Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale, Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

A hybrid-modeling approach to monoclonal antibody production process design using automated bioreactor equipment

Kosuke Nemoto[a], Sara Badr[a], Yusuke Hayashi[a], Yuki Yoshiyama[a], Kozue Okamura[a], Mizuki Morisasa[b], Junshin Iwabuchi[b], Hirokazu Sugiyama[a]*

[a] The University of Tokyo, Department of Chemical System Engineering, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan
[b] Chitose Laboratory Co., Ltd., Tech & Biz Development Division, KSP R&D C432, 3-2-1, Sakado, Takatsu-ku, Kawasaki, Kanagawa, Japan
* Corresponding Author: sugiyama@chemsys.t.u-tokyo.ac.jp

ABSTRACT

This work presents a hybrid-modeling approach to monoclonal antibody (mAb) production processes design using automated bioreactor equipment. Experimental data covering a reasonable yet broad range of cultivation conditions was collected by the equipment. Using the data, a model applicable to a wide range of cultivation conditions was developed. In the modeling, a data-driven model was applied to describe complicated/unknown phenomena that could not be captured by previously proposed mechanistic models. In the hybrid model, while maintaining the mass balance of the mechanistic model, coefficients of the equations were estimated with random forest regression. Overall, the model could describe the dynamic concentration profiles of product mAb and quality-relevant impurities depending on the media/glucose feeding conditions. The model was then applied to determine an optimal condition that maximized product mAb concentration and satisfied the impurity constraints. The work can further support model-based design of cell cultivation processes with considering multi-input & multi-output nature.

Keywords: Process Design, Dynamic Modelling, Biosystems

INTRODUCTION

Monoclonal antibody (mAb) drugs offer advantages such as higher affinity and specificity as compared to conventional drugs for the treatment of critical diseases. Such advantages have led to a rapid growth of the mAb market, along which new technologies have been developed including host cel lines such as a high-yield CHO-MK cell line [1]. In the cultivation step, not only the final product, mAb, but also impurities such as host cell proteins (HCP), DNA, and charge variants are produced. These impurities have a significant impact on quality in the time- and resource-intensive cultivation step, making this step a major factor influencing the overall cost, time, and quality [2]. Since these impurities and mAb are affected by many cultivation conditions, such as agitation speed, dissolved oxygen (DO), feed rate, and cultivation time [3], it is essential to select conditions appropriately. Therefore, mathematical models that describe the target process based on cultivation conditions and can reduce experimental burdens are useful in the design of this critical cultivation step [4].

In modeling the cultivation step involving many complex biological phenomena, it is challenging to describe all the phenomena with a mechanistic model, making hybrid models useful. Various hybrid models for the cultivation step have been developed using different methods and for different purposes. A hybrid model of a CHO-K1 perfusion process was developed to optimize media exchange rates [5]. Hybrid models using different machine learning algorithms were compared and utilized to predict viable cells and titer [6]. A hybrid model was developed to predict critical quality attributes such as mAb and glycan profiles [7]. A hybrid model for fed-batch process with Adaptive Moment Estimation Method was developed [8]. Different hybrid models were compared by introducing the concept of the degree of hybridization [9]. A hybrid modeling approach was proposed incorporating a data-driven model to improve the predictions of lactate which is important for predicting impurities [2]. However, hybrid models in previous studies were developed for limited cultivation conditions (for example, agitation rate and DO were constant in a previous study [8]). This limitation made it challenging to describe the relationships between various conditions and impurities as well as mAb, posing a significant obstacle to comprehensive multi-input, multi-output process design.

This work presents a hybrid-modeling approach to CHO cell cultivation process design. Experimental data covering a reasonable yet broad range of cultivation conditions was collected by the equipment. Using this data, a hybrid model applicable to a wide range of cultivation conditions was developed. The hybrid model combines mass balance equations (mechanistic component) with coefficients of the equations being estimated with random forest regression (data-driven component). Overall, the developed model could describe the concentrations of product mAb and quality-relevant impurities such as charge variants, HCP and DNA depending on the media/glucose feeding conditions. The model was then applied to determine an optimal condition that maximized product mAb concentration and met the impurity constraints considering 3,840 alternatives.

