\mathrm{ind}})' = \frac{d(X(t^*) - X_{\mathrm{ind}})}{dt^*} = \frac{dX(t^*)}{dt^*}$. Therefore,

$$\begin{gathered} \int_{t_0}^{t} (X(t^*) - X_{\mathrm{ind}})' \cdot P(t^*)dt^* = \int_{t_0}^{t} \frac{dX(t^*)}{dt^*} \cdot P(t^*)dt^* = \\ \int_{t_0}^{t} P(t^*)dX(t^*) \approx \sum_{l=1}^{m} (X_l - X_{l-1}) \cdot P_l, \end{gathered} \tag{11}$$

where the last integral of Equation (11) represents the left-hand Riemann sum [28], when the time's t sample has an index of m. Discrete DCW samples define variable $X_m \equiv X(t)$, where $m \in [1, n_m]$; n_m is the total number (hours) of offline sampling intervals and $X_0 \equiv X(t_0)$ is an initial biomass concentration after inoculation in the bioreactor. Introducing $\mathrm{cOUR}_m \equiv \int_{t_0}^{t} \mathrm{OUR}(t^*)\, dt^*$ and Equation (11) into Equation (10) yields:

$$\mathrm{cOUR}_m = \alpha \cdot (X_m - X_0) + k_\gamma \cdot (P_m \cdot (X_m - X_{\mathrm{ind}}) - \sum_{l=1}^{m} (X_l - X_{l-1}) \cdot P_l). \tag{12}$$

The final formula for offline model fitting of biomass concentration is as follows:

$$X_m = \frac{\mathrm{cOUR}_m + \alpha \cdot X_0 + k_\gamma \cdot P_m \cdot X_{\mathrm{ind}} + k_\gamma \cdot \sum_{l=1}^{m} (X_l - X_{l-1}) \cdot P_l}{\alpha + P_m \cdot k_\gamma}. \tag{13}$$

Equation (13) also represents the prediction value of the proposed model, i.e., it serves as the constraint over the probabilistic mean $\langle X_m \rangle$.

4.2. Procedure for Offline Analysis of Stoichiometry Parameters

Fitting the biomass concentration parameters to the gray box model means that the analysis of offline bioprocess data evaluates the stoichiometric parameters of the cell strain. Equation (13) shows that the essential data must consist of dry cell weight (DCW) or an optical density (OD) value (o.u.), which is converted to DCW by multiplying it by a factor of 0.4 g/L/o.u. [29], cumulative oxygen uptake rate (cOUR), and insoluble target protein values. However, the time duration of the process since inoculation is not required during this gray box model fitting procedure.

The model for fitting parameter values is a gray box model, because collected experimental data are combined with fundamental knowledge about bioprocesses [30]. The posterior distribution for the m-th offline sample is:

$$P_{\mathrm{posterior}}(X_m) \sim N\left(\langle X_m \rangle, \sigma^2_{\langle X \rangle}\right), \tag{14}$$

where every sampled prediction m has a constant variance $\sigma^2_{\langle X \rangle}$. Prior distribution also has the form of Gaussian distribution [31,32]:

$$P_{likelihood}(X_m) \sim N\left(X^y_m, \sigma^2_{X,m}\right), \tag{15}$$

where X^y_m is the mth observation value of the biomass concentration and its individual variance is $\sigma^2_{X,m}$. Integration of relative entropy [31] yields:

$$\begin{aligned} S_m\left(P_{posterior}, P_{likelihood}\right) &= -\int_{-\infty}^{\infty} P_{posterior}(X_m) \cdot \ln \frac{P_{posterior}(X_m)}{P_{likelihood}(X_m)} dX_m \\ &= -\frac{\left(\langle X_m \rangle - X^y_m\right)^2}{2 \cdot \sigma^2_{X,m}} + c, \end{aligned} \tag{16}$$

where a further procedure neglects coefficient c. In a previous study [31], the uncertainty of prior distribution was set as equal to the squared observed value. However, Reference [3] showed that there are trade-offs between the least squares approach and the squared mean absolute percentage error (MAPE) criterion. A separate tuning coefficient K_{exp} [3] is required to adjust uncertainty:

$$\sigma^2_{X,m} \sim \frac{X^2_m}{1 - K_{exp} + X^2_m \cdot K_{exp}}, \tag{17}$$

which yields the sum of two criteria after insertion into Equation (16)

$$\begin{aligned} S_m\left(P_{posterior}, P_{likelihood}\right) &= -\frac{\left(\langle X_m \rangle - X^y_m\right)^2}{2 \cdot \frac{X^2_m}{1 - K_{exp} + X^2_m \cdot K_{exp}}}, \\ &= -\frac{\left(\langle X_m \rangle - X^y_m\right)^2 \cdot \left(1 - K_{exp}\right)}{2 \cdot X^2_m} - \frac{\left(\langle X_m \rangle - X^y_m\right)^2 \cdot K_{exp}}{2}. \end{aligned} \tag{18}$$

The tuning coefficient K_{exp} $\left(0 \le K_{exp} \le 1\right)$ with a value of 1 recreates the least squares approach, which has a higher penalty for bigger criterion values. Meanwhile, the value of zero results in the squared MAPE criterion [31], which restricts estimation errors to smaller overall criterion values. Such criteria showed acceptable practical benefits in a generic case of a biomass model fitting procedure. As a result, the least squares method is combined with the squared MAPE to apply the advantages of both criteria and top overcome their disadvantages, where K_{exp} is an empirical "weight" coefficient between the two additive terms of the optimization criterion.

4.3. Model of Product Model Fitting

Product evaluation technology is a complex soft sensor and is important for the biotechnology industry, demonstrating process efficiency and saving time in protein measurements [9]. In this study, the basic idea of the protein model fitting comes from Levisauskas' research [33], claiming that relative protein synthesis is a function of the specific biomass growth rate:

$$\frac{dP_x}{dx} = q_{px}(\mu, P_x), \tag{19}$$

where q_{px} is a specific protein accumulation rate (U/g/h), μ is a specific biomass growth rate (1/h), and P_x is specific protein activity (U/g), where protein concentration is divided by biomass concentration, $P_x(t) = P(t)/X(t)$ [33]. Data analysis and studies have shown that production synthesis is linearly dependent on the specific growth rate (SGR) of the biomass and the product concentration acts as an inhibitor of product synthesis [34]:

$$\frac{dP_x}{dx} = q_{px}(\mu, P_x) = P_{max}(\mu, X) - k_t \cdot P_X(t). \tag{20}$$

In Equation (16), coefficient k_t is a corresponding time constant that is assumed to be a form of the self-inhibition effect [35]. P_{max} is a maximal specific product concentration value, which is asymptotically dependent on μ. The specific biomass growth rate and biomass concentration determine the maximum specific product concentration term [36], which represents the maximum possible protein concentration in the current process state:

$$P_{max}(\mu, X) = \mu(t) \cdot (k_{m0} + k_{m1} \cdot (X(t) - X_{ind})), \tag{21}$$

where k_{m0} and k_{m1} are empirical coefficients proposed by this study, k_{m0} relates to SGR and protein synthesis, and k_{m1} links the biomass concentration at the induction time and productivity [23]. Equation (21) is only valid after induction and biomass concentration at induction time is a prerequisite. Before IPTG injection into the bioreactor, coefficient k_{m1} is equal to zero and the maximum specific product concentration term becomes:

$$P_{max}(\mu, X) = \mu(t) \cdot k_{m0}. \tag{22}$$

The protein model fitting is comparable to the gray box model and the biomass concentration model. Prior to performing the coefficient evaluation, the gray box model, represented by Equation (20), integrates to:

$$P_X(t) = \int_{t_0}^{t} P_{max}(t^*) dt^* - k_t \cdot \int_{t_0}^{t} P_X(t^*) dt^* \tag{23}$$

The integrals of Equation (23) are expressed as the left-hand Riemann sum [28], i.e., $\int_{t_0}^{t} P_{max}(t^*) dt^* \approx \sum_{j=1}^{m} P_{max,j} \cdot \Delta t_{j,j-1}$; $\int_{t_0}^{t} P_X(t^*) dt^* \approx P_{X,m} \cdot \Delta t_{m,\,m-1} + \sum_{j=1}^{m-1} P_{X,j} \cdot \Delta t_{j,j-1}$; when time's t sample is indexed by m, discrete protein values define the variable $P_{X,m} \equiv P_X(t)$, where $m \in [1, n_m]$. The final formula of protein model fitting is as follows:

$$P_m = \frac{\left(\sum_{j=i}^{m} P_{max,j} \cdot \Delta t_{j,j-1} - k_t \cdot \sum_{j=1}^{m-1} P_{x,j} \cdot \Delta t_{j,j-1}\right) \cdot X_m}{1 + \Delta t_{m,m-1} \cdot k_t}. \tag{24}$$

Model fitting uses Equation (24) for a prediction value $\langle P_m \rangle$ and observed product concentrations P_m^y inside convex optimization.

4.4. Identification of E. Coli Parameters by Convex Optimization

The process of identifying *E. coli* BL21 (DE3) pET28a strain's stoichiometric parameters and protein model fitting coefficients is based on the convex optimization method and the maximization of entropy [31,37]. Figure 2 depicts the workflow of the optimization procedure.

Convex optimization uses the maximization of entropy as an indicator of local extremum detections [38]. Equation (18) helps with identification of stoichiometry parameters and Equation (25) does the same for the product model fitting:

$$S_{P,m} = -\frac{\left(\langle P_m \rangle - P_m^y\right)^2 \cdot \left(1 - K_{exp}\right)}{2 \cdot X_m^2} - \frac{\left(\langle P_m \rangle - P_m^y\right)^2 \cdot K_{exp}}{2} \tag{25}$$

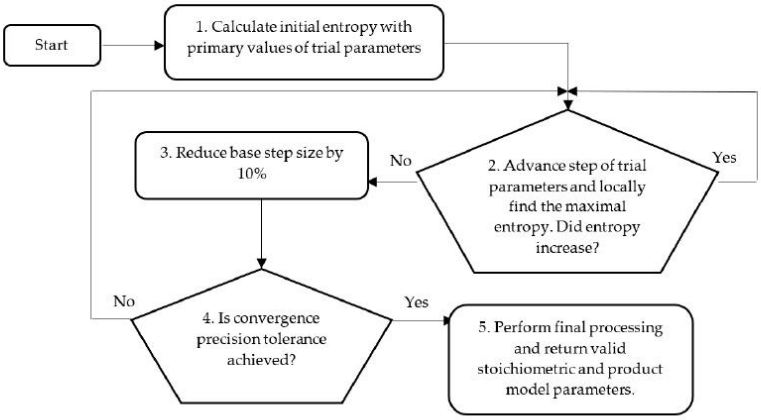


Figure 2. Workflow of structural scheme for convex optimization method identifying stoichiometric and product model fitting parameters.

5. Experimental Validation

For the comparison of results, the mean absolute error (MAE) and mean absolute percentage error (MAPE) are operated as evaluation criteria. The definition of MAE is [39]:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|}{n}, \tag{26}$$

where n is the number of data counts, $\hat{y}_i$ is the estimation result, and y_i is the observed value from the cultivation process. MAPE has the expression [40]:

$$\mathrm{MAPE} = \frac{100\,\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|. \tag{27}$$

Root mean square error (RMSE) represents the square root of the residuals of the differences between predicted and observed values. The formula is as follows [39]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n}}. \tag{28}$$

The experimental data of the fed-batch cultivation process of *Escherichia coli* were taken from Reference [6]. In order to test and validate the proposed models of this paper, data from 19 cultivation experiments were used in the system identification analysis. The start of this research included investigating a suitable expression describing stoichiometry parameters in biomass model fitting. Multiple tests employed various formulations, including previous assumptions on polynomial maintenance [3]. The purpose was to indicate the most suitable formula that describes cell stoichiometry. Table 1 describes the best-achieved coefficient values for the fitted model.

Table 1. Analysis results of biomass concentration models. MAE, mean absolute error; MAPE, mean absolute percentage error.

Model	α	$k_{\beta 0}$	$k_{\beta 1}$	$k_{\beta 2}$	k_{γ}	MAE	MAPE
Equation (3)	0.996	0.07	0.00084	0	—	1.422	8.85%
Equation (12)	0.997	0	0	0	2.705	0.68	6.92%

The MAE and MAPE values show the average from 19 experiments. The results of protein model fitting Equation (24) are presented in Table 2.

Table 2. Values of protein model parameters according to Equation (24).

E. Coli BL21 (DE3) pET28a
$k_{m0} = 0.2346$
$k_{m1} = -0.0172$
$k_t = 0.0687$

Table 2 presents the model parameters that produce the protein estimation results of this study. These parameters are only suitable for the genetically modified *E. coli* BL21 (DE3) pET28a cell strain investigated in this study and is mentioned in the Materials and Methods Section. Equation (24) mainly describes the recurrent procedure of offline estimation. Protein estimates were determined from previous protein estimates and offline biomass measurements. First, parameters k_{m0} and k_{m1} were used for determination of $P_{max,j}$ in Equation (21). This equation also used an approximate value of SGR, $\mu_j \cong \left(X_j - X_{j-1}\right)/\left(X_j \cdot \Delta t_{j,j-1}\right)$. Equation (24) was only dependent on offline biomass observations in this study, or online biomass estimates in future applications. After calculating the protein value using Equation (24), the "normalized" protein value $P_{x,j} = P_j/X_j$ served as input for the estimation of the next target protein value by Equation (24). In this way, model fitting used the equation in a recursive manner and had no dependency on target product related state variables.

Protein and biomass model fitting results are presented in Table 3 using the best-fit configurations of models parameters.

Table 3. Analysis of biomass and product concentration models. RMSE, root mean square error.

	Dry Biomass Concentration (Dry Cell Weight, DCW)			Product		
No.	**MAE (g/L)**	**MAPE (%)**	**RMSE (g)**	**MAE (g/L)**	**MAPE (%)**	**RMSE (g)**
1	0.728	6.802	5.212	0.139	5.378	0.571
2	0.762	4.997	6.621	0.231	6.095	0.647
3	0.860	11.022	6.172	0.473	52.526	2.91
4	0.388	4.458	3.085	0.184	13.265	1.248
5	0.798	8.02	6.107	0.527	82.075	3.258
6	0.512	8.82	3.703	0.113	6.7898	0.608
7	0.595	4.787	4.605	0.127	6.957	0.84
8	0.311	4.433	2.191	0.629	35.36	3.757
9	0.576	6.046	4.266	0.178	11.250	1.471
10	0.873	9.017	6.166	0.634	33.844	4.147
11	0.582	5.248	4.468	0.1407	8.286	0.872
12	0.61	5.884	5.264	0.31	19.407	1.946
13	0.7642	5.477	4.962	0.318	39.614	1.834
14	0.404	3.862	3.563	0.056	7.001	0.594
15	0.531	5.724	3.726	0.137	9.681	0.914
16	0.628	7.532	4.503	0.066	4.504	0.401
17	0.86	7.057	6.685	0.16	17.13	1.042
18	1.262	11.767	9.218	0.134	10.328	1.026
19	0.862	10.582	5.933	0.111	8.15	0.738

Therefore, the average MAE of biomass model fitting since the start of the bioprocess of inoculation is 0.679 g/L and that of product model fitting is 0.246 g/L. The overall average MAPE of biomass model fitting since the start of inoculation is 6.92% and that of product model fitting is 19.87%. The overall average RMSE of biomass model fitting since the start of inoculation is 5.07 g and that of product model fitting is 1.517 g. The MAPE, MAE, and RMSE of the product model fitting neglects the very first measurement after induction, since it has less meaning for MAPE when product synthesis starts.

To validate the identified model parameters shown in Table 2, data from six cultivation experiments of the same cell culture were processed.

According to the validation data shown in Table 4, the average MAE of biomass since the start of inoculation is 0.636 g/L and that of product is 0.099 g/L. The overall average MAPE of biomass since the start of inoculation is 7.09% and that of product is 8.22%. The overall average RMSE of biomass since the start of inoculation is 4.577% and that of product is 0.656%.

Table 4. Model validation results.

