proceedings

Proceedings

Non-Invasive Estimation of Acetates Using Off-Gas Information for Fed-Batch *E. Coli* Bioprocess †

Mindaugas Matukaitis, Deividas Masaitis *, Renaldas Urniežius, Lukas Zlatkus and Vygandas Vaitkus

Department of Automation, Kaunas University of Technology, LT-51367 Kaunas, Lithuania; mindaugas.matukaitis@ktu.lt (M.M.); renaldas.urniezius@ktu.lt (R.U.); lukas.zlatkus@ktu.lt (L.Z.); vygandas.vaitkus@ktu.lt (V.V.)

* Correspondence: deividas.masaitis@ktu.lt

† Presented at the 1st International Electronic Conference on Processes: Processes System Innovation, 17–31 May 2022; Available online: https://ecp2022.sciforum.net.

Abstract: Pharmaceutical industries widely use *Escherichia coli* cell strain to synthesize various target products. The main goal is to reach the highest possible product yield. However, throughout the cell growth stage, the formation of by-products is inevitable. Metabolic compounds such as acetates cause inhibition, particularly in later bioprocess stages. Therefore, the acetate accumulation model is necessary for planning bioprocesses to maximize cell biomass growth. The decision tree method was in possession to replicate the approach. Specific biomass growth at induction, broth weight, oxygen uptake rate, and consumed substrate weight were the inputs of model training [1,2]. Broth and consumed substrate weight had additional aging-related information incorporated as separate inputs to introduce the cumulative regularization [3].

Keywords: non-invasive; acetates; decision tree; off-gas; *E. coli*; oxygen uptake rate

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Proceedings* **2022**, *69*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Published: date

1. Introduction

Bioprocess data monitoring and control is one of bioengineers' problematic and time-consuming tasks. Bioprocesses have complex mathematical models, noisy, inconsistent data, and complicated control systems [4,5]. Offline measurements are time-delayed, require additional instruments, and are time-consuming. The development of soft sensors allows for improving and optimizing the process by facilitating real-time data collection. Process optimization leads to the primary goal of all pharmaceutical industries – to reach the highest possible product yield.

Metabolic compounds such as acetates interfere with target product synthesis by causing inhibition, particularly in later bioprocess stages. Furthermore, inhibition lengthens the lag phase leading to biomass production and growth rate loss [6]. By-product formation under aerobic conditions is mainly caused by the lack of dissolved oxygen and the imbalance between glucose uptake and its conversion to biomass [7]. To better understand these processes, real-time acetate estimation is necessary.

By-product estimation uses soft sensors that consist of numerous mathematical models [8]. These models vary from mechanistic and data-driven empirical models to hybrid models. Among the mechanistic models, the extended Kalman filter is one of the most popular approaches [9]. The EKF results are firmly related to the accuracy of the mathematical model. Therefore, hybrid models with data-driven subparts correct these inaccuracies [10,11].

This study gives an *E. coli* by-product estimation model based on gathered offline data with a black-box model. It discusses model inputs, their physical meanings, and the impact of incorporating age-related information. The novelty of this study is the proposition that off-gas analysis also carries information about the forming of by-products.

2. Materials and Methods

Data used for model training originates from fed-batch E.coli BL21(DE3) pET21-IFN-alfa-5 experiments. It consisted of 24 cultivation processes, some of which were with limited feeding. The cultivation medium throughout the experiments consisted of 46.55 g KH_2PO_4, 14 g $(NH_4)_2HPO_4$, 5.6 g $C_6H_8O_7.H_2O$, 3 mL of concentrated antifoam, 35 g $H_{14}MgO_{11}S$, and 105 gD (+) glucose monohydrate. The pressure and temperature of the system remained constant. 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.

3. Development of a Black-Box Model for Acetate Estimation

Decision trees serve for data classification and continuous data prediction [12]. Decision trees used in continuous data prediction are called regression trees. Later is used as a data-driven model in making this estimator. MATLAB software was the model development and data processing tool. Training and validation datasets were generated by dividing the sampled data from the experiments. The training dataset consisted of samples taken from 18 cultivation experiments and a validation dataset of 6.

3.1. Input Selection

Previous studies showed that specific growth rate μ is one of the best descriptors for estimating bioprocess parameters [3]. This parameter and oxygen uptake rate (OUR) carry much information about the growth and life of the cell [1,2]. The latter is estimated using a soft-sensor with off-gas information. The specific growth rate is expressed using OUR or offline-sampled biomass concentration X :

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

$$\mu = \frac{\mathrm{d}X}{\mathrm{d}t} \cdot \frac{1}{X(t)}, \tag{2}$$

where α – oxygen consumption parameter for biomass growth and the parameter β for maintenance. As an input, the specific growth rate is used only during the induction phase of the process (μ_{ind}). The value (μ_{ind}) is calculated during the induction moment using Eq. 2 formula shown above. After the induction, new biochemical reactions start, and the cells synthesize the target product [12]. During cultivations, substrate consumption affects cell development, and its inconsistent feeding may lead to cell conversion to metabolisms [13]. Substrate consumption and broth weight give model information about biomass growth, and broth weight also gives information about the dilution effect of substrate feeding. Additionally, broth and consumed substrate weight with supplementary aging-related information are separate inputs introducing the cumulative regularization [3]. The selected inputs were:

Table 1. Model inputs.