EXPERIMENTAL METHODS

Figure 1 represents an overview of the experimental methods. In the experiments, a newly developed CHO-MK cell line [1], known for exhibiting extremely complex metabolic phenomena, was used. Cells were extracted from a frozen stock, thawed, and cultured in flasks for one week to increase cell density. Subsequently, cultivation was carried out with automated bioreactor equipment (Ambr® 250) which contained twelve bioreactors, each with a 250 mL volume. Initial solution volume V_0 was set to 170 mL in all experiments. Three cycles of fed-batch cultivation were performed by varying agitation speed $agitation$, DO DO, normal feed (referred to as feed hereinafter) rate F_{in}, glucose feed rate $F_{\mathrm{in,glc}}$, feed start timing t_{in}, and glucose feed start timing $t_{\mathrm{in,glc}}$. Table 1 shows the minimum and maximum values of each cultivation condition. Experiments were conducted once under 20 different conditions, and each condition was determined to cover a reasonable yet broad range to extend the interpolation range of the model. During the experiments, feed and glucose feed were supplied at a constant rate. Items such as DO and agitation speed were measured online every few seconds. In addition, multiple items including viable cells, cell viability, mAb, metabolites, and impurities were measured offline as time series data once or twice per day.

Table 1: Minimum and maximum values of each cultivation condition.

Process parameters	Minimum	Maximum
t_{in} [h]	24	42
$t_{\mathrm{in,glc}}$ [h]	42	42
F_{in} [V_0 day^{-1}]	0.08	0.1
$F_{\mathrm{in,glc}}$ [g L^{-1} day^{-1}]	4.8	18.8
$agitation$ [rpm]	700	1400
DO [%]	20	50

MODEL DEVELOPMENT

Mechanistic model equations

First, mechanistic model equations were formulated based on model equations proposed in previous studies [10,11]. To develop a model applicable to a wide range of cultivation conditions, the equations were simplified as much as possible. The components included density/concentrations of viable cell, dead cell, glucose, lactate, mAb, main/acidic charge variant, HCP, and DNA as shown in Eqs. (1)–(9) in Table 2. In the equations, V [L], X_{v} [cells L^{-1}] , X_{d} [cells L^{-1}], $[\mathrm{GLC}]$ [mmol L^{-1}], $[\mathrm{LAC}]$ [mmol L^{-1}], P [g L^{-1}], $[\mathrm{Main}]$ [g L^{-1}], $[\mathrm{Acid}]$ [g L^{-1}], $[\mathrm{HCP}]$ [g L^{-1}], $[\mathrm{DNA}]$ [g L^{-1}] are the solution volume, density/concentrations of the viable/dead cells, glucose, lactate, mAb, main/acidic charge variant, HCP, and DNA. $F_{\mathrm{in}}/F_{\mathrm{in,glc}}$ [L h^{-1}] and $c_{\mathrm{in\_glc}}/c_{\mathrm{in,glc}}$ [mmol L^{-1}] are normal/glucose feed flow rate and glucose concentration of feed/glucose feed, respectively. μ [h^{-1}] and μ_{d} [h^{-1}] are specific cell growth and death rates, respectively. Q_{p} [g cell^{-1} h^{-1}] and Q_{lac} [mmol

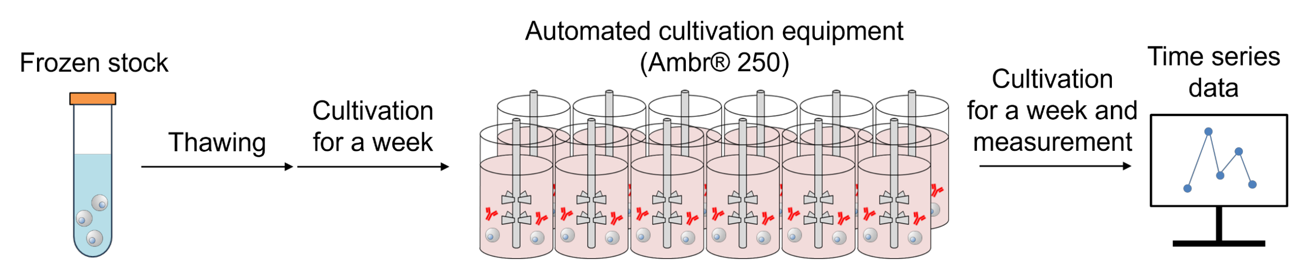

Figure 1: Overview of the cultivation experiments.

