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].

$$OUR(t) = \alpha \cdot X^{'}(t) + \beta \cdot X(t); \quad (1)$$

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:

$$X(t)=\frac{X_0+\int_0^t\frac{OUR(t_1)}{\alpha(t_1)}e^{\int_0^{t_1}\frac{\beta(t_2)}{\alpha(t_2)}\mathrm{d}t_2}\mathrm{d}t_1}{e^{\int_0^t\frac{\beta(t_3)}{\alpha(t_3)}\mathrm{d}t_3}}. \tag{4}$$

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,

$$\begin{cases} X(t)=X_0+\int_0^t\frac{OUR(t_1)}{\alpha(t_1)}\,\mathrm{d}t_1; & \text{if}\quad \int_0^t X(t_1)\,\mathrm{d}t_1\le k_{cX} \\ X(t)=\dfrac{X_0+\int_0^t\frac{OUR(t_1)+\beta(t)\cdot X_{cX}}{\alpha(t_1)}e^{\int_0^{t_1}\frac{\beta(t_2)}{\alpha(t_2)}\mathrm{d}t_2}\mathrm{d}t_1}{e^{\int_0^t\frac{\beta(t_3)}{\alpha(t_3)}\mathrm{d}t_3}}, & \text{otherwise.} \end{cases} \tag{7}$$

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,

$$\alpha_i \equiv \frac{\alpha_{max}}{1 - e^{-\frac{t_i}{Lag_{time}}}} \cdot \frac{Age_{i-1}}{t_{i-1}}, \tag{16}$$

$$\beta_i = \beta \cdot \frac{Age_{i-1}}{Age_{i-1} + k_{age}}, \tag{17}$$

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.

$$\begin{cases} \widehat{X}_i = X_0 + \sum_{j=1}^{i} \frac{OUR_j}{\alpha_j} \triangle t_j; & \text{if } \sum_{j=1}^{i-1} X_j \triangle t_j \le k_{cX} \\ \widehat{X}_i = \frac{X_0 + \sum_{j=1}^{i} \frac{OUR_j + \beta_j \cdot \widehat{X_{cX}}}{\alpha_j} e^{\sum_{k=1}^{t_j} \frac{\beta_k}{\alpha_k} \triangle t_k} \triangle t_j}{e^{\sum_{j=1}^{i} \frac{\beta_j}{\alpha_j} \triangle t_j}}, & \text{otherwise,} \end{cases} \tag{18}$$

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

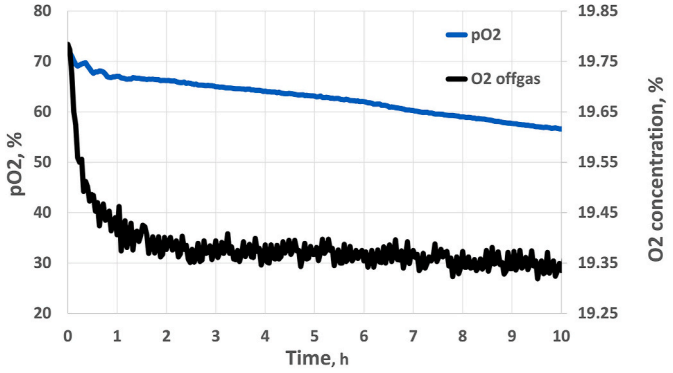

Fig. 1. The steep oxygen concentration drop at the exhaust gas outlet. The unique identity of this experiment was No 9.