	Dry Biomass Concentration (DCW)			Product		
No.	**MAE (g/L)**	**MAPE (%)**	**RMSE (g)**	**MAE (g/L)**	**MAPE (%)**	**RMSE (g)**
1	0.769	8.594	5.279	0.128	11.947	0.7222
2	0.481	7.39	2.916	0.0813	6.565	0.491
3	0.843	8.107	6.354	0.0563	7.86	0.397
4	0.727	5.25	5.975	0.05	4.996	0.323
5	0.596	7.199	4.17	0.134	8.715	0.821
6	0.402	6.033	2.768	0.149	9.26	1.185

Figure 3 portrays some typical biomass model fitting results and Figure 4 shows biomass validation results. These results show that estimation approaches for biomass concentration and product attained acceptable precision without compromising the simplicity of implementation. The proposed models show a simplistic structure while being accurate and a basis of fundamental knowledge. The main purpose of this paper is to show evidence that biomass and protein model fitting can be handled from the fundamental point of view based on stoichiometry Equation (1) and protein synthesis Equation (19), without the need for an artificial neural network (ANN) or other hybrid black box systems requiring data training [6,41–43]. Training procedures normally require huge amounts of training data, while this study proposes an approach that helps with the identification of the parameters once per strain. For comparison, the results of ANN and the model in this paper are compared in Table 5.

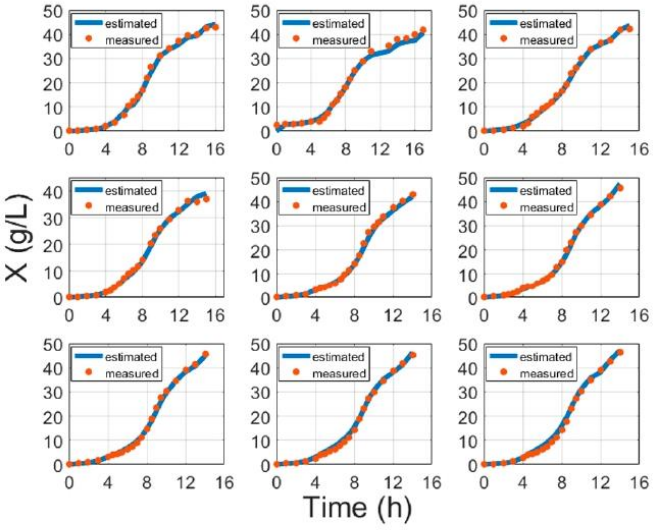


Figure 3. Biomass model fitting results with cultivation processes data, where time is the cultivation time since inoculation in the bioreactor.

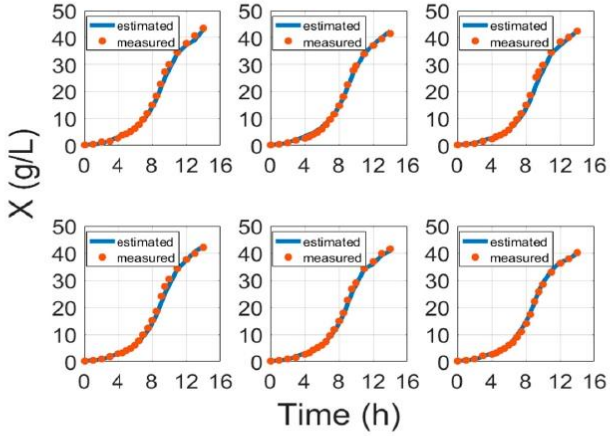


Figure 4. Biomass validation results with cultivation processes data, where time is the cultivation time since inoculation in the bioreactor.

Table 5. Comparison of prediction quality of the model in this paper and Gnoth et al. [6] model.

	RMSE (g)		
	Total Biomass	**Total Soluble Protein**	**Total Insoluble Protein**
Conventional model from Gnoth et al. [6]	10.81	1.78	0.87
Hybrid network from Gnoth et al. [6]	4.71	1.28	0.62
Model in this study	4.577	-	0.656

Moreover, instead of induction time [6], this study suggests using biomass concentration at induction, which better confirms conventional bioprocess development practices. The results of protein model fitting are shown in Figure 5 and are validated in Figure 6.

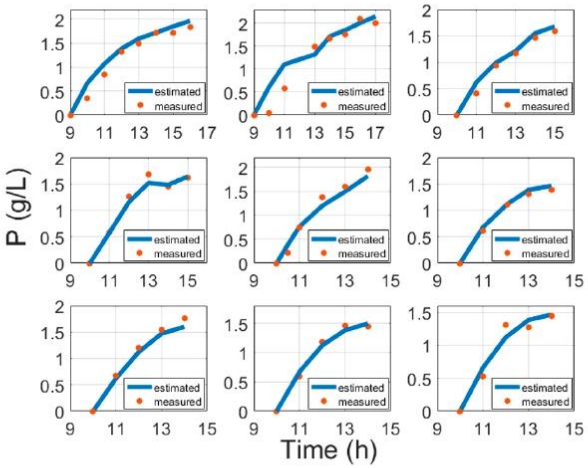


Figure 5. Protein model fitting results compared with cultivation experiment data, where time is the cultivation time since inoculation in the bioreactor.

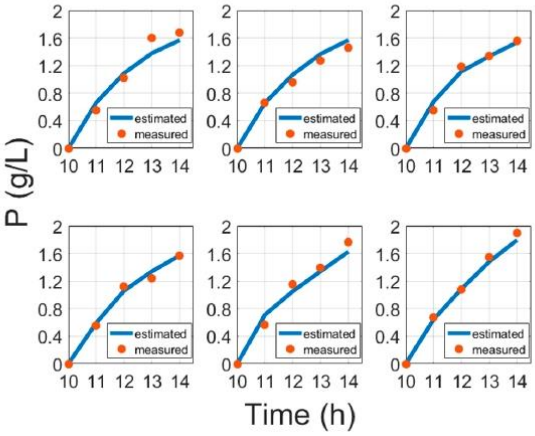


Figure 6. Protein validation results compared with cultivation experiment data, where time is the cultivation time since inoculation in the bioreactor.

6. Conclusions

This paper suggests two functional models for biomass and product concentration, which are crucial for the later development of online product and biomass estimators. The biomass model fitting approach uses the stoichiometry model proposed by Luedeking and Piret in 1959. This study assumed that the estimation routines are dependent on stoichiometry parameters of the strain and the biomass concentration at the time of induction. The proposed model fitting method utilizes only few inputs: Specific biomass growth rate and biomass concentration at time of induction. The approach is thus based on fundamental knowledge about biosynthesis. Analysis of process data from 19 cultivation experiments validated the routines. Evaluation errors confirmed that the approach is relevant for model fitting of the *Escherichia coli* BL21 (DE3) pET28a cell strain. The overall average MAE of biomass model fitting was 0.679 g/L and that of product model fitting was 0.246 g/L. The overall average MAPE of biomass model fitting was 6.92% and that of product model fitting was 19.87%. The suggested approach does not depend on any numeric initial optimization conditions and does not require any bioreactor parameters. The proposed approach has certain benefits compared to artificial neural networks. Training procedures normally require a huge amount of training data, while this study proposes an approach that helps with the identification (training) of parameters once per strain. This study suggests using a more general biomass concentration at induction, normally defined in contract or biotechnological protocols.

Author Contributions: Both of the authors contributed to the preparation of the manuscript and have read and approved the final manuscript. Conceptualization, R.U., A.S.; Methodology, R.U.; Software, R.U., A.S.; Validation, R.U., A.S.; Formal Analysis, R.U., A.S.; Investigation, R.U.; Resources, R.U.; Data Curation, A.S.; Writing—Original Draft Preparation, R.U., A.S.; Writing—Review and Editing, R.U.; Visualization, A.S.; Supervision, R.U.; Project Administration, R.U.; Funding Acquisition, R.U.

Funding: 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.

Acknowledgments: We are grateful to Rimvydas Simutis, Kaunas University of Technology, for kind encouragement to dig into model inference, which would have functional bioprocess meaning.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodwin, G.C. Predicting the performance of soft sensors as a route to low cost automation. *Annu. Rev. Control* **2000**, *24*, 55–66. [CrossRef]
2. Mansano, R.; Godoy, E.; Porto, A. The Benefits of Soft Sensor and Multi-Rate Control for the Implementation of Wireless Networked Control Systems. *Sensors* **2014**, *14*, 24441–24461. [CrossRef]
3. Urniezius, R.; Survyla, A.; Paulauskas, D.; Bumelis, V.A.; Galvanauskas, V. Generic estimator of biomass concentration for *Escherichia coli* and *Saccharomyces cerevisiae* fed-batch cultures based on cumulative oxygen consumption rate. *Microb. Cell Fact.* **2019**, *18*, 190. [CrossRef]
4. Garcia-Ochoa, F.; Gomez, E.; Santos, V.E.; Merchuk, J.C. Oxygen uptake rate in microbial processes: An overview. *Biochem. Eng. J.* **2010**, *49*, 289–307. [CrossRef]
5. Luedeking, R.; Piret, E.L. A kinetic study of the lactic acid fermentation. Batch process at controlled pH. *Biotechnol. Bioeng.* **1959**, *1*, 393–412. [CrossRef]
6. Gnoth, S.; Simutis, R.; Lübbert, A. Selective expression of the soluble product fraction in *Escherichia coli* cultures employed in recombinant protein production processes. *Appl. Microbiol. Biotechnol.* **2010**, *87*, 2047–2058. [CrossRef] [PubMed]
7. Jenzsch, M.; Simutis, R.; Luebbert, A. Generic model control of the specific growth rate in recombinant *Escherichia coli* cultivations. *J. Biotechnol.* **2006**, *122*, 483–493. [CrossRef] [PubMed]
8. Schaepe, S.; Kuprijanov, A.; Simutis, R.; Lübbert, A. Avoiding overfeeding in high cell density fed-batch cultures of *E. coli* during the production of heterologous proteins. *J. Biotechnol.* **2014**, *192*, 146–153. [CrossRef]
9. Kuprijanov, A.; Gnoth, S.; Simutis, R.; Lübbert, A. Advanced control of dissolved oxygen concentration in fed batch cultures during recombinant protein production. *Appl. Microbiol. Biotechnol.* **2009**, *82*, 221–229. [CrossRef]
10. Gnoth, S.; Kuprijanov, A.; Simutis, R.; Lübbert, A. Simple adaptive pH control in bioreactors using gain-scheduling methods. *Appl. Microbiol. Biotechnol.* **2010**, *85*, 955–964. [CrossRef]
11. Linko, P.; Zhu, Y. Neural network programming in bioprocess variable estimation and state prediction. *J. Biotechnol.* **1991**, *21*, 253–269. [CrossRef]
12. Simutis, R.; Lübbert, A. Bioreactor control improves bioprocess performance. *Biotechnol. J.* **2015**, *10*, 1115–1130. [CrossRef] [PubMed]
13. Schaepe, S.; Kuprijanov, A.; Sieblist, C.; Jenzsch, M.; Simutis, R.; Lübbert, A. Current Advances in Tools Improving Bioreactor Performance. *Curr. Biotechnol.* **2013**, *3*, 133–144. [CrossRef]
14. Lübbert, A.; Bay Jørgensen, S. Bioreactor performance: A more scientific approach for practice. *J. Biotechnol.* **2001**, *85*, 187–212. [CrossRef]
15. Simutis, R.; Galvanauskas, V.; Levisauskas, D.; Repsyte, J.; Vaitkus, V. Comparative Study of Intelligent Soft-Sensors for Bioprocess State Estimation. *JOLST* **2013**, 163–167. [CrossRef]
16. Wechselberger, P.; Sagmeister, P.; Herwig, C. Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. *Bioprocess Biosyst. Eng.* **2013**, *36*, 1205–1218. [CrossRef] [PubMed]
17. Gnoth, S.; Jenzsch, M.; Simutis, R.; Lübbert, A. Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. *J. Biotechnol.* **2007**, *132*, 180–186. [CrossRef]
18. Caramihai, M.; Severi, I. Bioprocess Modeling and Control. In *Biomass Now—Sustainable Growth and Use*; Matovic, M.D., Ed.; InTech: Vienna, Austria, 2013; ISBN 978-953-51-1105-4.
19. Çalik, P.; Yilgör, P.; Demir, A.S. Influence of controlled-pH and uncontrolled-pH operations on recombinant benzaldehyde lyase production by *Escherichia coli*. *Enzyme Microb. Technol.* **2006**, *38*, 617–627. [CrossRef]
20. Kocabaş, P.; Çalık, P.; Özdamar, T.H. Fermentation characteristics of l-tryptophan production by thermoacidophilic *Bacillus acidocaldarius* in a defined medium. *Enzyme Microb. Technol.* **2006**, *39*, 1077–1088. [CrossRef]
21. Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* **2009**, *18*, 936–948. [CrossRef]
22. Zymnis, A.; Boyd, S.; Gorinevsky, D. Mixed linear system estimation and identification. *Signal Process.* **2010**, *90*, 966–971. [CrossRef]

23. Babaeipour, V.; Shojaosadati, S.A.; Maghsoudi, N. Maximizing Production of Human Interferon-γ in HCDC of Recombinant *E. coli*. *Iran. J. Pharm Res.* **2013**, *12*, 563–572. [PubMed]
24. Jenzsch, M.; Gnoth, S.; Kleinschmidt, M.; Simutis, R.; Lübbert, A. Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production. *J. Biotechnol.* **2007**, *128*, 858–867. [CrossRef] [PubMed]
25. Jenzsch, M.; Simutis, R.; Eisbrenner, G.; Stückrath, I.; Lübbert, A. Estimation of biomass concentrations in fermentation processes for recombinant protein production. *Bioprocess Biosyst. Eng.* **2006**, *29*, 19–27. [CrossRef]
26. Petkov, S.B.; Davis, R.A. On-line biomass estimation using a modified oxygen utilization rate. *Bioprocess Eng.* **1996**, *15*, 43–45. [CrossRef]
27. Brand, L. *TotalBoox. TBX Advanced Calculus*; Dover Publications: Mineola, NY, USA, 2013; ISBN 978-0-486-15799-3.
28. Swokowski, E.W. *Calculus with Analytic Geometry*, 2nd ed.; Prindle, Weber & Schmidt: Boston, MA, USA, 1979; ISBN 978-0-87150-268-1.
29. Shiloach, J.; Fass, R. Growing E. coli to high cell density—A historical perspective on method development. *Biotechnol. Adv.* **2005**, *23*, 345–357. [CrossRef]
30. Bohlin, T. *Practical Grey-Box Process Identification: Theory and Applications*; Advances in Industrial Control; Springer: London, UK, 2006; ISBN 978-1-84628-402-1.
31. Urniezius, R.; Galvanauskas, V.; Survyla, A.; Simutis, R.; Levisauskas, D. From Physics to Bioengineering: Microbial Cultivation Process Design and Feeding Rate Control Based on Relative Entropy Using Nuisance Time. *Entropy* **2018**, *20*, 779. [CrossRef]
32. Giffin, A.; Urniezius, R. The Kalman Filter Revisited Using Maximum Relative Entropy. *Entropy* **2014**, *16*, 1047–1069. [CrossRef]
33. Levisauskas, D.; Galvanauskas, V.; Henrich, S.; Wilhelm, K.; Volk, N.; Lübbert, A. Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. *Bioprocess Biosyst. Eng.* **2003**, *25*, 255–262. [CrossRef]
34. Galvanauskas, V.; Volk, N.; Simutis, R.; Lübbert, A. Design of recombinant protein production processes. *Chem. Eng. Commun.* **2004**, *191*, 732–748. [CrossRef]
35. Miao, F.; Kompala, D.S. Overexpression of cloned genes using recombinant Escherichia coli regulated by a T7 promoter: I. Batch cultures and kinetic modeling. *Biotechnol. Bioeng.* **1992**, *40*, 787–796. [CrossRef] [PubMed]
36. Schuler, M.M.; Marison, I.W. Real-time monitoring and control of microbial bioprocesses with focus on the specific growth rate: Current state and perspectives. *Appl. Microbiol. Biotechnol.* **2012**, *94*, 1469–1482. [CrossRef] [PubMed]
37. Urniezius, R. Convex programming for semi-globally optimal resource allocation. In *AIP Conference Proceedings*; AIP Publishing: Beirut, Lebanon, 2016; p. 040002.
38. Giffin, A.; Urniezius, R. Simultaneous State and Parameter Estimation Using Maximum Relative Entropy with Nonhomogenous Differential Equation Constraints. *Entropy* **2014**, *16*, 4974–4991. [CrossRef]
39. Willmott, C.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
40. de Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. Mean Absolute Percentage Error for regression models. *Neurocomputing* **2016**, *192*, 38–48. [CrossRef]
41. Gnoth, S.; Jenzsch, M.; Simutis, R.; Lübbert, A. Control of cultivation processes for recombinant protein production: A review. *Bioprocess Biosyst. Eng.* **2008**, *31*, 21–39. [CrossRef]
42. Galvanauskas, V.; Simutis, R.; Lübbert, A. Hybrid process models for process optimisation, monitoring and control. *Bioprocess Biosyst. Eng.* **2004**, *26*, 393–400. [CrossRef]
43. Gnoth, S.; Jenzsch, M.; Simutis, R.; Lübbert, A. Product formation kinetics in genetically modified E. coli bacteria: Inclusion body formation. *Bioprocess Biosyst. Eng.* **2008**, *31*, 41–46. [CrossRef]

Computational and Structural Biotechnology Journal 19 (2021) 5856–5863

ELSEVIER

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

An oxygen-uptake-rate-based estimator of the specific growth rate in *Escherichia coli* BL21 strains cultivation processes

Arnas Survyla [a], Donatas Levisauskas [a], Renaldas Urniezius [a,*], Rimvydas Simutis [a]

[a] Department of Automation, Kaunas University of Technology, Studentu 48, LT-51367 Kaunas, Lithuania

ARTICLE INFO

Article history:
Received 15 July 2021
Received in revised form 5 October 2021
Accepted 9 October 2021
Available online 19 October 2021

Keywords:
Cultivation process
Oxygen uptake rate
Specific growth rate
Real-time simulation
In-process monitoring

ABSTRACT

The cell cultivation process in a bioreactor is a high-value manufacturing process that requires excessive monitoring and control compatibility. The specific cell growth rate is a crucial parameter that describes the online quality of the cultivation process. Most methods and algorithms developed for online estimations of the specific growth rate controls in batch and fed-batch microbial cultivation processes rely on biomass growth models. In this paper, we present a soft sensor – a specific growth rate estimator that does not require a particular bioprocess model. The approach for online estimation of the specific growth rate is based on an online measurement of the oxygen uptake rate. The feasibility of the estimator developed in this study was determined in two ways. First, we used numerical simulations on a virtual platform, where the cell culture processes were theoretically modeled. Next, we performed experimental validation based on laboratory-scale (7, 12, 15 L) bioreactor experiments with three different Escherichia coli BL21 cell strains.