Table 2: Mechanistic model equations.

$$\frac{d(VX_\mathrm{v})}{dt} = (\mu - \mu_\mathrm{d})VX_\mathrm{v} \tag{1}$$

$$\frac{d(VX_\mathrm{d})}{dt} = \mu_\mathrm{d}VX_\mathrm{v} - k_{X_\mathrm{d},\mathrm{lysis}}VX_\mathrm{d} \tag{2}$$

$$\frac{d(V[\mathrm{GLC}])}{dt} = -\left(\frac{\mu-\mu_\mathrm{d}}{Y_{X_\mathrm{v}/\mathrm{glc}}} + m_\mathrm{glc}\right)VX_\mathrm{V} + F_\mathrm{in}c_\mathrm{in\_glc} + F_\mathrm{in,glc}c_\mathrm{in,glc} \tag{3}$$

$$\frac{d(V[\mathrm{LAC}])}{dt} = Q_\mathrm{lac}VX_\mathrm{v} \tag{4}$$

$$\frac{d(VP)}{dt} = \begin{cases} Q_\mathrm{p,a}VX_\mathrm{v}\big|_{[\mathrm{GLC}]\geq[\mathrm{GLC}]_\mathrm{threshold}} \\ Q_\mathrm{p,b}VX_\mathrm{v}\big|_{[\mathrm{GLC}]<[\mathrm{GLC}]_\mathrm{threshold}} \end{cases} \tag{5}$$

$$\frac{d(V[\mathrm{Main}])}{dt} = \frac{d(VP)}{dt} - (k_\mathrm{acid} + k_\mathrm{base})V[\mathrm{Main}] \tag{6}$$

$$\frac{d(V[\mathrm{Acid}])}{dt} = k_\mathrm{acid}V[\mathrm{Main}] \tag{7}$$

$$\frac{d(V[\mathrm{DNA}])}{dt} = Y_{\mathrm{dna}/X_\mathrm{d}}k_{X_\mathrm{d},\mathrm{lysis}}VX_\mathrm{d} - k_\mathrm{dna,lysis}V[\mathrm{DNA}] \tag{8}$$

$$\frac{d(V[\mathrm{HCP}])}{dt} = Y_{\mathrm{hcp}/X_\mathrm{v}}VX_\mathrm{v} + Y_{\mathrm{hcp}/X_\mathrm{d}}k_{X_\mathrm{d},\mathrm{lysis}}VX_\mathrm{d} - k_\mathrm{hcp,lysis}(V[\mathrm{HCP}])^\mathrm{n} \tag{9}$$

cell^{-1} h^{-1}] are specific production rates of mAb and lactate, respectively, with Q_p being classified based on whether glucose is depleted or not. $Y_{X_\mathrm{v}/\mathrm{glc}}$ [cells mmol^{-1}] and m_glc [mmol cell^{-1} h^{-1}] are specific glucose consumption for cell growth and cell maintenance, respectively. $k_{i,\mathrm{lysis}}$ is the rate constant of the dead cell lyses and the impurity dissolution, where i includes dead cell, HCP, and DNA. k_acid and k_base are the rate constants for changes of main charge variants into acidic and basic charge variants. $Y_{i/X_\mathrm{v},X_\mathrm{d}}$ [mg cell^{-1}] is the mass of impurity i released from a viable cell and a lysed dead cell, where i includes HCP and DNA.

Parameter estimation of the mechanistic model

Time series data from 20 experiments across three cycles were used for parameter estimation, and a part of the results are shown in Figure 2. The experimental data indicated that lactate repeated cycles of production and consumption, and viable cell density saturated in the later stages. However, due to the constant Q_lac previously addressed for improvement [2], and the exponential form of the viable cell density equation, the mechanistic model could not capture the behavior of lactate and viable cell density.