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

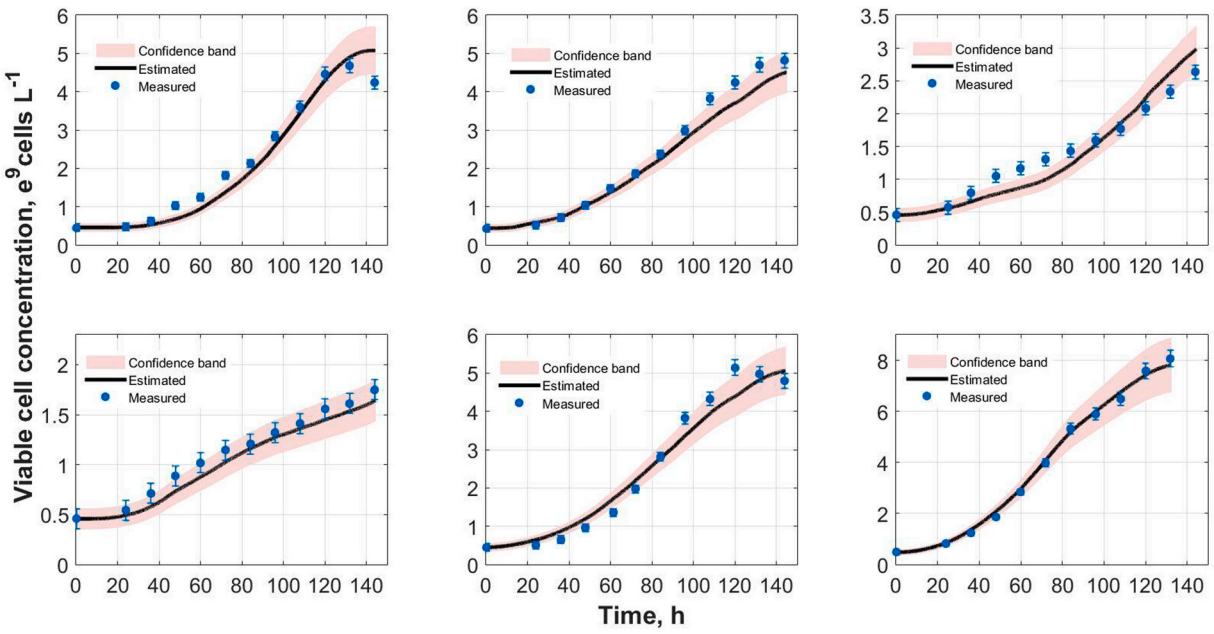

Fig. 2. CHO viable cell estimation results from experiments No 1–6. Vertical error bars indicate a total error. The purple shadow represents the prediction band. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