1. Introduction

Currently, the production of therapeutic proteins, drugs, and vaccines to treat diseases has been carried out in large-scale industrial bioreactors [1]. The cultivation processes in large-scale reactors are high-value manufacturing processes in which failure is intolerable. Moreover, the efficiency of these processes must be high and not compromised by their control simplicity. Monitoring and control algorithms with feedback signals are necessary to reduce errors and increase the efficiency of biotechnological processes [2]. The real-time monitoring and controlling tools for maintaining production processes within certain boundaries [3] will eventually become mandatory in upstream and downstream development, scale-up and scale-down reiterations, and contract development and manufacturing organization technology transfer services.

In microbial cultivation processes, specifically with recombinant *Escherichia coli*, one essential procedure is to monitor and control the growth characteristics of the culture. The specific growth rate (SGR) is an essential cultivation process variable because it represents a characteristic of the physiological state of the cell culture. The SGR is also related to the biosynthesis of the target product [4,5]. In addition, the quality of the desired product and the entire cultivation process can be defined by the specific growth rate of the biomass [6–8]. The SGR value can be obtained in two ways. The first method for calculating the SGR was based on the rate of change in the dry biomass samples. This procedure can take several hours or days. Hence, this method cannot be used as feedback for a control system. The second method to acquire the SGR is to use soft sensors, that is, estimating the SGR value by using other measurable online parameters such as the oxygen uptake rate. Such a calculation approach of SGR provides real-time values that serve as feedback to the control system.

This study explores the development of a specific growth rate estimator based on the stoichiometric parameters (more specifically, on the single ratio of those parameters) of the cell culture and oxygen uptake rate signal. Kinetics information does not vary and does not depend on the environment or other growth process conditions. Thus, constant stoichiometric parameters serve as inputs for the SGR estimator. Furthermore, the constant coefficients α define the oxygen demand for biomass growth, and the maintenance term β relates to the oxygen consumption by the biomass. Consequently, the off-gas analyzer's oxygen uptake rate (OUR) signal is beneficial for determining the specific growth rate.

Section 2 reviews the literature related to this study. Section 3 describes the materials, strains, and operating conditions of the bioreactor system. Section 4 outlines the developmental path of

* Corresponding author.
E-mail address: renaldas.urniezius@ktu.lt (R. Urniezius).

https://doi.org/10.1016/j.csbj.2021.10.015

the SGR estimation algorithm. Section 5 provides an investigation of the SGR estimator's performance, detailing the estimator's advantages, and estimation results. The final section, Conclusions, discusses the results, and provides the final statements of this study.

2. Related work

Many important cultivation process variables, such as SGR and biomass concentrations, cannot be directly measured in real time because biotechnology processes have complex relationships between the processes and variables. The best way to express unmeasurable parameters in real time is to use appropriate soft sensors/estimators [9].

One of the attractive ways to estimate SGR is the direct use of biomass concentration measurements. However, this approach faces difficulties in online measurements of biomass concentration, which is a challenging state variable to measure accurately in a liquid culture [10] noninvasively when various cultivation conditions are to be tested online. This is particularly true for non-stationary processes at the upstream bioprocess development stage. A dielectric spectrometer was used to estimate the biomass concentration and implement an observer-based estimator of the SGR [11]. However, the developed estimation algorithm requires an accurate tuning of the estimator parameters. Moreover, oscillations and instability in estimator performance occur at low biomass concentrations. An SGR estimation approach using biomass concentration measurements obtained through dielectric spectroscopy was presented in [12]. However, the calculated SGR values suffered from biomass measurement uncertainty, which could be reduced by increasing the observation window. However, a large observation window increased the SGR signal delay.

Because online analyzers of biomass concentration are often unavailable or not sufficiently reliable, the SGR needs to be estimated through directly measurable variables, such as the substrate consumption, oxygen uptake rate, carbon dioxide production rate, and base consumption rate [7,10]. For example, the successful implementation of unscented Kalman filters (UKF) combined with an artificial neural network for estimating the SGR based on cumulative oxygen consumption and carbon dioxide production measurements was reported by Simutis and Lübbert [13]. An advanced Kalman filter (EKF) is also suitable for SGR and biomass concentration estimations, where the oxygen uptake rate is one of the input signals [14,15].

Rocha et al. [6] presented a biomass observer that involved the development of an SGR estimator for the fed-batch bioprocess of recombinant *E. coli*, for which online measurements of the dissolved oxygen, oxygen transfer rate, and culture weights were used. The observer and estimator algorithms are based on the asymptotic observer approach, a mathematical model, and the assumption that the model parameters are known. The development of a complex SGR estimation algorithm requires specific knowledge and is a time-consuming task. For example, SGR estimation, data-driven models such as artificial neural networks (ANNs), and hybrid models can be employed, especially in industrial processes. A large amount of data can be used to train and validate ANN-based models [16,17]. However, the ANN and hybrid model approaches entail considerable performance trade-offs and design costs to select proper experimental data and training for ANNs. In addition, each ANN-based estimator applies only to a specific cultivation process. Therefore, the complex approaches that result in complex algorithms are not attractive for developing robust SGR estimators for industrial applications.

In this study, we developed a robust and straightforward estimator of cell biomass SGR in batch and fed-batch cultivation processes based on online estimates of the oxygen uptake rate. The algorithm is simple because it requires only two inputs: the OUR and a tuning parameter that uses the stoichiometric parameter ratio. The reliability and simplicity of the SGR estimator make it easy to implement it into the control system as feedback (1).

3. Materials and methods

3.1. Medium and culture conditions

In this work, due to data availability, three types of *Escherichia coli* cell-strain cultivation data were studied to verify the SGR estimates and determine their reliability and versatility. The *E. coli* BL21(DE3) pET9a-IdeS, *E. coli* BL21 (DE3) pET21-IFN-alfa-5, and *E. coli* BL21(DE3) pLysS were chosen as the study subjects. All three cell strains were cultivated in several independent R&D laboratories.

The cell strain of *E. coli* BL21 (DE3) pET21-IFN-alfa-5 was cultivated in a 7 L bioreactor. The cultivation medium featured minimal mineral concentrations, including 46.55 g potassium dihydrogen phosphate, 14 g ammonium phosphate dibasic, 5.6 g citric acid monohydrate, 3 ml of concentrated antifoam, 35 g magnesium sulfate heptahydrate, and 105 g D (+) glucose monohydrate. The initial weight of the medium was 3.7 kg. The environmental parameters of the cultivation process remained constant throughout the experiment. The temperature was set to 37 °C, the DOT was set to 20% of air saturation, and the pH was maintained at pH 6.8 through the addition of NaOH(aq). The stirrer speed ranged from 800 to 1200 rpm. The airflow scope ranged from 1.75 to 3.75 L/min. During the cultivation process, pure oxygen flow from 0 to 7.5 L/min was used to increase the oxygen transfer rate in the bioreactor.

E. coli BL21 (DE3) pET9a-IdeS cell strain was cultivated in a 15 L bioreactor. The cultivation medium was introduced according to the minimum requirements of a mineral medium. During the cultivation process, the environmental parameters were as follows: temperature, 37 °C; DOT, 30% of air saturation; and pH maintained at 6.98 via the addition of NaOH(aq). The stirrer speed ranged from 300 to 750 rpm. The operating airflow range ranged from 0.3 to 15 L/min. Pure oxygen flow was provided to the bioreactor during the cultivation process from 0 to 7.5 L/min. During cultivation of the *E. coli* BL21 (DE3) pET21-IFN-alfa-5 and *E. coli* BL21 (DE3) pET9a-IdeS cell strains. During the *E. coli* BL21 (DE3) pET21-IFN-alfa-5 and *E. coli* BL21 (DE3) pET9a-IdeS cell cultivation processes, the oxygen concentration in the off-gas from the bioreactor was measured online using a BlueSens BlueInOne Ferm gas analyzer, which had a measuring range from 0 to 100%.

The *E. coli* (BL21(DE3) pLysS) cell strain was cultivated in minimal mineral medium. This medium was composed of (NH4)2SO4,

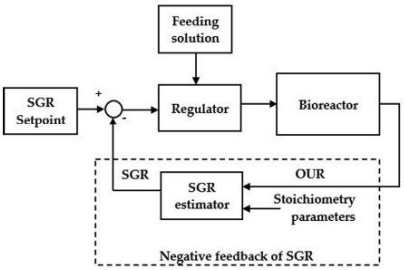


Fig. 1. Principal scheme of the SGR control system.

2.46 g/L; NH4Cl, 0.5 g/L; NaH2PO4 × H2O, 3.6 g/L; Na2SO4, 2 g/L; K2HPO4, 14.6 g/L; (NH4)2-citrate, 1 g/L; 1 M MgSO4 solution, 5 mL/L; trace element solution, 2 mL/L; and no glucose. The initial mass of all cultivations was 5 kg. The pH was kept constant at pH 7, and the temperature was set to 30 °C. DO was set to 30% saturation. The bioreactor had a working volume of 15 L (Biostat C, Sartorius Stedim Biotech), and the stirrer speed varied from 100 to 1400 rpm. The oxygen uptake rate, OUR, was measured online with a paramagnetic oxygen sensor placed in the reactor's vent line behind the offgas cooler (Sidor, Sick-Maihak, Hamburg).

4. Development of the SGR estimation algorithm

Cells are living organisms that breathe and consume food (glucose), so respiratory data can express the state of the cell culture in the bioreactor. The higher the biomass content, the more evident is the respiration data. The main parameters of respiratory data are the oxygen uptake rate (OUR) and carbon dioxide production rate (CPR). In this study, the algorithm for the online estimation of SGR during microbial cultivation processes relies upon reasonable estimates of the OUR, as the OUR signal is less sensitive to cellular metabolism and other negative cell growth phenomena than the CPR signal. The oxygen uptake rate can be calculated online from the difference between the oxygen concentration entering the bioreactor and the oxygen concentration leaving the bioreactor [18,19]:

$$OUR(t) = Q \cdot (O_2^{in} - O_2^{out}); \tag{1}$$

where O_2^{in} and O_2^{out} are the oxygen concentrations at the inlet and outlet gas streams, and Q is the gas flow rate. The relationship between the OUR and biomass growth in microbial cultures can be modeled using Luedeking/Piret-type relationships [20,21]:

$$OUR(t) = \alpha \cdot Xn(t) + \beta \cdot X(t); \tag{2}$$

$$\frac{dX}{dt} = \mu \cdot X(t), \tag{3}$$

where X is the amount of cell biomass in the bioreactor, μ is the SGR, t is time, and α and β are stoichiometry parameters.

The stoichiometric coefficients α and β define the cell metabolism of oxygen consumption. Stoichiometry means that the same cell strain has the same coefficients or forms. In Eq. (2), the coefficient α describes a specific cell's oxygen consumption yield ($\alpha \equiv Y_{O_2/X}$) for growth, while β is a coefficient representing the oxygen consumption for maintenance ($\beta \equiv m_{O_2/X}$) [22,23,7].

Taking the derivative of Eq. (2) with respect to time and combining it with Eq. (3), we obtain:

$$\frac{dOUR(t)}{dt} = \alpha \cdot \frac{d\mu}{dt} \cdot X(t) + OUR(t) \cdot \mu. \tag{4}$$

Eq. (4) can then be reconstructed to eliminate the biomass parameter X to make the algorithm dependent only on OUR and stoichiometry:

$$\frac{1}{OUR(t)} \cdot \frac{dOUR(t)}{dt} = \frac{1}{\mu + \beta/\alpha} \cdot \frac{d\mu}{dt} + \mu. \tag{5}$$

Parameter R denotes the dynamics of oxygen consumption:

$$R = \frac{1}{OUR(t)} \cdot \frac{dOUR(t)}{dt}. \tag{6}$$

The differential equation can then be used to represent the dynamic relationship between the SGR values μ and R:

$$\frac{1}{\mu + \beta/\alpha} \cdot \frac{d\mu}{dt} + \mu = R. \tag{7}$$

One can also use the first-order transfer function:

$$G_{\mu/R}(s) = \frac{1}{Ts + 1}, \tag{8}$$

where s is the Laplace variable, and T is the time constant related to the SGR.

$$T = \frac{1}{\mu + \beta/\alpha}. \tag{9}$$

The dynamic relationship between OUR and R is defined by the differentiator transfer function

$$G_{R/OUR}(s) = ks, \tag{10}$$

where $k = 1/OUR$. The resulting transfer function relating the OUR to the SGR is as follows:

$$G_{\mu/OUR}(s) = \frac{ks}{Ts + 1}. \tag{11}$$

The discrete OUR measurement-based SGR estimation algorithm is then obtained from the transfer function (11) by applying the z-transform. The discrete algorithm of the SGR estimator is illustrated in Fig. 2.

In the structure scheme of the SGR estimator (Fig. 2), the first part is intended for calculating the parameter R_n, which conveys the dynamics of oxygen consumption:

$$R_n = OUR_n \cdot \frac{k(1 - z^{-1})}{\Delta t}. \tag{12}$$

By applying the z-transform for Eq. (12), where $k = 1/OUR_n$,the results yield

$$R_n = \frac{1}{OUR_n} \cdot \frac{OUR_n - OUR_{n-1}}{\Delta t}. \tag{13}$$

The last part of the structure diagram of the SGR estimator shows the relationship between the dynamics of oxygen consumption and the time constant, which gives the value of the SGR:

$$\mu_n = R_n \cdot \frac{\Delta t}{T(1 - z^{-1}) + \Delta t}. \tag{14}$$

The final formula for the SGR estimator is obtained by applying the z-transform to Eq. (14):

$$\mu_n = R_n \cdot \frac{\Delta t}{T + \Delta t} + \mu_{n-1} \cdot \frac{T}{T + \Delta t}, \tag{15}$$

where $T = 1/(\mu_{n-1} + (\beta/\alpha))$.

The presented SGR estimator is versatile and can be applied to the monitoring of various cultivation processes. A single turning parameter is the stoichiometric parameter ratio β/α, which is specific to a particular strain of microorganisms and can be found in reference books or estimated from early batch culture experiments [24]. For many cell strains and in many cultivation processes, the maintenance term β is negligible. The ratio β/α (typically 0.01–0.04) is usually smaller than the SGR by orders of magnitude. Therefore, even using a zero value for β/α in the estimation algorithm provides interpretable SGR estimation results.