Hybrid modeling

To address the issues mentioned above, a hybrid model was developed. Utilizing a hybrid model can contribute to overcoming the challenge of fully elucidating the biological phenomena causing lactate shifts and cell growth saturation, and to avoid increasing the number of estimated parameters. The hybrid model, whose overview is shown in Figure 3, maintained the mass balance of the mechanistic model while estimating the coefficients $\mu - \mu_\mathrm{d}$ and Q_lac with a data-driven model as shown in Eq. (10).

$$Coef = f\left(t, V, t_\mathrm{in}, t_\mathrm{in,glc}, F_\mathrm{in}, F_\mathrm{in,glc}, agitation, DO\right) \tag{10}$$

where $Coef$ represents $\mu - \mu_\mathrm{d}$ and Q_lac. Since the values of Q_lac and $\mu - \mu_\mathrm{d}$ change dynamically, the model inputs included not only constant experimental conditions (agitation speed, DO, feed/glucose feed rate, and feed/glucose feed start timing), but also dynamically changing

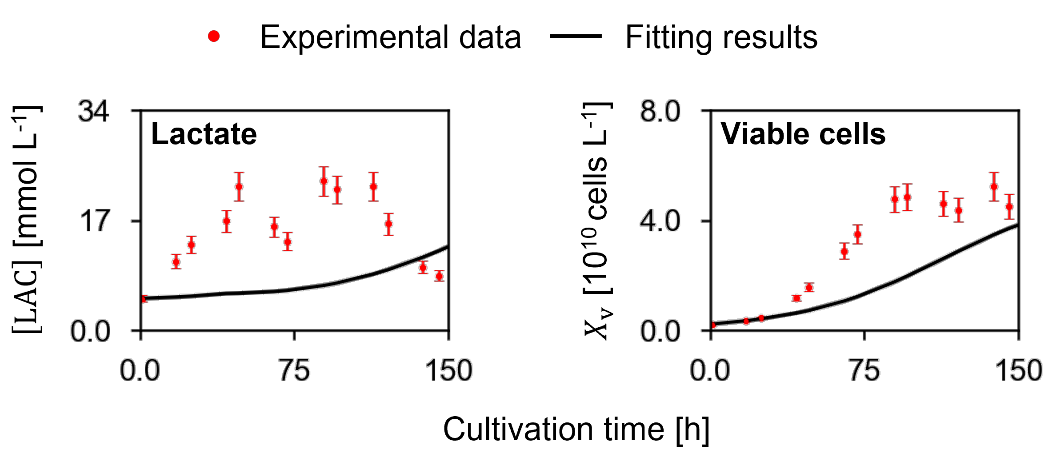

Figure 2: A part of the modeling results based on the mechanistic model.

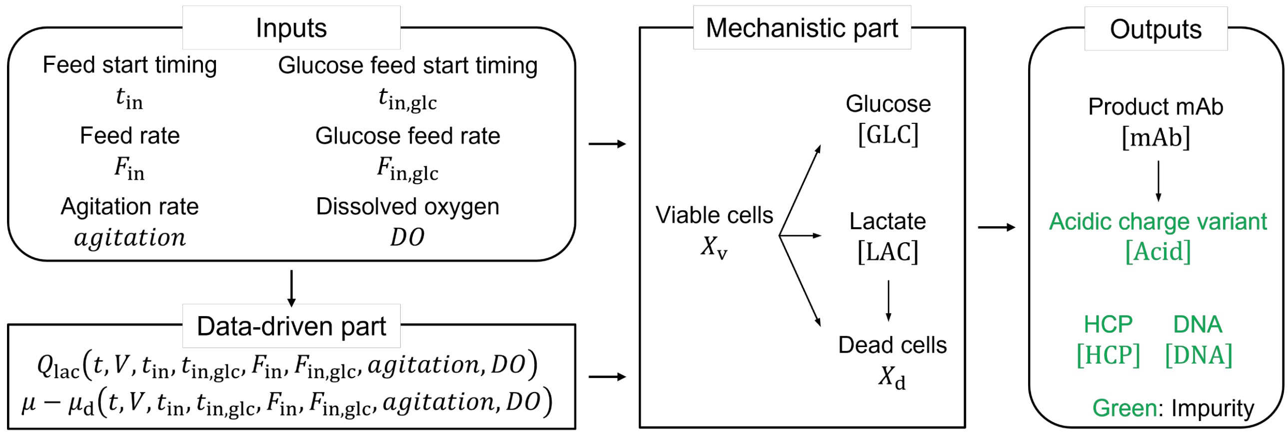

Figure 3: Overview of the developed hybrid model.

cultivation time and solution volume. Random forest regression was selected as a method for the data-driven model due to its high accuracy. In model development, experimental data were interpolated to compensate for the limited amount of data, and the values of $\mu - \mu_{\text{d}}$ and Q_{lac} were calculated based on the interpolated values using Eqs. (1) and (4), respectively. To avoid overfitting, a portion of the calculated values was extracted, and the model was developed based on the values. Time series data from 20 experiments across three cycles were used for parameter estimation, and a part of the results are shown in Figure 4. Table 3 presents the average values