Time	h
The specific growth rate during induction μ_{ind}	1/h
Broth weight	kg
Consumed substrate weight	g
Oxygen uptake rate *OUR*	g/(h*kg)
Broth weight with age information	kg/h
Consumed substrate weight with age information	g/h

3.2. Model Errors

Model parameters fitting was based on minimizing the errors. Errors were based on measured and estimated acetates sums of squared residuals (RSS) and mean absolute error (MAE) results

$$RSS = \sum_{i=1}^{n} (A_i^* - A_i)^2, \tag{3}$$

$$MAE = \frac{\sum_{i=1}^{n} |A_i^* - A_i|}{n}, \tag{4}$$

where A_i^* is acetate *i-th* observation, and the value (A_i) is a black-box model acetate estimate. The values RSS and MAE are shown with two different sets of inputs in Table 2. The first set included time (μ_{ind}), the *OUR*, broth and consumed substrate weights, and the second set contained additional broth and consumed substrate weights with cumulative regularization. As shown in Table 2, these additional inputs improved the results.

Table 2. Model errors with age-related info and without.

Inputs	MAE	RSS
Without age-related info	0.192	1.739
With age-related info	0.155	1.264

4. Results and Discussion

Model results show that the regression tree model is applicable for estimating E. coli cell metabolic compounds in fed-batch cultivation. Figure 1 shows the difference between measured and estimated acetate values. By comparing results from inputs with age-related information (dark blue color) and inputs without age-related information (light blue color), it is clear that introducing cumulative regularization improved by-product estimation. These extra inputs resolved the errors in the first half of the bioprocess and smoothened the big spikes in the second half of the process. After bioreactor inoculation, sudden spikes in acetate estimation can be related to new biochemical processes. This phenomenon requires further, more in-depth data analysis.

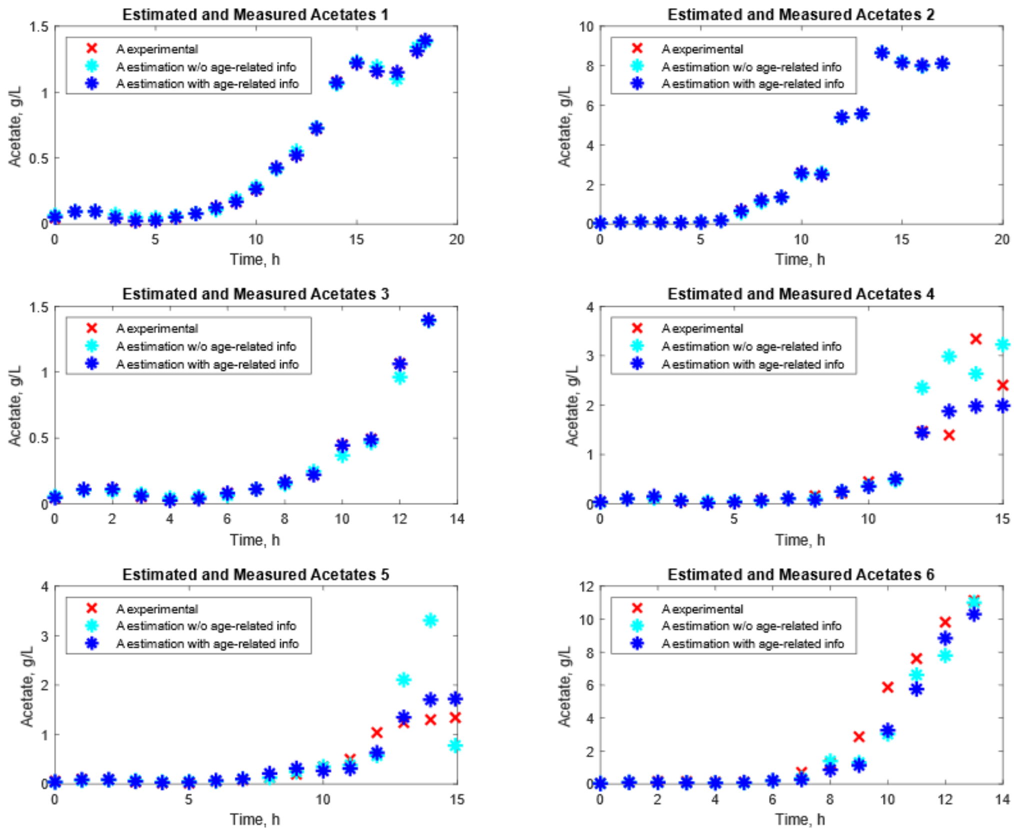

Figure 1. Measured and estimated acetates using a validation dataset.