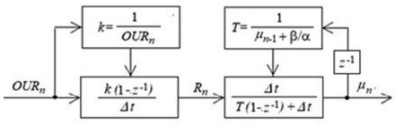


Fig. 2. Block scheme of the SGR estimation algorithm (z^{-1} is the backward-shift operator, Δt is the sampling time, and n is the number time discretization point).

5. Investigation of the SGR estimator performance

5.1. Computer simulation

The first step in assessing the performance of the SGR estimator was chosen by computer simulation using the MATLAB/Simulink platform. In the simulation, different fed-batch cultivation processes with different SGR time profiles were modeled by varying the feeding speed of the feeding solution. A mathematical model of the *E. coli* cultivation process is described in [25,26]. The following differential equation describes the biomass concentration (g/L):

$$\frac{dx}{dt} = \mu(s) \cdot x - F \cdot \frac{x}{V}, \tag{16}$$

where x is the biomass concentration, g/L; μ is the specific biomass growth rate (1/h); V is the volume of the liquid culture, L; F is the substrate feeding rate, g/h; and t is the process time, h. Another differential equation describing the glucose concentration in a bioreactor (in g/L) is as follows:

$$\frac{ds}{dt} = -q_s(s) \cdot x + F \cdot \frac{s_f - s}{V}, \tag{17}$$

where q_s is the specific substrate consumption rate, g/(g·h), and s_f is the substrate concentration in the feeding solution (g/L). The volume of the medium in the bioreactor depends directly on the feeding speed of the feeding solution.

$$\frac{dV}{dt} = F. \tag{18}$$

The oxygen uptake rate signal is calculated using the Luedeking/Piret model (g/(h·V)):

$$OUR = \alpha \cdot \mu(s) \cdot x \cdot V + \beta \cdot x \cdot V. \tag{19}$$

The dependence of the relative growth rate on the substrate concentration can be mathematically expressed using the Monod model [27,28]:

$$\mu(s) = \mu_{max} \cdot \frac{s}{k_s + s} \cdot \frac{k_i}{k_i + s}, \tag{20}$$

where μ_{max} is the maximum possible specific growth rate of the specific cell culture, and k_s and k_i are the Monod expression parameters indicating the inhibition of the cell culture by overfeeding. Finally, the simulation's mathematical expression describing the relative substrate consumption rate (in g/(g·h)) is as follows:

$$q_s(s) = \frac{\mu(s)}{Y_{x/s}} + m, \tag{21}$$

where $Y_{x/s}$ is a specific cell culture yield factor that describes the need for a certain amount of food (glucose) for a certain amount of biomass (g/g), and m refers to the model parameters that define the food requirements for biomass maintenance, g/(g·h).

The parameters of the model Eqs. (12)–(17) and the initial values of the state variables in the simulation experiments are listed in Table (1).

In the simulation experiments, various time profiles of the SGR were obtained by manipulating the feed rate. Feed-rate interruption disturbances were also added to the feed-rate time profiles to simulate the complicated process control conditions. As the actual measurements of the OUR are usually corrupted by noise, the measurements applied in the recursive estimation algorithm (Fig. 2) were simulated by adding white Gaussian noise:

$$OURm_n = OUR_n + \sigma \cdot OUR_n \cdot Rand, \tag{22}$$

where *OURm* is the measured value of OUR; the percentage standard deviation of the absolute *OUR* value estimated from observations is σ (σ = 3%), *Rand* is a number from the Gaussian random number sequence with zero mean and unit variance, and subscript n denotes the count of discrete measurement points.

Table 1
Values of model parameters.

Parameter	Value	Dimension
k_i	85	g/L
k_s	0.7	g/L
m	0.02	g/(g·h)
s_f	150	g/L
$Y_{x/s}$	0.8	g/g
α	0.82	g/g
β	0.01	g/(g·h)
μ_{max}	1.1	1/h
$x(0)$	0.5	g/L
$s(0)$	5.0	g/L
$V(0)$	8.0	L

In the simulation experiments, the time discretization step of the recursive estimation algorithm was set to Δt = 0.0025 h, and the ratio β/α (tuning parameter) value was determined to be β/α = 0.01.

Preliminary simulation experiments showed problems in the estimator's performance during the initial stage of the cultivation process. The convergence rate to the actual value of SGR at the beginning of the process was sensitive to the initial value of the SGR entered into the recursive algorithm (Fig. 2). The OUR estimation errors significantly corrupted the estimator's performance in the initial stage of the cultivation process when the OUR signal-to-noise ratio was low [6]. It was discovered that the initial measurement-based estimate of the variable (Eq. 7 provides a valid first-approach value for the SGR in the recursive estimation algorithm. Therefore, the initial SGR value can also be estimated from the early cultivation experiments. The above estimator performance problems can be resolved by switching the estimator output after some time once the OUR increases, and the estimation algorithm captures the actual value of the SGR. A proper time point for enabling the estimator was found to be 1–3 h into the cultivation process in the simulation platform.

The results of the simulation experiments under various cultivation conditions are presented in Fig. 3 (Experiments I and II). The time trajectories of the feeding rate applied to simulate different cultivation conditions are shown in Fig. 3a. The trajectories of biomass growth are shown in Fig. 3b. The simulated values of the OUR measurements $OURm_n$, upon which the SGR estimation is based, are shown in Fig. 3c. The estimator's performance in tracking time-varying biomass SGR is illustrated in Fig. 3d, in which the estimated SGR trajectories (solid lines) are compared with the actual trajectories (dotted lines).

The simulation results presented in Fig. 3 (Experiments I and II) show that the proposed estimator offers accurate SGR estimates during fed-batch cultivation processes under feeding rate disturbances and OUR measurement noise.

5.2. Experimental testing

The SGR estimator's performance and reliability were investigated using actual *E. coli* cultivation process data. Experimental SGR values and OUR data for the oxygen uptake rate were collected from fed-batch experiments of *Escherichia coli* obtained from [29] and industrial R&D laboratories. To cover more practical scenarios, three types of *E. coli* cell strains were used in different sizes of bioreactors:

1. The *E. coli* BL21 (DE3) pET21-IFN-alfa-5 cell strain was cultivated in a 7 L bioreactor. Eleven fed-batch cultivations were performed, where eight cultivations were carried out with a growth-limiting feed rate and three without a growth-limiting feed rate, that is, batch or repetitive batch processes.

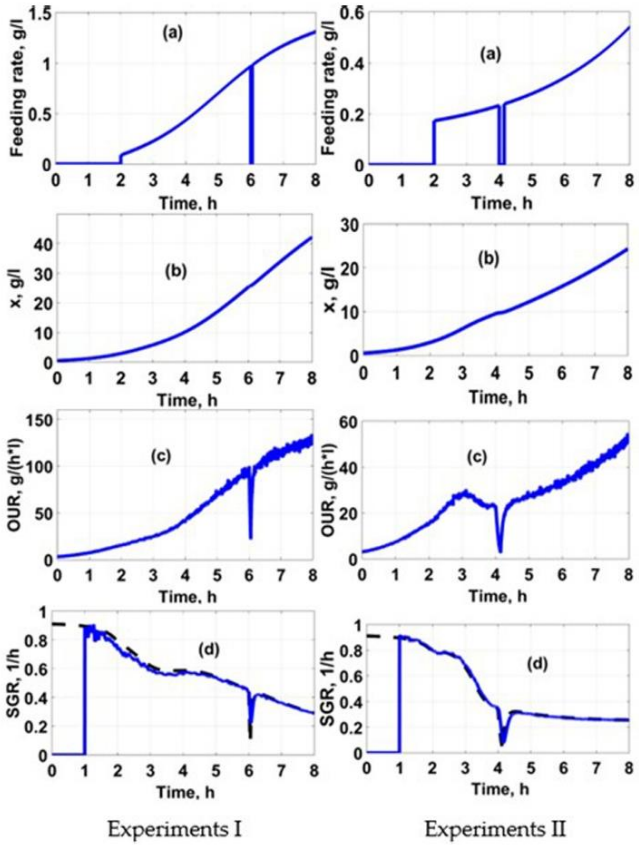


Fig. 3. Simulation results of the SGR estimator performance by tracking various SGR time trajectories (Experiments I, II): (a), (b), (c) feeding rate, biomass growth, and oxygen uptake rate curves, respectively; d) comparison of the simulated SGR versus estimated SGR curves (dotted and solid lines, respectively).

2. The *E. coli* BL21 (DE3) pET9a-IdeS cell strain was cultivated in a 12 L bioreactor. Two cultivations were performed at a growth-limiting rate.
3. The *E. coli* (BL21(DE3) pLysS) cell strain was cultivated in a 15 L bioreactor. Seven fed-batch cultivations under growth-limiting rate conditions were selected for SGR estimator inspection.

The offline biomass concentration values in the cultivation experiments were determined from the measurements of the optical density OD (in o.u.) multiplied by the coefficient of the biomass concentration (approximately 0.4 g/L/o.u.) [30]. The experimental SGR values were calculated from offline biomass concentration values collected from the sample measurements. The stoichiometry parameters of cell cultures α and β were determined using the literature and experimental data [24]. In this work, the values for the stoichiometry parameters of oxygen consumption remained the same in all *E. coli* cell cultures (ratio $\beta/\alpha = 0.04$).

The precision and reliability of the SGR estimator were evaluated by comparing the estimator predictions with the SGR values calculated offline from the biomass growth curve approximating the biomass concentration measurements. To describe the SGR estimator results, the indicators of mean absolute error (MAE) and root mean square error (RMSE) were applied. The MAE method evaluates the errors between the estimated and observed biomass values during the cultivation process. The MAE approach is defined as follows [31]:

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}, \tag{23}$$

where n is the number of data counts, and $\hat{y}_i$ is the estimation result compared to y_i, which is the value determined through the cultivation process. The root mean square error represents the square root

of the residuals of the differences between the predicted and observed values. The RMSE formula is as follows [31]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}. \quad (24)$$

The SGR estimator results for the three different cell strains are shown below.

Three experiments (9–11) were performed with a dose-unlimited substrate feeding. The rest of the experiments were provided limited feeding using the various control strategies described in [32], with multiple substrate-limited feeding profiles. The human factors and equipment influenced the results, as shown in Tables 2–4. Because the samples were taken manually, the SGR experimental values featured errors that affected the outcomes of the estimates. The overall average MAE of the SGR was 0.042 1/h, and the overall average RMSE of the SGR estimation was 0.051 1/h. These results show that this approach is acceptable for both limited and unlimited fed-batch cultivation processes with various *E. coli* cell strains.

Table 2
Analysis of *E. coli* BL21 (DE3) pET21-IFN-alfa-5, 7 L bioreactor.

Exp. No.	RMSE, 1/h	MAE, 1/h
1	0.034	0.027
2	0.054	0.047
3	0.056	0.040
4	0.051	0.036
5	0.050	0.038
6	0.045	0.038
7	0.054	0.046
8	0.057	0.040
9	0.040	0.035
10	0.041	0.030
11	0.034	0.030

Table 3
Analysis of *E. coli*(BL21(DE3) pLysS), 15 L bioreactor.

Exp. No.	RMSE, 1/h	MAE, 1/h
1	0.058	0.054
2	0.056	0.048
3	0.049	0.036
4	0.058	0.053
5	0.050	0.042
6	0.062	0.048
7	0.059	0.048

Table 4
Analysis of *E. coli*(BL21(DE3) pET9a-ldeS, 12 L bioreactor.

Exp. No.	RMSE, 1/h	MAE, 1/h
1	0.060	0.050
2	0.053	0.044

At the beginning of the cultivation process, the SGR estimator requires an initial SGR value. This can be done in two ways to obtain an initial SGR value. The first method uses two biomass concentration values taken from the measurement samples at the beginning of the cultivation process and calculate the initial SGR value. This method allows the use of the SGR estimator at the beginning of the cultivation process when two samples are taken at an interval of at least half an hour. This method is suitable if data monitoring does not start from the beginning of inoculation and when offline OD values are available. The second method (recommended by the authors of this study) uses the initial value of the SGR value set to zero. This method can be used when data monitoring of the cultivation process data started immediately after the inoculation or when the cells were still dormant. At the inoculation moment, the cells have the stress of a new environment and must be prepared for reproduction. This phase is called the lag phase. The cells prepare ferments to start reproduction; hence, in the lag phase, the specific growth rate is equal to zero [33,34].This method allows the use of the SGR estimator at the beginning of the cultivation process after inoculation without any measurements of the biomass. As shown in Figs. 4–6, the SGR estimator begins to run the start of the cultivation process.

During online monitoring of the cultivation experiments, the SGR estimator demonstrated robust behavior and consistency between the SGR online estimates and the rough SGR observations obtained from the discrete offline biomass concentration measurements.

6. Conclusions

In this study, an estimator of the biomass-specific growth rate was developed for online monitoring of microbial cultivation processes. The estimation algorithm is based on a functional model and measurements of the oxygen uptake rate.

The computer simulation of our specific-growth-rate estimator revealed robust behavior of the recursive estimation algorithm and sufficiently accurate tracking of the specific-growth-rate time trajectories under process disturbances and measurement errors of

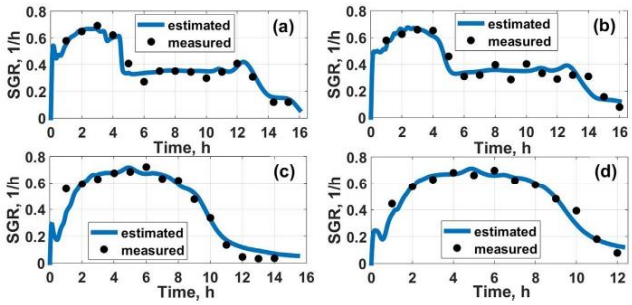


Fig. 4. SGR estimation results with cultivation process data: a) Exp. 1 Table 2, limited fed-batch cultivation processes; b) Exp. 2 Table 2, limited fed-batch cultivation process; and c) Exp. 9 Table 2, unlimited fed-batch cultivation process; d) Exp. 10 Table 2 and the unlimited fed-batch cultivation process.

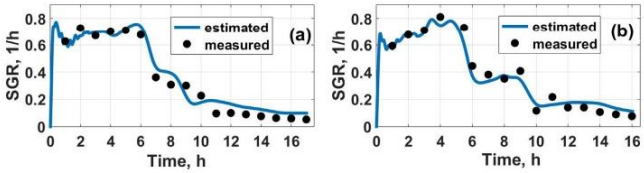


Fig. 5. SGR estimation results with limited feeding solution fed-batch cultivation process data: a) Exp. 1 Table 3; b) Exp. 2 Table 3.

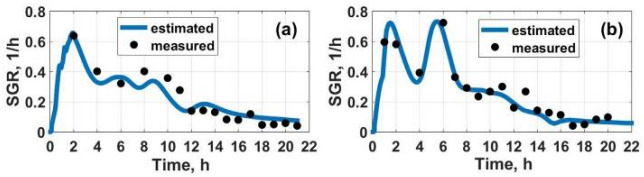


Fig. 6. SGR estimation results with limited feeding solution fed-batch cultivation process data: a) Exp. 1 Table 4; b) Exp. 2 Table 4.

the off-gas parameters. The experimental investigation estimator was established using three different *E. coli* strains in bioreactors with several different working volumes. The overall average MAE of the SGR was 0.044 1/h, and the overall average RMSE of the SGR estimation was 0.074 1/h.

This estimator can be applied to the online monitoring of various cultivation processes with limited or unlimited substrate feeding. This method requires adjusting only a single tuning parameter, that is, the ratio of β/α, to adapt the estimator to a particular process. An approximate or zero value of the tuning parameter provided satisfactory estimation results. Thus, the presented estimator can provide a proper feedback signal for advanced SGR automatic control systems.

Funding

This project was funded by the European Regional Development Fund (project No 01.2.2-LMT-K-718–03-0039) under a grant agreement with the Research Council of Lithuania (LMTLT).

CRediT authorship contribution statement

Arnas Survyla: Methodology, Software, Validation, Data curation, Writing – original draft, Visualization. **Donatas Levisauskas:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft. **Renaldas Urniezius:** Methodology, Software, Validation, Resources, Writing – original draft, Writing – review & editing, Supervision. **Rimvydas Simutis:** Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Tripathi NK, Shrivastava A. Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. Front Bioeng Biotechnol 2019;7:420. https://doi.org/10.3389/fbioe.2019.00420.