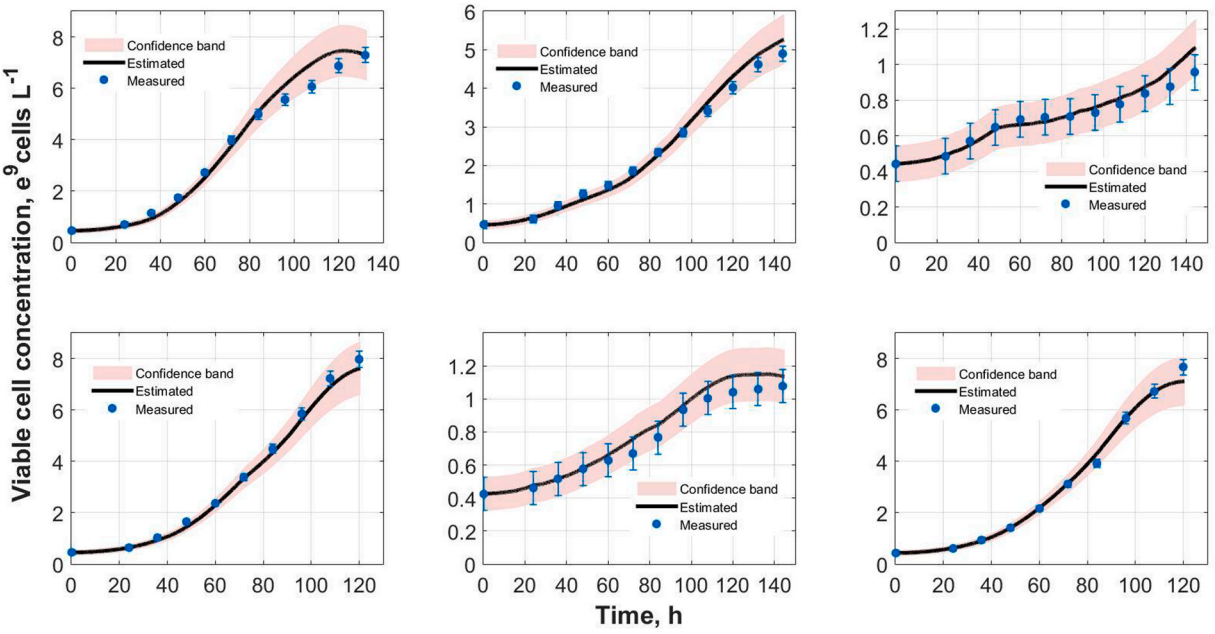

Fig. 3. CHO viable cell estimation results from experiments No 7–12. Vertical error bars indicate a total error. The purple shadow represents the prediction band. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

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

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

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

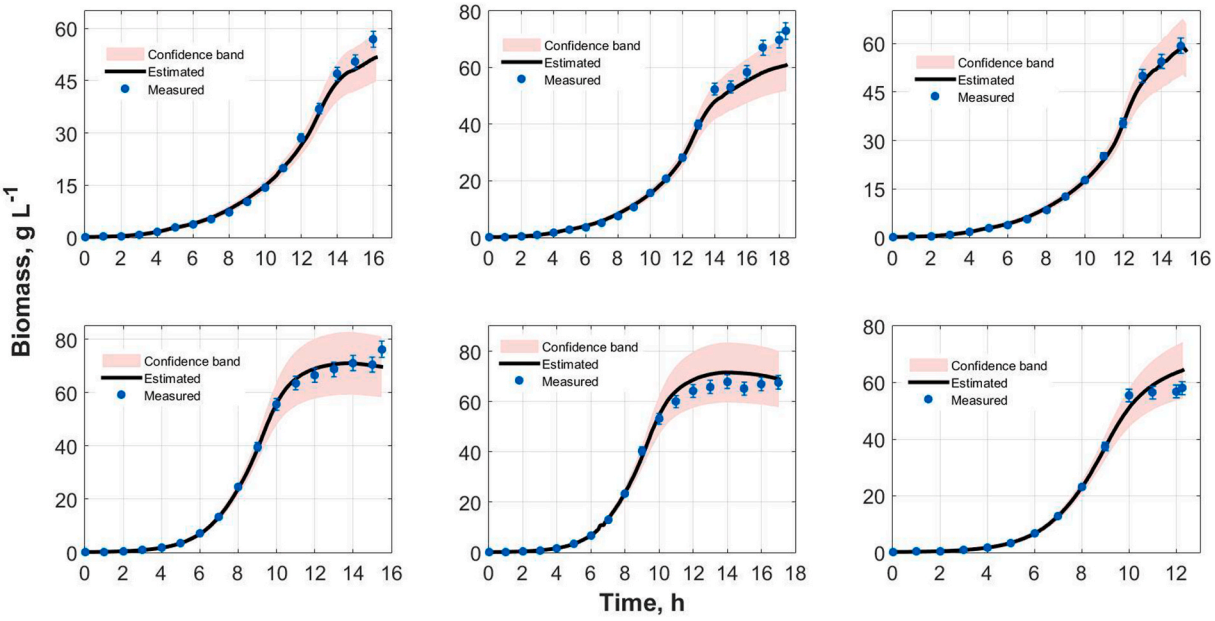

Fig. 4. *E. coli* bacteria biomass estimation results of the first six experiments. Vertical error bars indicate a total error. The purple shadow represents the prediction band. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

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/10.2174/22115501026613121723524 6.
[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.