On the other hand, by comparing this model to the traditional EKF model's lowest MAE, the black box model estimation lowered MAE by approximately 5% [9]. Also, this study used only two samples of biomass concentration, whereas EKF based model used biomass observed once every 5s. Additionally, experiments used for this model validation and training contained data involving the induction. Such an event causes a disturbance for bioprocess dynamics, making it more challenging to estimate acetates.

In the future, the proposed model will serve as the feedback that has the potential to improve the quantity and quality of a synthesized product. Altering the main parameters responsible for metabolic pathways such as substrate feed or/and oxygen transfer rate enhances the growth and well-being of the cell. Bioprocess improvements lead to easier process control and managing, providing a better workspace for future optimizations.

5. Conclusions

This study proposed an acetate estimator using the regression tree method. The model training dataset consisted of 18 cultivation experiments, and a dataset of 6 cultivation experiments validated the chosen inputs and model parameters. The regression tree model had the best results by using samples integrated with aging-related information, and it achieved satisfactory results estimating acetates reaching MAE of 0.155 and RSS of 1.264.

Funding: This project received funding from 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).

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

References

1. R. Urniezius and A. Survyla, "Identification of Functional Bioprocess Model for Recombinant E. Coli Cultivation Process," Entropy, vol. 21, no. 12, p. 1221, Dec. 2019, DOI: 10.3390/e21121221.
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, pp. 5856–5863, 2021, DOI: 10.1016/j.csbj.2021.10.015.
3. 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, Aug. 2021, DOI: 10.3390/e23081057.
4. J. Randek and C.-F. Mandenius, "Online soft sensing in upstream bioprocessing," *Critical Reviews in Biotechnology*, vol. 38, no. 1, pp. 106–121, Jan. 2018, doi: 10.1080/07388551.2017.1312271.
5. A. S. Rathore, S. Mishra, S. Nikita, and P. Priyanka, "Bioprocess Control: Current Progress and Future Perspectives," *Life*, vol. 11, no. 6, p. 557, Jun. 2021, doi: 10.3390/life11060557.
6. C. M. Takahashi, D. F. Takahashi, M. L. C. Carvalhal, and F. Alterthum, "Effects of Acetate on the Growth and Fermentation Performance of Escherichia coli KO11," *ABAB*, vol. 81, no. 3, pp. 193–204, 1999, doi: 10.1385/ABAB:81:3:193.
7. M. De Mey, S. De Maeseneire, W. Soetaert, and E. Vandamme, "Minimizing acetate formation in E. coli fermentations," *J Ind Microbiol Biotechnol*, vol. 34, no. 11, pp. 689–700, Nov. 2007, doi: 10.1007/s10295-007-0244-2.
8. H. Zhang, "Software Sensors and Their Applications in Bioprocess," in *Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control*, vol. 218, M. do Carmo Nicoletti and L. C. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 25–56. doi: 10.1007/978-3-642-01888-6_2.
9. L. Dewasme, G. Goffaux, A.-L. Hantson, and A. V. Wouwer, "Experimental validation of an Extended Kalman Filter estimating acetate concentration in E. coli cultures," *Journal of Process Control*, vol. 23, no. 2, pp. 148–157, Feb. 2013, doi: 10.1016/j.jprocont.2012.09.004.
10. L. Bárzaga-Martell, M. A. Duarte-Mermoud, F. Ibáñez-Espinel, B. Gamboa-Labbé, P. A. Saa, and J. R. Pérez-Correa, "A robust hybrid observer for monitoring high-cell density cultures exhibiting overflow metabolism," *Journal of Process Control*, vol. 104, pp. 112–125, Aug. 2021, doi: 10.1016/j.jprocont.2021.06.006.
11. H. Narayanan *et al.*, "Hybrid-EKF: Hybrid model coupled with extended Kalman filter for real-time monitoring and control of mammalian cell culture," *Biotechnology and Bioengineering*, vol. 117, no. 9, pp. 2703–2714, Sep. 2020, doi: 10.1002/bit.27437.
12. M. Xu, P. Watanachaturaporn, P. Varshney, and M. Arora, "Decision tree regression for soft classification of remote sensing data," Remote Sensing of Environment, vol. 97, no. 3, pp. 322–336, Aug. 2005, doi: 10.1016/j.rse.2005.05.008.
13. 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, Dec. 2019, doi: 10.1186/s12934-019-1241-7.
14. 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, Oct. 2018, doi: 10.3390/e20100779.