[2] Schaepe S, Kuprijanov A, Sieblist C, Jenzsch M, Simutis R, Lübbert A. Current advances in tools improving bioreactor performance. Curr Biotechnol 2013;3(2):133–44. https://doi.org/10.2174/22115501026661312172352 46.

[3] Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, manufacturing, and Quality Assurance 19. URL: https://www.fda.gov/media/71012/download..

[4] Urniezius R, Survyla A. Identification of functional bioprocess model for recombinant E. coli cultivation process. Entropy 2019;21(12):1221. https://doi.org/10.3390/e21121221. URL: https://www.mdpi.com/1099-4300/21/12/1221.

[5] Urniezius R, Kemesis B, Simutis R. Bridging offline functional model carrying aging-specific growth rate information and recombinant protein expression: entropic extension of akaike information criterion. Entropy 2021;23(8):1057. https://doi.org/10.3390/e23081057. URL: https://www.mdpi.com/1099-4300/23/8/1057.

[6] Rocha I, Veloso A, Carneiro S, Costa R, Ferreira E. Implementation of a specific rate controller in a fed-batch E. coli fermentation. IFAC Proc Vol 2008;41(2):15565–70. https://doi.org/10.3182/20080706-5-kr-1001.02632.

[7] Gnoth S, Jenzsch M, Simutis R, Lübbert A. Control of cultivation processes for recombinant protein production: a review. Bioprocess Biosyst Eng 2007;31(1):21–39. https://doi.org/10.1007/s00449-007-0163-7.

[8] Galvanauskas Simutis, Urniežius Levišauskas. Practical solutions for specific growth rate control systems in industrial bioreactors. Processes 2019;7(10):693. https://doi.org/10.3390/pr7100693.

[9] Goodwin GC. Predicting the performance of soft sensors as a route to low cost automation. Annu Rev Control 2000;24:55–66. https://doi.org/10.1016/s1367-5788(00)90013-0.

[10] Schuler MM, Marison IW. Real-time monitoring and control of microbial bioprocesses with focus on the specific growth rate: current state and perspectives. Appl Microbiol Biotechnol 2012;94(6):1469–82. https://doi.org/10.1007/s00253-012-4095-z.

[11] Claes JE, Impe JFV. On-line estimation of the specific growth rate based on viable biomass measurements: experimental validation. Bioprocess Eng 1999;21(5):389. https://doi.org/10.1007/s004490050692.

[12] Brignoli Y, Freeland B, Cunningham D, Dabros M. Control of specific growth rate in fed-batch bioprocesses: novel controller design for improved noise management. Processes 2020;8(6):679. https://doi.org/10.3390/pr8060679.

[13] Simutis R, Lubbert A. Exploratory analysis of bioprocesses using artificial neural network-based methods. Biotechnol Prog 1997;13(4):479–87. https://doi.org/10.1021/bp9700364.

[14] Soons ZITA, van Straten G, van der Pol LA, van Boxtel AJB. Online automatic tuning and control for fed-batch cultivation. Bioprocess Biosyst Eng 2007;31(5):453–67. https://doi.org/10.1007/s00449-007-0182-4.

[15] Simutis R, Lübbert A. Bioreactor control improves bioprocess performance. Biotechnol J 2015;10(8):1115–30. https://doi.org/10.1002/biot.201500016.

[16] Desai KM, Vaidya BK, Singhal RS, Bhagwat SS. Use of an artificial neural network in modeling yeast biomass and yield of β)glucan. Process Biochem 2005;40(5):1617–26. https://doi.org/10.1016/j.procbio.2004.06.015.

[17] Aehle M, Simutis R, Lübbert A. Comparison of viable cell concentration estimation methods for a mammalian cell cultivation process. Cytotechnology 2010;62(5):413–22. https://doi.org/10.1007/s10616-010-9291-z.

[18] Garcia-Ochoa F, Gomez E, Santos VE, Merchuk JC. Oxygen uptake rate in microbial processes: an overview. Biochem Eng J 2010;49:289–307. https://doi.org/10.1016/j.bej.2010.01.011.

[19] Pappenreiter M, Sissolak B, Sommeregger W, Striedner G. Oxygen uptake rate soft-sensing via dynamic kla computation: cell volume and metabolic transition prediction in mammalian bioprocesses. Front Bioeng Biotechnol 2019;7:195. https://doi.org/10.3389/fbioe.2019.00195.
[20] Luedeking R, Piret EL. Transient and steady states in continuous fermentaion. theory and experiment. J Biochem Microbiol Technol Eng 1959;1(4):431–59. https://doi.org/10.1002/jbmte.390010408.
[21] Luedeking R, Piret EL. A kinetic study of the lactic acid fermentation. batch process at controlled pH. J Biochem Microbiol Technol Eng 1959;1 (4):393–412. https://doi.org/10.1002/jbmte.390010406.
[22] Unrean P. Bioprocess modelling for the design and optimization of lignocellulosic biomass fermentation. Bioresour Bioprocess 2016;3(1). https://doi.org/10.1186/s40643-015-0079-z.
[23] Caramihai M, Severi I. Bioprocess modeling and control. In: Biomass Now – Sustainable Growth and Use: InTech; 2013. doi:10.5772/55362..
[24] Urniezius R, Survyla A, Paulauskas D, Bumelis VA, Galvanauskas V. Generic estimator of biomass concentration for escherichia coli and saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate. Microbial Cell Factories 2019;18(1). https://doi.org/10.1186/s12934-019-1241-7.
[25] Levisauskas D. Biotechnol Lett 2001;23(15):1189–95. https://doi.org/10.1023/a:1010528915228.
[26] Brion K. Measurement and control in bioprocessing. London New York: Elsevier Applied Science; 1991.
[27] Kovárová-Kovar K, Egli T. Growth kinetics of suspended microbial cells: From single-substrate-controlled growth to mixed-substrate kinetics. Microbiol Mol Biol Rev 1998;62(3):646–66. https://doi.org/10.1128/mmbr.62.3.646-666.1998.
[28] Monod J. The growth of bacterial cultures. Annu Rev Microbiol 1949;3 (1):371–94. https://doi.org/10.1146/annurev.mi.03.100149.002103.
[29] Schaepe S, Kuprijanov A, Simutis R, Lübbert A. Avoiding overfeeding in high cell density fed-batch cultures of e. coli during the production of heterologous proteins. J Biotechnol 2014;192:146–53. https://doi.org/10.1016/j.jbiotec.2014.09.002.
[30] Shiloach J, Fass R. Growing e. coli to high cell density—a historical perspective on method development. Biotechnol Adv 2005;23(5):345–57. https://doi.org/10.1016/j.biotechadv.2005.04.004.
[31] Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79–82. https://doi.org/10.3354/cr030079.
[32] Urniezius R, Galvanauskas V, Survyla A, Simutis R, Levisauskas D. From physics to bioengineering: microbial cultivation process design and feeding rate control based on relative entropy using nuisance time. Entropy 2018;20 (10):779. https://doi.org/10.3390/e20100779.
[33] Rolfe MD, Rice CJ, Lucchini S, Pin C, Thompson A, Cameron ADS, Alston M, Stringer MF, Betts RP, Baranyi J, Peck MW, Hinton JCD. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. J Bacteriol 2012;194:686–701. https://doi.org/10.1128/JB.06112-11.
[34] Madar D, Dekel E, Bren A, Zimmer A, Porat Z, Alon U. Promoter activity dynamics in the lag phase of Escherichia coli. BMC Syst Biol 2013;7:136. https://doi.org/10.1186/1752-0509-7-136.

Talanta 254 (2023) 124121

Contents lists available at ScienceDirect

Talanta

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

Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional

Arnas Survyla, Renaldas Urniezius *, Rimvydas Simutis

Department of Automation, Kaunas University of Technology, Studentu 48, LT-51367, Kaunas, Lithuania

ARTICLE INFO

Keywords:
Mammalian cells
Stoichiometry
Aging
Online estimation
Viable cells
Soft sensor

ABSTRACT

This study developed an estimation routine for counting the viable cells in an in vitro fed-batch Chinese hamster ovary cultivation that relies on off-gas information and inlet gas mixture knowledge. We computed the oxygen uptake rate bound to the bioreactor exhaust gas outlet when the inlet gas mixture was stationary. Our mammalian biosynthesis analysis determined the stoichiometric parameters as a function of the average population age. We cross-validated an identical algorithm for mammalian and microbial cultivations and found that the' 99% confidence band of the model generally overlapped with the error bars defined from observations. The resulting RMSE and MAE averages were 0.188 and 0.14e^{9}cells L^{-1}, respectively, when estimating the viable mammalian cell count. The validation for the estimation of total bacterial biomass yielded an MAE and RMSE of 1.78 g L^{-1} and 2.53 g L^{-1}, respectively. Moreover, our proposed approach provides an online estimation of the average population age for both aerobically cultivated microorganisms.

1. Introduction

Over a decade ago, investigators showed that the cumulative oxygen uptake rate (OUR) is a reliable indicator of the cell viability repeatability in mammalian fed-batch biosynthesis [1]. In the same year, they showed that simple and efficient substrate-feeding control based on the OUR signal is a promising tool for validating the variability of viable cell counts using an off-gas analyzer [2]. Similarly, another team showed that gas analyzer information and bioreactor parameters can further help optimize the target product in mammalian fermentation [3]. Later efforts estimated the intermediate state variables of bioreactors; however, no report on noninvasive viable cell estimation in fed-batch mammalian biosynthesis has yet to be published. Animal cells are the closest strain to human cells, producing many high-quality and specific proteins that are used in unique medical applications [4]; for example, the Chinese hamster ovary (CHO) is a well-known mammalian cell strain used to produce glycoproteins [5].

Cultivating bioprocesses with mammalian cells to complete target-product fermentation with high efficiency is challenging. Animal-cell-based biosynthesis is at a relatively higher risk than that based on microbial cells [6] because of the longevity of the process, the seed of the strain, and the nutrition medium. To reduce the risks of this process, bioreactor control [7,8] must depend on reliable real-time estimations of the culture state. Thus, monitoring the main characteristic parameter, that is, the number of viable cells, is crucial; however, contemporary viable cell measurements are performed offline, which is time-consuming and human resource intensive.

This paper presents a soft sensor (i.e., estimator) as a tool for estimating viable CHO cells using noninvasive off-gas [9] measurements that depend on oxygen consumption rate [10] information. The method is based on stoichiometric parameters and the Luedeking–Piret model with an aging term introduced [11–13]. The primary off-gas signal reported by the viable cell estimator defined the oxygen uptake rate input. Exhaust gas analyses have provided information about cultures in media that is indirectly related to oxygen-consuming viable cells [14].

Section 1 discusses the motives for using the noninvasive cell counter as a novel functionality of gas analyzers; Section 2 reviews literature related to this study; Section 3 describes the bioreactor system materials and the protocol conditions; Section 4 outlines the development of a viable cell-estimation algorithm; Section 5 lays out the hypothetical functional model, aging-specific parameters, and the motivation behind the assumptions; and the final section discusses the conclusions of this study.

* Corresponding author. Department of Automation, Kaunas University of Technology, Studentu 50, LT-51368, Kaunas, Lithuania.
E-mail address: renaldas.urniezius@ktu.lt (R. Urniezius).

https://doi.org/10.1016/j.talanta.2022.124121
Received 10 October 2022; Received in revised form 17 November 2022; Accepted 19 November 2022
Available online 25 November 2022

2. Related work

The Luedeking–Piret model was used to estimate the viable cells of a mammalian strain in a bioreactor. Earlier researchers have shown that a standard stoichiometric model is crucial for estimating the microbial biomass of *Escherichia coli*, as it provides acceptable results [15]. The main difference between microbial and mammalian cells is the specific growth rate, which is relatively low for the CHO strain. Consequently, the long cultivation time leads to instability in the oxygen consumption rate by viable cells. The average age of the cell population was used to describe its yield dynamics. At low ages, the potential specific growth rates of the cells were high, and the inhibition of the cell count had a negligible effect. Furthermore, cultures of more considerable ages have their cell growth inhibited. The maintenance of aging cells [16] determines excess oxygen intake.

Cell fermentation is a complex process [17] that requires sophisticated control [18,19] and monitoring of biotechnological phenomena [20]; therefore, data collection and model management are critical [21], as they provide essential variables in real time with acceptable accuracies [22] to identify the state of the process and its adaptive control [23]. Several developments have been made to obtain real-time measurements of the amount of viable cells using equipment installed in a bioreactor. One of these developments includes using image analysis, such as microscopy [24,25]. The equipment necessary to count the cells requires maintenance (recalibration), and image data might be disturbed owing to enriched media supplements or homogeneity-related effects. Furthermore, the predicted cell counts often deviate from the ground truth, even for simple test images [26].

Another way to indirectly estimate viable cell counts in culture is to use in situ mid-infrared spectrometry and online glucose concentration, which has been established [27]. Using a non-stationary growth yield parameter, viable cells can be computed from direct glucose concentration measurements; however, one major disadvantage of such an approach is the necessity for in situ equipment to obtain the glucose concentration online. Moreover, the function of glucose consumption faces accuracy challenges when determined in the culture death phase, that is, the decline of the cell population instead has steady-state properties [27].

Soft sensors [28] are analytical tools for the online observation of in-situ parameters [10,29]. Many different models of soft biomass sensors have been proposed for microbial biosynthesis, such as dielectric spectroscopy [30]; however, their use for cell count estimation remains challenging in mammalian and stem cell bioprocessing [31,32].

Cultivation processes with animal cells are more complex than those with microorganism cells because of the meager specific growth rate and strict requirements for the composition of the medium and precise maintenance of environmental parameters [33]. The shallow specific growth rate of mammalian cells significantly prolongs the cultivation process, which can greatly disturb the in-situ data with relatively higher transient constants, and noise has a considerable influence on the signal [1].

To eliminate signal noise and resolve the complexity of estimating the viable cells of animal strains, a hybrid model with artificial neural networks (ANN) was proposed as a viable cell estimation tool [34], with a hybrid model that includes exhaust gas analyses and a base (NaOH for pH setpoint control) providing the most acceptable results; however, obtaining sufficient accuracy for ANNs requires considerable data for model calibration. Moreover, hybrid model approaches require significant performance trade-offs and design-space maintenance. Additionally, the resulting model applies exclusively to a specific bioprocess [34,35].

This study estimates cell counts based only on exhaust gas data and the OUR, which is closely associated with dissolved oxygen [36] and is crucial for aerobic cultivation (Table 1).

The chosen viable cell estimation method is based on stoichiometry and aging theory [11,12] and avoids the use of a data-driven black box (ANN). The selected model was based on knowledge combined with off-gas analytics.

Table 1
Comparison between proposed mammalian viable cells estimation.

Source	Inputs	Model-based	Equipment
Joeris et al. [24], Shah et al. [25]	Visual material	Image processing software	Microscope
Ducommun et al. [27]	Glucose concentration	Parametric optimization	Mid-infrared spectrometry
Aehle et al. [34]	*OUR*, *CPR*, *Base*	Data-driven (ANNs), recurrent model	Exhaust gas analyzer, balance for base
This study	*OUR*	Functional optimization	Exhaust gas analyzer

3. Materials and methods

3.1. Cultivation conditions

The viable cell estimator for mammalian cell culture proposed in this study used data from the cultivation process of CHO–K1 (CHO–S, No. 11619–012, Karlsruhe, Germany). Prior research [37] presents the Biostat B bioreactor cultivation processes, the information of which is presented in Table 2.

An automated cell counter was used to measure offline viable and total cell concentrations (CASY TT; Roche Innovatis AG, Mannheim). Exhaust gas analyses were performed using a quadrupole mass spectrometer (Balzers QMA 200; Balzers, Liechtenstein).

Furthermore, data from bioprocesses with the *Escherichia coli* strain were used to estimate the biomass. The bacteria *E. coli* BL21 (DE3) pET21-IFN-alfa-5 (Table 3) were cultivated in a minimal-mineral medium [38].

The BlueSens BlueInOne Ferm gas analyzer (oxygen concentrations from 0 to 100%) and airflow information from Applikon BioBundle bioreactor enabled the oxygen uptake rate assessment.