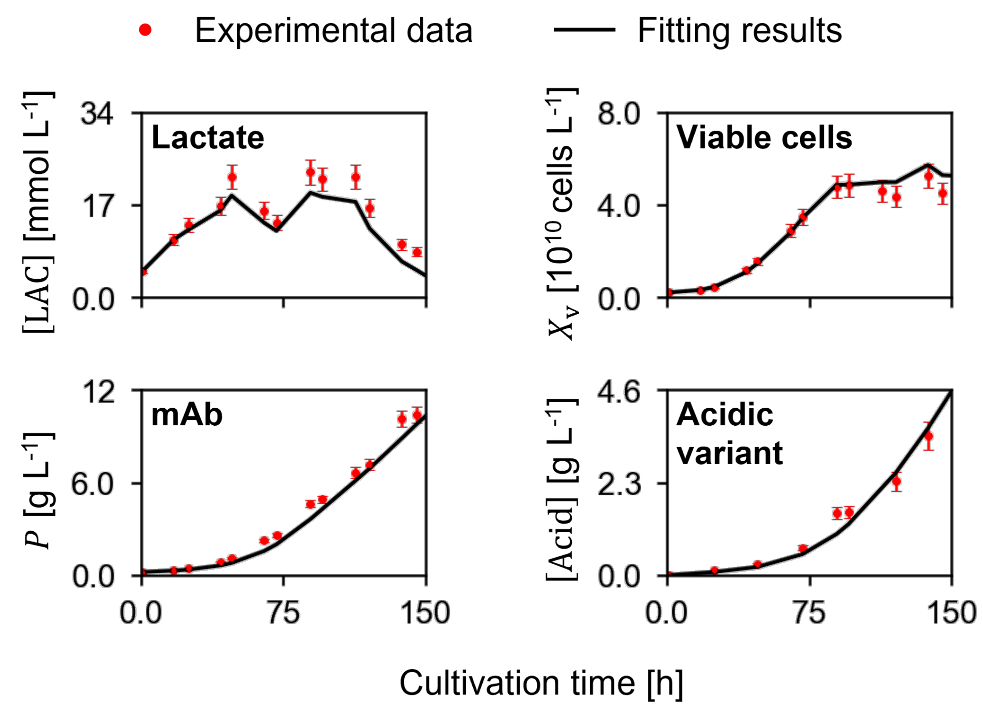

Figure 4: A part of the modeling results based on the hybrid model.

Table 3: Average values of R^2 and NRMSE for each component.

	X_{v}	X_{d}	[GLC]	[LAC]	P	[Main]	[Acid]	[HCP]	[DNA]
R^2	0.99	0.70	0.85	0.86	0.95	0.92	0.92	0.68	0.69
NRMSE	0.035	0.16	0.10	0.10	0.078	0.099	0.091	0.19	0.18

of R^2 and NRMSE for all training datasets used in the estimation for each component. The results indicated that high R^2 values and low NRMSE values were obtained for all components, showing that the model could capture not only the behavior of mAb but also impurities across a wide range of cultivation conditions.

MODEL APPLICATION

A model-based assessment of cultivation conditions was conducted by using the developed hybrid model. In the assessment, a design problem was formulated to maximize mAb while keeping impurity levels low as shown in the following equation:

$$\max\ [\mathrm{mAb}](t, t_{\mathrm{in}}, t_{\mathrm{in,glc}}, F_{\mathrm{in}}, F_{\mathrm{in,glc}}, agitation, DO) \qquad (11)$$

subject to

$$\begin{array}{ccc}
t\ [\mathrm{h}] & \le & 145 \\
[\mathrm{HCP}]\ [\mathrm{g\ L^{-1}}] & \le & 2.2 \\
[\mathrm{DNA}]\ [\mathrm{g\ L^{-1}}] & \le & 0.42 \\
[\mathrm{Acid}]/P\ [\%] & \le & 40 \\
t_{\mathrm{in}}\ [\mathrm{h}] & \in & \{42, 48\} \\
t_{\mathrm{in,glc}}\ [\mathrm{h}] & \in & \{42, 48\} \\
F_{\mathrm{in}}\ [V_0\ \mathrm{day^{-1}}] & \in & \{0.08, 0.1\} \\
F_{\mathrm{in,glc}}\ [\mathrm{g\ L^{-1}\ day^{-1}}] & \in & \{5, 6, \dots, 17, 18, 18.8\} \\
Agitation\ [\mathrm{rpm}] & \in & \{700, 800, \dots, 1300, 1400\} \\
DO\ [\%] & \in & \{20, 30, 40, 50\}
\end{array}$$