3.2. Development of the viable cells estimation algorithm

Off-gas analysis is founded on the basis that its information source is cumulative. The entire bioreactor medium, with an inevitable time delay, determines the gas mixture content at the condenser outlet, that is, accumulated carbon dioxide needing to be removed from nutrient media explains the reason that off-gas analysis with the oxygen uptake rate signal is a rational intuitive candidate invariant to the homogeneity of the bioreactor medium. A typical off-gas-based candidate for the stoichiometric relationship between the total OUR and the biomass growth and maintenance is the Luedeking–Piret-type model [11,12].

$$OUR(t) = \alpha \cdot \dot{X}(t) + \beta \cdot X(t); \tag{1}$$

where X is the total count of viable cells, t is the time, and α and β are parameters that determine the corresponding stoichiometric relationship with the growth and maintenance of viable cells. To introduce a generic estimator for the number of viable cells, the time dependence of both kinetic parameters indicates a general inhomogeneous first-order differential equation [39].

Table 2
Mammalian cell cultivation details.

Condition	State	Condition	State
Bioreactor volume	2 L	Broth volume	1 L
Temperature	37 °C	PH	7.15
pO2	20%	airflow	0.1 L min^{-1}
Stirrer	60-400 RPM	Feeding start	at 75 h

Table 3
Recombinant *E. coli* cultivation details.

Condition	State	Condition	State
Bioreactor volume	7 L	Broth volume	3.7 L
Temperature	37 °C	PH	6.8
pO2	20%	Feeding start	at 5–7 h

$$X^{'}(t)+\frac{\beta(t)}{\alpha(t)}\cdot X(t)=\frac{OUR(t)}{\alpha(t)}. \tag{2}$$

The generic solution to 2 is as follows:

$$X(t)=\frac{c+\int\frac{OUR(t)}{\alpha(t)}e^{\int\frac{\beta(t)}{\alpha(t)}\mathrm{d}t}\mathrm{d}t}{e^{\int\frac{\beta(t)}{\alpha(t)}\mathrm{d}t}}, \tag{3}$$

where integration constant c must account for the boundary condition. Specifically, the exact form of the answer with initial condition $X(0)\equiv X_0$ is as follows:

$$X(t)=\frac{X_0+\int_0^t\frac{OUR(t_1)}{\alpha(t_1)}e^{\int_0^{t_1}\frac{\beta(t_2)}{\alpha(t_2)}\mathrm{d}t_2}\mathrm{d}t_1}{e^{\int_0^t\frac{\beta(t_1)}{\alpha(t_1)}\mathrm{d}t_1}}. \tag{4}$$

3.3. Maintenance component

Previously, the authors [15] demonstrated that the boundary condition has to be computationally resolved when the maintenance term is negligible prior to the induction phase of biosynthesis during microbial bioprocesses. As there is no induction in mammalian upstream development, the age-related threshold of viable cell populations serves as the rational hypothesis to assume the start of the cell maintenance effect. In the interim, the verge is defined as follows: $k_{cX}\equiv\int_0^{t_{cX}}X(t)\,\mathrm{d}t$, where the time instant t_{cX} and its biomass ($X_{cX}\equiv X(t_{cX})$) are unknown in advance. However, the approximate value of k_{cX} was assessed or estimated in the model training phase. Then, 1 generalizes to the following:

$$\begin{cases} OUR(t)=\alpha(t)\cdot X^{'}(t); & \text{if}\quad \int_0^t X(t_1)\,\mathrm{d}t_1\le k_{cX}\\ OUR(t)=\alpha(t)\cdot X^{'}(t)+\beta(t)\cdot(X(t)-X_{cX}), & \text{otherwise,}\end{cases} \tag{5}$$

where the second equality, similarly to 2 has an alternative arrangement

$$X^{'}(t)+\frac{\beta(t)}{\alpha(t)}\cdot X(t)=\frac{OUR(t)+\beta(t)\cdot X_{cX}}{\alpha(t)}. \tag{6}$$

The solution of 6 is the extended form of 4. Then,

$$\begin{cases} X(t)=X_0+\int_0^t\frac{OUR(t_1)}{\alpha(t_1)}\,\mathrm{d}t_1; & \text{if}\quad \int_0^t X(t_1)\,\mathrm{d}t_1\le k_{cX}\\ X(t)=\frac{X_0+\int_0^t\frac{OUR(t_1)+\beta(t)\cdot X_{cX}}{\alpha(t_1)}e^{\int_0^{t_1}\frac{\beta(t_2)}{\alpha(t_2)}\mathrm{d}t_2}\,\mathrm{d}t_1}{e^{\int_0^t\frac{\beta(t_1)}{\alpha(t_1)}\mathrm{d}t_1}}, & \text{otherwise.}\end{cases} \tag{7}$$

Equation (7) indicates that, as time approaches infinity, the initial value of the biomass is negligible. Second, because the aging effect is noticeable in $OUR(t)$, at least for fed-batch bioprocesses, such an integral form has potential benefits for estimators based on finite differences.

3.4. Hypothetic functional model of kinetic parameters

Generally, both kinetic parameters ($\alpha(t)$ and $\beta(t)$) of the Luedeking–Piret model are functions of time 7; however, we hypothesized, based on previous research results [12,40,41], that in prolonged fed-batch aerobic bioprocesses, such as mammalian cell cultivation, kinetic parameters are functionals that depend on the average age of the cell population.

$$\overline{Age}(t)=\frac{X_0\cdot t+\int_0^t(t-t_1)\cdot X^{'}(t_1)\mathrm{d}t_1}{X(t)}. \tag{8}$$

The main drawback of this expression, when used for continuous biosynthesis analysis, is that it is time-dependent. It is more suitable to make it more generalizable so that both fed-batch and continuous biosynthesis use the same form. The arrangement and integration give the following:

$$\overline{Age}(t)=\frac{X(t)\cdot t-\int_0^t t_1\mathrm{d}X(t_1)}{X(t)}. \tag{9}$$

After applying the integration by parts formula to the numerator, it becomes the following:

$$Age(t)\equiv\overline{Age}(t)=\frac{\int_0^t X(t_1)\mathrm{d}t_1}{X(t)}, \tag{10}$$

which is equally convenient for non-invasive estimation in both fed-batch and continuous biosynthesis. The age expression depends on the state rather than time. Such an assumption is relevant for perfusion bioprocesses [42], when the biomass concentration (microbial) or number of viable cells (mammalian) might be age-invariant.

The choice was to introduce a parametric hypothesis for fed-batch mammalian cultivation. The following functionals served as the model fitting classes to enable non-invasive online estimation of the kinetic coefficient ($\alpha(t)$) at runtime:

$$\alpha(t)\equiv\frac{\alpha_{max}}{1-e^{-\frac{t}{Lag_{time}}}}\cdot\frac{Age(t)}{t}, \tag{11}$$

where the maximal growth-based oxygen consumption yield ($\alpha_{\max}$) for cells represents the theoretical aerobic oxidative capacity and the lag time (Lag_{time}) is related to exponential decay [43] and defines the moment when the lag phase approaches the end and cells enter the exponential growth phase. The rightmost multiplier 11 also has a physical meaning of the relative time ratio, designated for the last stage of biosynthesis. The oxygen consumption yield ($\beta(t)$) allows cells to remain alive

$$\beta(t)=\beta\cdot\frac{Age(t)}{Age(t)+k_{age}}, \tag{12}$$

where aging-specific coefficient k_{age} is "half-age-constant" if the maintenance coefficient β is treated as the maximal maintenance value.

3.5. Online numeric estimation of viable cells count

Finite differences allow integral routines to be simplified; however, the computational inertia of accumulating errors and algorithmic performance challenges must be avoided. The first biomass value could be zero if the fed-batch bioprocess is considerably longer, similar to mammalian cultivation. Moreover, the initial biomass, that is, the number of viable cells, is typically known after bioreactor inoculation in industrial installations. The initial count of viable cells is the result of offline analyses in this study and serves as the initial estimate. In this work, the age estimate is a function of the prior viable cell estimates, as follows:

$$\widehat{Age}_i\equiv\frac{\sum_{i=1}^n\widehat{X}_i\cdot\triangle t_{i,i-1}}{\widehat{X}_i}, \tag{13}$$

where n denotes the total number of discrete observations (estimates). The critical time at which maintenance starts is defined as follows:

$$t_{cX}\equiv\left(\int_0^{t_{cX}\ \text{or the total duration}}X(t)\mathrm{d}t\right)^{-1}(k_{cX}), \tag{14}$$

where the notation "-1" is an inverse function or t_{cX} is such that the cumulative biomass equals the threshold k_{cX}. The discrete formula is expressed as follows:

$$\widehat{t_{cX}} \equiv \left(\sum_{i=1}^{n} \widehat{X}_i \cdot \triangle t_{i,i-1} \right)^{-1} (k_{cX}), \tag{15}$$

where the cumulative biomass $\widehat{X_{cX}}$ is the boundary value for the iterative routine. The stoichiometry parameters depend on the most recent estimates. Then, the kinetic coefficient $(\alpha(t))$ at runtime,

$$\alpha_i \equiv \frac{\alpha_{max}}{1 - e^{-\frac{1}{Lag_{time}} \cdot \frac{Age_{i-1}}{t_{i-1}}}}, \tag{16}$$

$$\beta_i = \beta \cdot \frac{Age_{i-1}}{Age_{i-1} + k_{age}}, \tag{17}$$

which does not degrade the convergence of the overall algorithm. Therefore, it is recommended that the sampling interval $\triangle t_i \equiv \triangle t_{i,i-1}$ be noticeably smaller than the expected transient constant of the population age dynamics. One minute (0.167 h) represented a discretization step during the experiments because of the dependency on the off-gas OUR observations. Moreover, the algorithm is sufficient for varying sampling intervals, that is, when measurements are temporarily lost, because the aging state of the culture does not change abruptly in fed-batch cultivations.

$$\begin{cases} \widehat{X}_i = X_0 + \sum_{j=1}^{i} \frac{OUR_j}{\alpha_j} \triangle t_j; & \text{if } \sum_{j=1}^{i-1} X_j \triangle t_j \le k_{cX} \\ \widehat{X}_i = \dfrac{X_0 + \sum_{j=1}^{i} \frac{OUR_j + \beta_j \cdot \widehat{X_{cX}} \cdot e^{\sum_{l=1}^{j} \frac{\beta_l}{\alpha_l} \triangle t_l}}{\alpha_j} \triangle t_j}{e^{\sum_{j=1}^{i} \frac{\beta_j}{\alpha_j} \triangle t_j}}, & \text{otherwise,} \end{cases} \tag{18}$$

where Algorithm 18 did not show convergence issues when the time variable t approached higher values than the microbial analysis. Non-invasive online observations decrease over time in fed-batch cultivation when nutrient medium perfusion is absent.

4. Results and discussion

Viable CHO cells and the biomass of the recombinant *E. coli* strain were selected to determine the reliability and performance of the estimation. The stoichiometric parameters of the cell culture (α, β) were assumed to be independent of the experimental analysis in which they lie. The discrete check compares the offline and online analysis results using the mean absolute error (MAE) and root mean square error (RMSE). The MAE is defined as [44]:

$$MAE_i = \frac{\sum_{j=1}^{n} \left| \widehat{y}_i - y_{i,j} \right|}{n}, \tag{19}$$

where the average $\widehat{y}_i$ is the total number (n) of observations y_i obtained through offline sampling. The RMSE formula is defined as [44]:

$$RMSE_i = \sqrt{\frac{\sum_{j=1}^{n} \left(\widehat{y}_i - y_{i,j} \right)^2}{n}}. \tag{20}$$

Overall, ten cross-validation steps returned ten estimation sets. Next, the ensemble averaging [45] scheme inferred a single optimal set of parameters for the estimation. The principal purpose of the chosen technique was to acquire a result from the submitted candidates by averaging them according to a weight that depends on the relevance of the item. Specifically, the weights of ten candidate sets were dependent on RMSE by applying the ensemble averaging equation:

$$\widehat{y} = \sum_{i=1}^{n} w_i y_i(\mathbf{x}), \tag{21}$$

where the final guess of parameter $\widehat{y}$ and the weight w_i of parameter y_i are RMSE-based functions resulting from the parameters listed in Table 5:

$$w_i = \frac{\sum_{j=1}^{n} RMSE_j - RMSE_i}{\sum_{j=1}^{n} RMSE_j \cdot (n-1)}. \tag{22}$$

In Equation (22), the number of parameter sets $n = 10$ leads to the final and optimal parameter sets presented in Table 4.

4.1. Estimation of mammalian cells viability

Experimental data on viable CHO cells and oxygen consumption are presented in detail in Ref. [37]. Every 12 h, manual offline sampling was used to quantify the count of viable cells using an automated cell counter (CASY TT).

The development of a method for estimating CHO viable cells and stoichiometric and inhibition parameters passed cross-validation using data from 12 cultivations of a CHO mammalian strain [46]. The 12 presented experiments were unique in terms of growth profile similarity. In the cross-validation method, 80% of the process data points helped in model fitting (training stage), and the remaining 20% concluded trials. Uniform Random indexes to skip in the model calibration originated from the "Random" function (C#) with default seed values. The experiment consisted of ten random data sets. Table 5 contains the model calibration and validation results using RMSE and MAE; Table 4 lists the optimal parameter values.

During parameter identification, the interference from signal noise and device calibration inaccuracy of the exhaust gas analysis sensors had a significant impact. The specific growth rate of mammalian cells is meager; thus, the signal-to-noise ratio is sufficiently high to cause issues in estimation precision [47]. To increase the parameter estimation accuracy, the choice was to introduce an offset for the oxygen concentration signal at the gas mixture inlet of the bioreactor. Such an improvement considers practical experience and knowledge of how the volumetric oxygen transfer coefficient (kLa) varies in the bioreactor. It was assumed that the acceleration of oxygen consumption could not exceed the dynamics of the pO2 signal, considering that the airflow and stirrer values were stationary [48] (Fig. 1).

In conclusion, an abrupt change in the oxygen consumption rate results in the quality of sensor calibration in this specific context. Hence, the offset values (Table 6) were re-fitted for all 12 experiments to increase confidence in the proposed approach.

Overall, the estimation of viable mammalian cells provided accurate predictions. The average RMSE and MAE values were 0.188 and 0.14, respectively.

The average RMSE and MAE validation values were 0.158 and 0.139, respectively. Table 7 compares 'our results with a hybrid model. Figs. 2 and 3 represent the performance of the model estimations and the confidence band with $\alpha = 0.01$. Classification of the error values between the measured and calculated points for all data from the 12 experiments over a range of viable cell concentrations was sufficient to

Table 4
Final and optimal values of model parameters.

Parameter	Value	Unit
Lag_{time}	20.489	h
α_{max}	0.727	g e^9cells^{-1}
β	0.034	g e^9cells^{-1} h^{-1})
k_{cX}	29.99	e^9cells·h L^{-1}
k_{age}	102.05	h

Table 5
Individual model fitting the cross-validation results.

Iteration	Model calibration, (e^9cells L^{-1})		Validation, (e^9cells L^{-1})	
No.	**MAE**	**RMSE**	**MAE**	**RMSE**
1	0.137	0.191	0.154	0.169
2	0.145	0.198	0.122	0.135
3	0.141	0.189	0.154	0.166
4	0.142	0.189	0.123	0.146
5	0.149	0.198	0.105	0.122
6	0.143	0.19	0.129	0.158
7	0.14	0.173	0.18	0.202
8	0.144	0.193	0.12	0.132
9	0.14	0.188	0.131	0.146
10	0.14	0.176	0.178	0.199

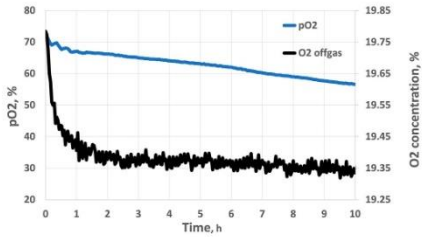


Fig. 1. The steep oxygen concentration drop at the exhaust gas outlet. The unique identity of this experiment was No 9.

Table 6
The oxygen consumption signal characteristics at the gas mixture inlet.

Exp. No.	Inlet O_2(%)	Offset ((%)	Exp. No.	Inlet O_2(%)	Offset ((%)
1	20.584	0.143	7	19.327	−0.045
2	19.785	0.034	8	21.546	−0.12
3	19.689	0.11	9	21.195	0
4	19.773	0.415	10	21.757	−0.030
5	19.476	0.043	11	21.698	−0.022
6	19.353	−0.102	12	21.776	−0.028

Table 7
Compression of CHO viable cells estimation techniques.

Author	**Estimation** technique	**RMSE** of training (e^9cells L^{-1})	**RMSE** of validation (e^9cells L^{-1})
Aehle et al. [34]	Hybrid model	0.16	0.154
This study	Functional optimization	0.188	0.158

determine the confidence band of the method. The statistical histogram and degrees of freedom are presented in Table 8. The error classification method shows the error dependence of the viable cell concentration (VC) as a model fit:

$$\widehat{e}(VC) = 0.0068VC^2 + 0.0604VC. \quad (23)$$

The purple shadow depicts the confidence band $\alpha = 0.01$ in Figs. 2 and 3. This area has a high (pessimistic) bound to the error statistics. The error bars were assumed to result from applying a systematic error of 0.1^9 cells L^{-1} and a random error of 4%. These errors consisted of experimentation-related additive errors and device characteristics.

4.2. Biomass estimation on E. coli bacteria

The chosen method also evaluated the biomass estimation of the bacterial strains to verify the 'versatility of the algorithm. Half of the 12 experiments contained a growth-limiting substrate feed. Offline biomass concentration values helped to identify the parameters of the estimation model for *E. coli* bacteria and evaluate the validation results. The offline measurements consisted of optical density (OD) (in o.u.) samples (Eppendorf BioSpectrometer basic) multiplied by the coefficient of the biomass concentration (approximately 0.4 g L^{-1} o.u.$^{-1}$) [49].

Table 9 presents the parameter set for the bacterial strains. Parameter identification results in 1.67 g L^{-1} MAE and 2.87 g L^{-1} RMSE. The validation process produced 1.78 g L^{-1} MAE and 2.53 g L^{-1} RMSE. Fig. 4 represents an analogical methodology applied to microbial analysis, and the confidence band relationship for bacterial examination is identical to that for mammalian analysis.

The histogram statistics are presented in Table 10 and the obtained prediction error dependencies by biomass (X) are as follows:

$$\widehat{e}(X) = 0.00172X^2 + 0.03431X + 0.56058. \quad (24)$$

In Fig. 4, the purple shadow indicates the confidence band $\alpha = 0.01$ in Fig. 4. The error bars consisted of a systematic error of 0.2 g L^{-1} and a random error of 4%. These errors reflect the bounds of experimentation-related errors and device characteristics.

5. Conclusions

This study proposes a model for estimating viable cells in a mammalian CHO strain to indirectly monitor the crucial state variable of the cultivation process. The proposed method was developed using functional optimization, including aging information and off-gas observations, based on the OUR at the outlet of the bioreactor. Experimental cross-validation was performed for both microbial and mammalian strains. A total of 12 experiments for each strain allowed the same model training and validation procedures to pass through. The final average MAE of the viable cell estimation (mammalian scenario) result was 0.139e^9cells L^{-1} and the overall mean RMSE result was 0.158e^9cells L^{-1}. These numerical precision results match the original hybrid findings [37] but with additional benefits. First, the number of parameters used in the functional approach was minimal. Second, these parameters have physiological implications, allowing optimal planning control for future digital twin technology.

Furthermore, a universality check was performed on the experimental data of the *Escherichia coli* recombinant strain. The procedure was identical except that the microbial stoichiometry parameters did not directly depend on the average aging of the bacterial population. The results were satisfactory: the final mean MAE was 1.78 g L^{-1}, and the overall RMSE was 2.53 g L^{-1}.

A comparison of the results for CHO and *E. coli* cells shows that the aging-specific formulation has a considerable influence on modeling when the longevity of the bioprocess is demonstrative, that is, in mammalian biosynthesis.

Author contributions

Conceptualization, R.U., A.S.; methodology, R.U., A.S.; validation, A. S., R.U.; formal analysis, A.S. and R.U.; investigation, A.S.; resources, R. U., R.S.; data curation, R.S., and A.S.; writing—original draft preparation, A.S., R.U.; writing—review and editing, R.U.; visualization, A.S.; supervision, R.U. All authors have read and agreed to the published version of the manuscript.

Funding

This project received funding from the European Regional

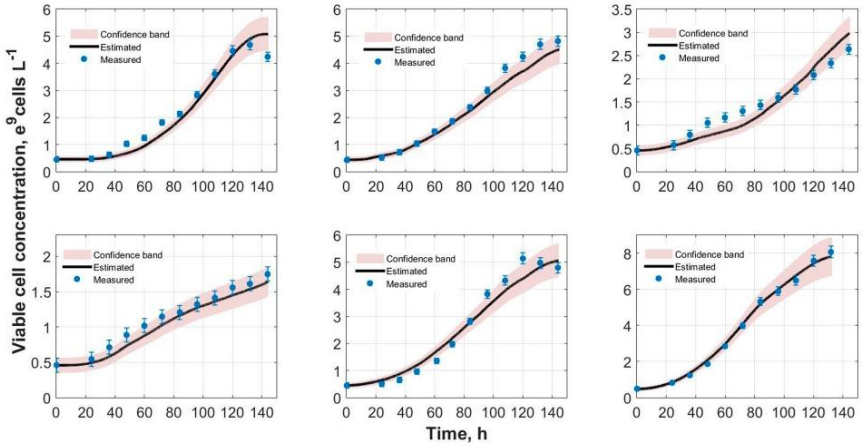


Fig. 2. CHO viable cell estimation results from experiments No 1–6. Vertical error bars indicate a total error. The purple shadow represents the prediction band. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

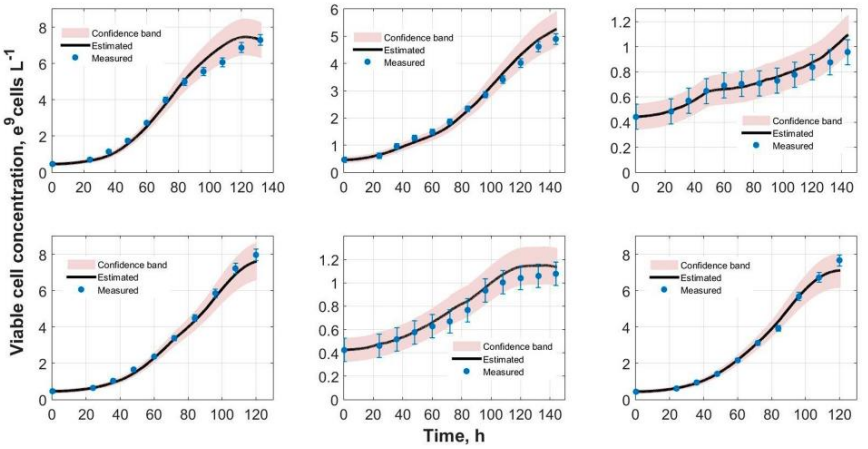


Fig. 3. CHO viable cell estimation results from experiments No 7–12. Vertical error bars indicate a total error. The purple shadow represents the prediction band. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 8
Numbers of freedom at a specific range of viable cell concentration.

Range, e^9cells L^{-1}	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8
No. of freedom	36	35	13	8	13	5	4	4

Table 9
The model parameters for *E. coli* bacteria biomass estimation.

Parameter	Value	Unit
Lag_{time}	0	h
α_{max}	0.75	g g^{-1}
β	0.16	g g^{-1} h^{-1})
k_{cX}	17	g h L^{-1}
k_{age}	0	h

Development Fund (project No 01.2.2-LMT-K-718-03-0039) under grant agreement with the Research Council of Lithuania (LMTLT).

Declaration of competing interest

The authors declare that they have no known competing financial

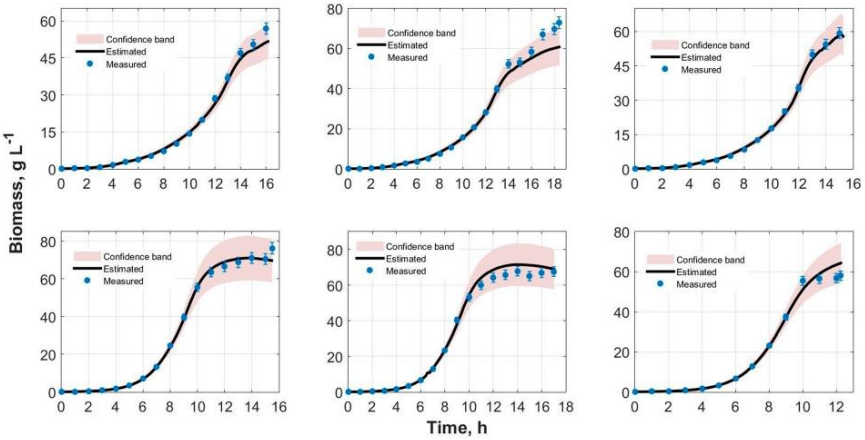


Fig. 4. *E. coli* bacteria biomass estimation results of the first six experiments. Vertical error bars indicate a total error. The purple shadow represents the prediction band. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 10
Numbers of freedom at a specific range of biomass concentration.

Range, X L^{-1}	0–5	5–10	10–15	15–20	20–30
No. of freedom	66	17	13	5	12
Range, X L^{-1}	30–40	40–50	50–60	60–70	70–80
No. of freedom	10	7	29	16	6

interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] M. Aehle, A. Kuprijanov, S. Schaepe, R. Simutis, A. Lübbert, Simplified off-gas analyses in animal cell cultures for process monitoring and control purposes, Biotechnol. Lett. 33 (11) (2011) 2103–2110, https://doi.org/10.1007/s10529-011-0686-5.

[2] M. Aehle, S. Schaepe, A. Kuprijanov, R. Simutis, A. Lübbert, Simple and efficient control of CHO cell cultures, J. Biotechnol. 153 (1–2) (2011) 56–61, https://doi.org/10.1016/j.jbiotec.2011.03.006.

[3] F. David, A. Berger, R. Hänsch, M. Rohde, E. Franco-Lara, Single cell analysis applied to antibody fragment production with Bacillus megaterium: development of advanced physiology and bioprocess state estimation tools, Microb. Cell Factories 10 (1) (2011) 23, https://doi.org/10.1186/1475-2859-10-23.

[4] G. Walsh, Biopharmaceutical benchmarks 2006, Nat. Biotechnol. 24 (7) (2006) 769–776, https://doi.org/10.1038/nbt0706-769.

[5] P. Hossler, S.F. Khattak, Z.J. Li, Optimal and consistent protein glycosylation in mammalian cell culture, Glycobiology 19 (9) (2009) 936–949, https://doi.org/10.1093/glycob/cwp079.

[6] R. Urniezius, A. Survyla, Identification of functional bioprocess model for recombinant e. coli cultivation process, Entropy 21 (12) (2019), https://doi.org/10.3390/e21121221.

[7] D. Levisauskas, R. Simutis, V. Galvanauskas, R. Urniezius, Simple control systems for set-point control of dissolved oxygen concentration in batch fermentation processes, Chem. Eng. Transact. 74 (2019) 127–132, https://doi.org/10.3303/CET1974022.

[8] Simutis Galvanauskas, Urniežius Levišauskas, Practical solutions for specific growth rate control systems in industrial bioreactors, Processes 7 (10) (2019), https://doi.org/10.3390/pr7100693.

[9] A. Survyla, R. Urniezius, V. Vaitkus, D. Levisauskas, L. Jankauskaite, D. Lukminaite, G. Laucaityte, Noninvasive continuous tracking of partial pressure of oxygen in arterial blood: adapting microorganisms bioprocess soft sensor technology for holistic analysis of human respiratory system, in: 2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI), 2021, https://doi.org/10.1109/mfi52462.2021.9591182.

[10] G.C. Goodwin, Predicting the performance of soft sensors as a route to low cost automation, Annu. Rev. Control 24 (2000) 55–66, https://doi.org/10.1016/s1367-5788(00)90013-0.

[11] R. Luedeking, E.L. Piret, Transient and steady states in continuous fermentaion. theory and experiment, J. Biochem. Microbiol. Technol. Eng. 1 (4) (1959) 431–459, https://doi.org/10.1002/jbmte.390010408.

[12] R. Urniezius, B. Kemesis, R. Simutis, Bridging offline functional model carrying aging-specific growth rate information and recombinant protein expression: entropic extension of akaike information criterion, Entropy 23 (8) (2021), https://doi.org/10.3390/e23081057.

[13] M. Matukaitis, D. Masaitis, R. Urniezius, L. Zlatkus, V. Vaitkus, Non-invasive estimation of acetates using off-gas information for fed-batch e. coli bioprocess, in: ECP 2022, 2022, https://doi.org/10.3390/ecp2022-12668.

[14] H. Gjerkes, J. Malensek, A. Sitar, I. Golobic, Product identification in industrial batch fermentation using a variable forgetting factor, Control Eng. Pract. 19 (10) (2011) 1208–1215, https://doi.org/10.1016/j.conengprac.2011.06.011.

[15] R. Urniezius, A. Survyla, D. Paulauskas, V.A. Bumelis, V. Galvanauskas, Generic estimator of biomass concentration for escherichia coli and saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate, Microb. Cell Factories 18 (1) (2019), https://doi.org/10.1186/s12934-019-1241-7.

[16] C. Tomasetti, J. Poling, N.J. Roberts, N.R. London, M.E. Pittman, M.C. Haffner, A. Rizzo, A. Baras, B. Karim, A. Kim, C.M. Heaphy, A.K. Meeker, R.H. Hruban, C. A. Iacobuzio-Donahue, B. Vogelstein, Cell division rates decrease with age, providing a potential explanation for the age-dependent deceleration in cancer incidence, Proc. Natl. Acad. Sci. USA 116 (41) (2019) 20482–20488, https://doi.org/10.1073/pnas.1905722116.

[17] E. Bender, Stem-cell start-ups seek to crack the mass-production problem, Nature 597 (7878) (2021) S20–S21, https://doi.org/10.1038/d41586-021-02627-y.

[18] S. Craven, J. Whelan, B. Glennon, Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller, J. Process Control 24 (4) (2014) 344–357, https://doi.org/10.1016/j.jprocont.2014.02.007.

[19] B. Wang, Z. Wang, T. Chen, X. Zhao, Development of novel bioreactor control systems based on smart sensors and actuators, Front. Bioeng. Biotechnol. 8 (2020), https://doi.org/10.3389/fbioe.2020.00007.

[20] J. Randek, C.-F. Mandenius, On-line soft sensing in upstream bioprocessing, Crit. Rev. Biotechnol. 38 (1) (2017) 106–121, https://doi.org/10.1080/07388551.2017.1312271.

[21] P. Noll, M. Henkel, History and evolution of modeling in biotechnology: modeling & simulation, application and hardware performance, Comput. Struct. Biotechnol. J. 18 (2020) 3309–3323, https://doi.org/10.1016/j.csbj.2020.10.018.

[22] P. Sagmeister, P. Wechselberger, M. Jazini, A. Meitz, T. Langemann, C. Herwig, Soft sensor assisted dynamic bioprocess control: efficient tools for bioprocess development, Chem. Eng. Sci. 96 (2013) 190–198, https://doi.org/10.1016/j.ces.2013.02.069.

[23] M.M. Schuler, I.W. Marison, Real-time monitoring and control of microbial bioprocesses with focus on the specific growth rate: current state and perspectives, Appl. Microbiol. Biotechnol. 94 (6) (2012) 1469–1482, https://doi.org/10.1007/s00253-012-4095-z.