where t [h] represents cultivation time. In the design problem, impurity values in the constraint were set according to expert opinions. Based on the design problem, 3,840 cultivation conditions were explored, with results evaluated at each one-hour interval of cultivation time. Table 4 shows the top three conditions with the highest mAb values satisfying the constraints, along with the corresponding cultivation time.

Figure 5 shows the simulation results for the top three conditions. In the simulations, glucose levels remained low but non-depleted until the later stages of the cultivation and only became depleted at the end of the cultivation. Glucose is widely recognized as a critical factor for both cell growth and mAb production; however, glucose levels are proportional to osmolarity, and high osmolarity is known to have a negative impact on cell growth [12]. Therefore, it was found that setting conditions to maintain appropriate glucose levels was important. Furthermore, the simulation results revealed that both mAb and impurities increased over time, indicating that setting an appropriate cultivation duration was crucial to obtaining more mAb while keeping impurity levels low.

Table 4: Top three conditions with the highest mAb values satisfying the constraints.

Rank	t_{in}	$t_{\mathrm{in,glc}}$	F_{in}	$F_{\mathrm{in,glc}}$	$Agitation$	DO	t
1	48	42	0.1	15	700	40	135
2	48	42	0.1	15	1000	50	135
3	42	48	0.1	14	700	40	136

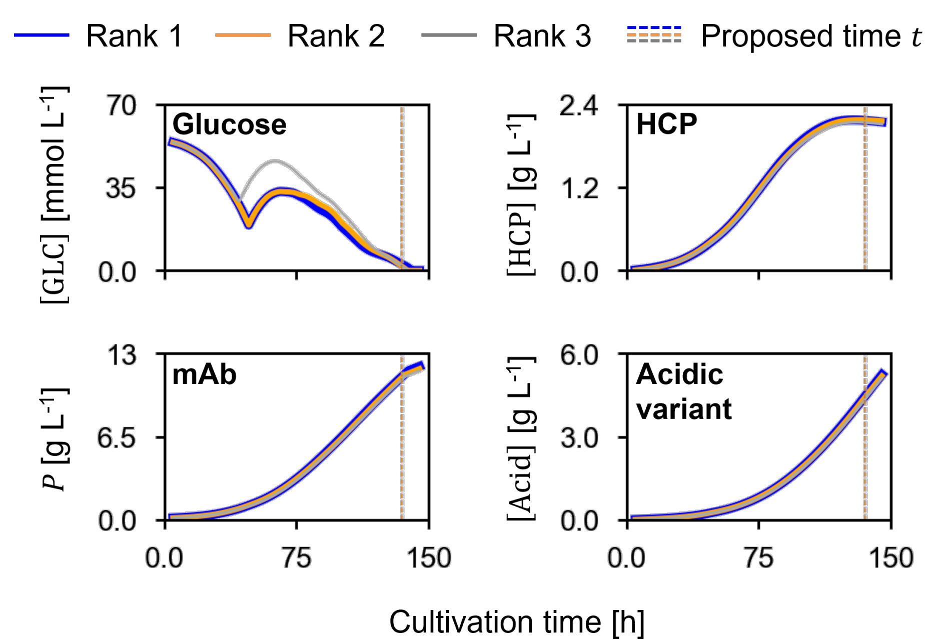

Figure 5: Simulation results for the top three conditions.

CONCLUSIONS AND OUTLOOK

This work presented a hybrid-modeling approach to mAb production process design. Using automated bioreactor equipment enabled the acquisition of experimental data that covers a reasonable yet broad range. Using this data, a model applicable to a wide range of cultivation conditions was developed. In the hybrid model, the coefficients of the mechanistic model were estimated with a data-driven model, while maintaining the mass balance of the systems components. In the case study, the model enabled determination of a cultivation condition that maximized the product mAb concentration with meeting the impurity constraints out of 3,840 candidates. This work supports model-based design of cell cultivation processes with considering the multi-input & multi-output nature of the process. The future work includes expansion to dynamic optimization, operation support, and integration with purification processes.

ACKNOWLEDGEMENTS

This work was supported by the Japan Agency for Medical Research and Development (AMED) under Grant Number JP21ae0121015.

REFERENCES

1. Masuda, K., Kubota, M., Nakazawa, Y., Iwama, C., Watanabe, K., Ishikawa, N., Tanabe, Y., Kono, S., Tanemura, H., Takahashi, S., Makino, T., Okumura, T., Horiuchi, T., Nonaka, K., Murakami, S., Kamihira, M., Omasa, T. Establishment of a novel cell line, CHO-MK, derived from Chinese hamster ovary tissues for biologics manufacturing. *J. Biosci. Bioeng.* 137:471-479 (2024) https://doi.org/10.1016/j.jbiosc.2024.02.005
2. Okamura, K., Badr, S., Murakami, S., Sugiyama, H. Hybrid Modeling of CHO Cell Cultivation in Monoclonal Antibody Production with an Impurity Generation Module. *Ind. Eng. Chem. Res.* 61:14898-14909 (2022) https://doi.org/10.1021/acs.iecr.2c00736
3. Ha, T.K., Kim, D., Kim, C.L., Grav, L.M., Lee, G.M. Factors affecting the quality of therapeutic proteins in recombinant Chinese hamster ovary cell culture. *Biotechnol Adv.* 54:107831 (2022) https://doi.org/10.1016/j.biotechadv.2021.107831
4. Badr, S., Sugiyama, H. A PSE perspective for the efficient production of monoclonal antibodies: integration of process, cell, and product design aspects. *Curr. Opin. Chem. Eng.* 27:121-128 (2020) https://doi.org/10.1016/j.coche.2020.01.003
5. Agarwal, P., McCready, C., Ng, S.K., Ng, J.C., van de Laar, J., Pennings, M., Zijlstra, G. Hybrid modeling for in silico optimization of a dynamic perfusion cell culture process. *Biotechnol. Prog.* (2024) https://doi.org/10.1002/btpr.3503
6. Yatipanthalawa, B.S., Wallace Fitzsimons, S.E., Horning, T., Lee, Y.Y., Gras, S.L. Development and validation of a hybrid model for prediction of viable cell density, titer and cumulative glucose consumption in a mammalian cell culture system. *Comput. Chem. Eng.* 184:108648 (2024) https://doi.org/10.1016/j.compchemeng.2024.108648
7. Polak, J., Huang, Z., Sokolov, M., von Stosch, M., Butté, A., Hodgman, C., Borys, M., Khetan, A. An innovative hybrid modeling approach for simultaneous prediction of cell culture process dynamics and product quality. *Biotechnol. J.* 19:2300473 (2024) https://doi.org/10.1002/biot.202300473
8. Pinto, J., Ramos, J.R.C., Costa, R.S., Rossell, S., Dumas, P., Oliveira, R. Hybrid deep modeling of a CHO-K1 fed-batch process: combining first-principles with deep neural networks. *Front. Bioeng. Biotechnol.* 11: 1237963 (2023) https://doi.org/10.1016/j.coche.2020.01.003
9. Narayanan, H., Luna, M., Sokolov, M., Butté, A., Morbidelli, M. Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Cell Culture Processes. *Ind. Eng. Chem. Res.* 61:8658-8672 (2022) https://doi.org/10.1021/acs.iecr.1c04507
10. Xing, Z., Bishop, N., Leister, K., Li, Z.J. Modeling kinetics of a large-scale fed-batch CHO cell culture by markov chain monte carlo method. *Biotechnol. Prog.* 26:208-219 (2010) https://doi.org/10.1002/btpr.284
11. Badr, S., Okamura, K., Takahashi, N., Ubbenjans, V., Shirahata, H., Sugiyama, H. Integrated design of biopharmaceutical manufacturing processes: Operation modes and process configurations for monoclonal antibody production. *Comput. Chem. Eng.* 153:107422 (2021) https://doi.org/10.1016/j.compchemeng.2021.107422
12. Alhuthali, S., Kotidis, P., Kontoravdi, C., Osmolality effects on cho cell growth, cell volume, antibody productivity and glycosylation. *Int. J. Mol. Sci.* 22:3290 (2021) https://doi.org/10.3390/ijms22073290