[24] K. Joeris, J.-G. Frerichs, K. Konstantinov, T. Scheper, CBIOT 38 (1/3) (2002) 129–134, https://doi.org/10.1023/a:1021170502775.
[25] D. Shah, M. Naciri, P. Clee, M. Al-Rubeai, NucleoCounter—an efficient technique for the determination of cell number and viability in animal cell culture processes, CBIOT 51 (1) (2006) 39–44, https://doi.org/10.1007/s10616-006-9012-9.
[26] X. Ding, Q. Zhang, W.J. Welch, Classification Beats Regression: Counting of Cells from Greyscale Microscopic Images Based on Annotation-free Training Samples, 2020, https://doi.org/10.48550/ARXIV.2010.14782 arXiv.
[27] P. Ducommun, I. Bolzonella, M. Rhiel, P. Pugeaud, U. von Stockar, I.W. Marison, On-line determination of animal cell concentration, Biotechnol. Bioeng. 72 (5) (2001) 515–522, https://doi.org/10.1002/1097-0290(20010305)72:5<515::aid-bit1015>3.0.co;2-q.
[28] R. Luttmann, D.G. Bracewell, G. Cornelissen, K.V. Gernaey, J. Glassey, V.C. Hass, C. Kaiser, C. Preusse, G. Striedner, C.-F. Mandenius, Soft sensors in bioprocessing: a status report and recommendations, Biotechnol. J. 7 (8) (2012) 1040–1048, https://doi.org/10.1002/biot.201100506.
[29] R. Mansano, E. Godoy, A. Porto, The benefits of soft sensor and multi-rate control for the implementation of wireless networked control systems, Sensors 14 (12) (2014) 24441–24461, https://doi.org/10.3390/s141224441.
[30] R.E. Madrid, C.J. Felice, Microbial biomass estimation, Crit. Rev. Biotechnol. 25 (3) (2005) 97–112, https://doi.org/10.1080/07388550500248563.
[31] J. Chen, Y. Xu, Y. Gao, L. Sun, X. Meng, K. Gu, Y. Zhang, X. Ning, A mitochondria-specific fluorescent probe for rapidly assessing cell viability, Talanta 221 (2021), https://doi.org/10.1016/j.talanta.2020.121653.
[32] P. O'Mara, A. Farrell, J. Bones, K. Twomey, Staying alive! sensors used for monitoring cell health in bioreactors, Talanta 176 (2018) 130–139, https://doi.org/10.1016/j.talanta.2017.07.088.
[33] A. Verma, M. Verma, A. Singh, Animal tissue culture principles and applications, Anim. Biotechnol. (2020) 269–293, https://doi.org/10.1016/b978-0-12-811710-1.00012-4.
[34] M. Aehle, R. Simutis, A. Lübbert, Comparison of viable cell concentration estimation methods for a mammalian cell cultivation process, CBIOT 62 (5) (2010) 413–422, https://doi.org/10.1007/s10616-010-9291-z.
[35] K.M. Desai, B.K. Vaidya, R.S. Singhal, S.S. Bhagwat, Use of an artificial neural network in modeling yeast biomass and yield of β-glucan, Process Biochem. 40 (5) (2005) 1617–1626, https://doi.org/10.1016/j.procbio.2004.06.015.
[36] O. Johnsson, D. Sahlin, J. Linde, G. Lidén, T. Hägglund, A mid-ranging control strategy for non-stationary processes and its application to dissolved oxygen control in a bioprocess, Control Eng. Pract. 42 (2015) 89–94, https://doi.org/10.1016/j.conengprac.2015.03.003.
[37] M. Aehle, A. Kuprijanov, S. Schaepe, R. Simutis, A. Lübbert, Increasing batch-to-batch reproducibility of CHO cultures by robust open-loop control, CBIOT 63 (1) (2010) 41–47, https://doi.org/10.1007/s10616-010-9320-y.
[38] A. Survyla, D. Levisauskas, R. Urniezius, R. Simutis, An oxygen-uptake-rate-based estimator of the specific growth rate in escherichia coli BL21 strains cultivation processes, Comput. Struct. Biotechnol. J. 19 (2021) 5856–5863, https://doi.org/10.1016/j.csbj.2021.10.015.
[39] G. Simmons, Differential Equations with Applications and Historical Notes, third ed., 2016, https://doi.org/10.1201/9781315371825.
[40] B. Kemesis, R. Urniezius, T. Kondratas, L. Jankauskaite, D. Masaitis, P. Babilius, Bridging functional model of arterial oxygen with information of venous blood gas: validating bioprocess soft sensor on human respiration, in: Intelligent and Safe Computer Systems in Control and Diagnostics, 2022, pp. 42–51, https://doi.org/10.1007/978-3-031-16159-9_4.
[41] A. Survyla, R. Urniezius, B. Kemesis, L. Zlatkus, D. Masaitis, V. Galvanauskas, Modeling the specific glucose consumption rate for the recombinant e.coli bioprocesses based on aging-specific growth rate, Chem. Eng. Transact. 93 (2022) 265–270, https://doi.org/10.3303/CET2293045.
[42] M. Sbarciog, D. Coutinho, A.V. Wouwer, A simple output-feedback strategy for the control of perfused mammalian cell cultures, Control Eng. Pract. 32 (2014) 123–135, https://doi.org/10.1016/j.conengprac.2014.08.002.
[43] V. Galvanauskas, R. Simutis, S.C. Nath, M. Kino-oka, Kinetic modeling of human induced pluripotent stem cell expansion in suspension culture, Regenerative Therapy 12 (2019) 88–93, https://doi.org/10.1016/j.reth.2019.04.007.
[44] C. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. 30 (2005) 79–82, https://doi.org/10.3354/cr030079.
[45] S. Haykin, Neural Networks: a Comprehensive Foundation, 1999.
[46] B. Efron, R. Tibshirani, Improvements on cross-validation: the .632+ bootstrap method, J. Am. Stat. Assoc. 92 (438) (1997), https://doi.org/10.2307/2965703.
[47] S. Schaepe, A. Kuprijanov, C. Sieblist, M. Jenzsch, R. Simutis, A. Lübbert, Current advances in tools improving bioreactor performance, CBIOT 3 (2) (2013) 133–144, https://doi.org/10.2174/22115501026661312172352 46.
[48] M. Pappenreiter, B. Sissolak, W. Sommeregger, G. Striedner, Oxygen uptake rate soft-sensing via dynamic kLa computation: cell volume and metabolic transition prediction in mammalian bioprocesses, Front. Bioeng. Biotechnol. 7 (2019), https://doi.org/10.3389/fbioe.2019.00195.
[49] J. Shiloach, R. Fass, Growing E. coli to high cell density—a historical perspective on method development, Biotechnol. Adv. 23 (5) (2005) 345–357, https://doi.org/10.1016/j.biotechadv.2005.04.004.

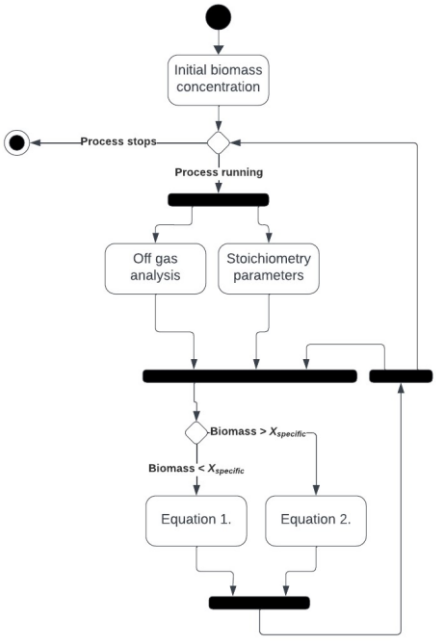
APPENDIXES

Appendix 1. Activity diagram in UML for biomass concentration estimation algorithm in the article *Generic estimator of biomass concentration for Escherichia coli and Saccharomyces cerevisiae fed-batch cultures based on cumulative oxygen consumption rate*

The process of biomass concentration estimation algorithms begins with the initial biomass concentration measurements taken at the inoculation moment. When the initial reference value is known, the estimation approach begins its iteration until the cultivation process has been stopped. The iterations consist of the following steps:

1. Off-gas analysis; we calculate the oxygen uptake rate
2. Depending on the previous or initial biomass concentration, the algorithm's equation is selected
3. We estimate the biomass concentration by using the oxygen uptake rate and stoichiometry parameters

The activity diagram in UML is presented in the following figure.

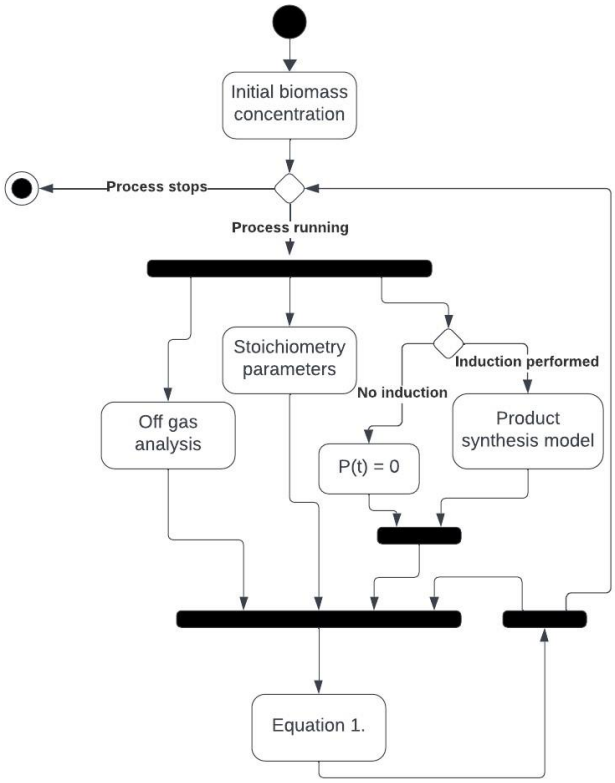


Appendix 2. Activity diagram in UML for biomass concentration estimation algorithm in the article *Identification of functional bioprocess model for recombinant E. Coli cultivation process*

The process of biomass concentration estimation algorithms begins with the initial biomass concentration measurements taken at the inoculation moment. When the initial reference value is known, the estimation approach begins its iteration until the cultivation process is stopped. The iterations consist of the following steps:

1. Off-gas analysis; we calculate the oxygen uptake rate
2. We check if induction has been performed. If yes, then, we evaluate the current protein concentration from the target product model described in Appendix 3; otherwise, protein concentration is equal to zero
3. We estimate the biomass concentration by using the oxygen uptake rate, stoichiometry parameters, and protein concentration

The activity diagram in UML is presented in the following figure.

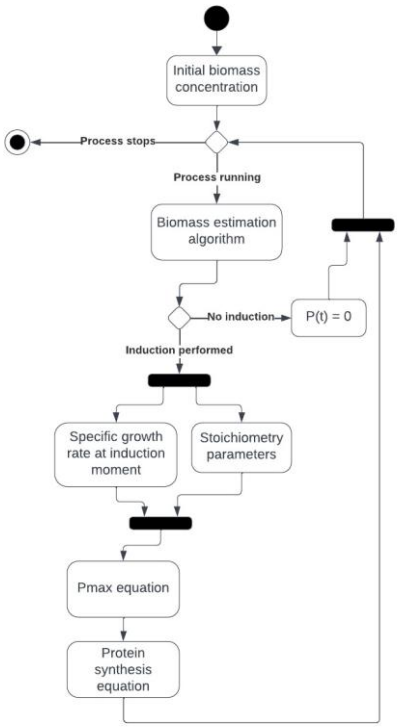


Appendix 3. Activity diagram in UML for target product model in the article *Identification of functional bioprocess model for recombinant E. Coli cultivation process*

The process of the product concentration evaluation model begins with the initial biomass concentration measurements taken at the inoculation moment. When the initial reference value is known, the evaluation model begins its iteration until the cultivation process is stopped. The iterations consist of the following steps:

1. The biomass concentration estimation procedure described in Appendix 2
2. We check if induction has been performed. If the inductor is not injected, the product concentration is equal to zero, and the next iteration is initiated. Otherwise, the current iteration continues
3. We calculate the specific growth rate at the induction moment
4. We calculate the P_{max} maximum possible protein concentration at the end of the cultivation variable
5. We calculate the current target protein concentration, and the next iteration is initiated

The activity diagram in UML is presented in the following figure.

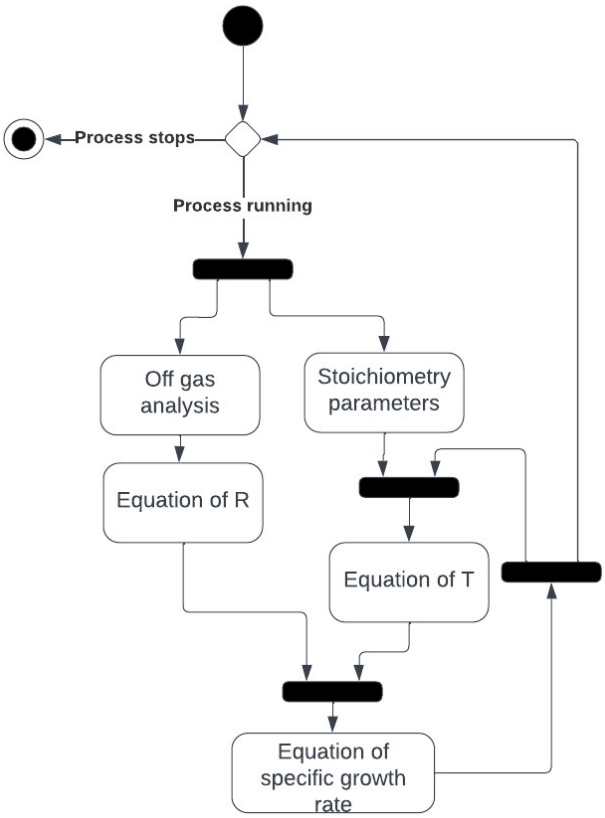


Appendix 4. Activity diagram in UML for specific growth rate estimation algorithm in the article *Oxygen uptake-rate-based estimator of the specific growth rate in Escherichia coli BL21 strains cultivation processes*

The process of the specific growth rate estimation algorithm starts with the cultivation process. This approach is running until the bioprocess is stopped. Each iteration consists of 2 steps:

1. Off-gas data analysis. We conduct the oxygen uptake rate calculation
2. We run the specific growth rate evaluation by using stoichiometry and the previous iterations of the specific growth rate value

The activity diagram in UML is presented in the following figure.

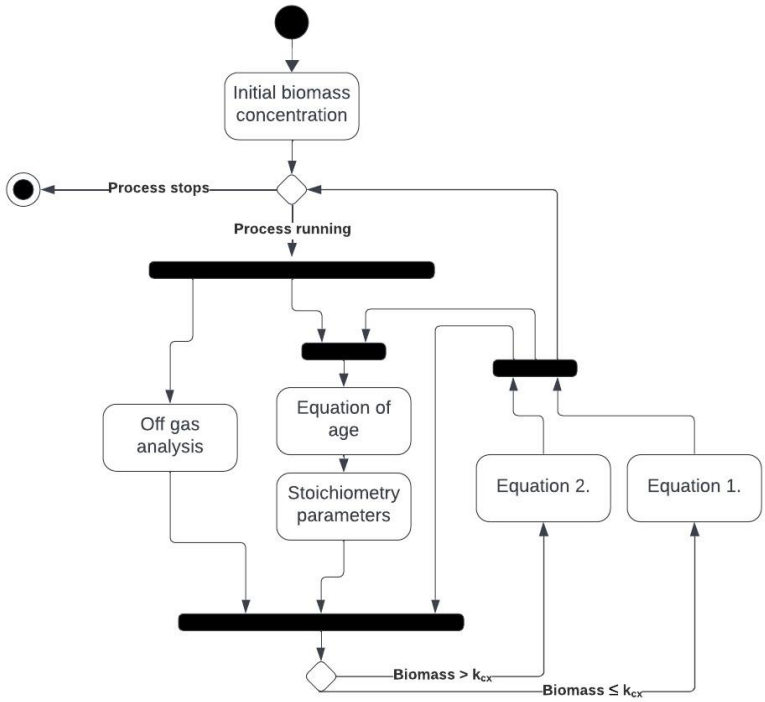


Appendix 5. Activity diagram in UML for viable cells concentration estimation algorithm in the article *Viable cell estimation of mammalian cells using off-gas-based oxygen uptake rate and aging-specific functional*

The process of viable cells concentration estimation algorithms begins with the initial biomass concentration measurements taken at the inoculation moment. When the initial reference value is known, the estimation approach begins its iteration until the cultivation process is stopped. The iterations consist of the following steps:

1. Off-gas analysis; we calculate the oxygen uptake rate
2. We evaluate the age of cell cultures
3. We adapt stoichiometry parameters regarding the cell current culture age
4. Depending on the previous or initial biomass concentration, the algorithm's equation is selected
5. We estimate the biomass concentration by using the oxygen uptake rate and stoichiometry parameters

The activity diagram in UML is presented in the following figure.



UDK 602.1:53.082.9+602.1:681.586](043.3)

SL344. 20xx-xx-xx, xx leidyb. apsk. l. Tiražas 14 egz. Užsakymas xxx.
Išleido Kauno technologijos universitetas, K. Donelaičio g. 73, 44249 Kaunas
Spausdino leidyklos „Technologija“ spaustuvė, Studentų g. 54, 51424 Kaunas