Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 59-1 (2025) 391–396

Digital twin development of yeast fed-batch cultures for vaccine production

Micaela Benavides *,** **Pascal Gerkens** ** **Gaël de Lannoy** **
Laurent Dewasme * **Alain Vande Wouwer** *

* *Systems, Estimation, Control and Optimization (SECO), University of Mons, 7000 Mons, Belgium, (e-mail:alain.vandewouwer@umons.ac.be)*

** *GSK, 1330 Rixensart, Belgium*

Abstract: This paper reports on designing a digital twin of a yeast culture process in an industrial context. Using a rich experimental data set corresponding to different input profiles, an original dynamic model based on the assumption of cell overflow metabolism is derived. The kinetic laws are represented by smooth functions combining Monod and Jerusalimski factors, which model rate activation and inhibition, respectively. Parameter estimation is achieved following an iterative procedure including practical identifiability analyses which allow the design of informative experiments, enhancing parameter precision and accuracy.

Keywords: Modeling, Parameter estimation, Biotechnology, Yeast cultures

1. INTRODUCTION

Digital twins are now increasingly popular in the biopharmaceutical sector and have renewed the interest in mathematical modeling in all possible forms, from standard physics-based models to machine learning paradigms (Amribt et al., 2013; von Stosch et al., 2014; Retamal et al., 2018). For developing dynamic models of cultures in bioreactors, physics-based models, or more precisely, macroscopic biology-inspired models based on mass balance ordinary differential equations are still the solution of choice given their relatively small number of parameters. Indeed, collecting large amounts of experimental data remains time- and money-expensive, hampering the use of machine learning models. While Raman or near-infrared probes combined with chemometric models can provide more information on the evolution of the culture metabolites (Yousefi-Darani et al., 2021), conducting experiments under various operating conditions specifically designed to unveil the dynamic models remains uncommon.

The primary focus of this article is to present a novel dynamic model for a vaccine production bioprocess, which is based on a genetically modified yeast strain. The challenge begins with creating an appropriate model structure, followed by identifying the unknown model parameters from experimental data. The yeast strain's metabolic behavior is characterized by overflow metabolism (Crabtree, 1929), and our approach is distinct in that we avoid the use of switches in the kinetic model (Sonnleitner and Käppeli, 1986; Dewasme et al., 2011; Retamal et al., 2018), opting instead for a combination of modulation factors that represent activation and inhibition effects, thereby enabling continuous transitions between metabolic modes.

Even though rigorous experiment design procedures have been proposed in the literature (for instance, based on the determinant or the condition number of the Fisher Information Matrix - FIM (Telen et al., 2012)), they usually lead to computationally expensive optimization, which can become unpractical as the number of parameters increases (Bhonsale et al., 2022). Hence, in this study, we follow a simple, pragmatic approach, where, starting from one culture run, a first model structure and its parameters are deduced, as well as various information, including a posteriori sensitivity analysis, FIM, and confidence intervals, as well as validation results. Next, a second experiment is suggested, and the additional information is used to question the previous proposal and to hypothesize on different structural and/or parametric changes. The reasoning is pursued along this route, leading to the realization of a limited number of experiments, typically 4-5 bioreactor runs.

In our specific case study, the achievement amounts to being quite satisfactory. It offers a relatively parsimonious dynamic model of the cultures that could be further used for the design of state observers (software sensors, Bogaerts and Vande Wouwer (2003)) or model-based controllers (Abadli et al., 2022), opening up new possibilities in the biopharmaceutical sector. This paper aims to present and discuss this case study, highlighting the pitfalls and positive findings in a real-industrial context.

This paper is organized as follows. The next section is dedicated to the presentation of the modeling methodology and section 3 describes the proposed dynamic model. Section 4 presents the experimental setup and the data sets that were used to identify the model in section 5. Model validations are discussed in section 6 and conclusions are drawn in section 7.

2. METHODOLOGY

The iterative methodology of Banga and Balsa-Canto (2008) has been followed to develop a mechanistic model

10.1016/j.ifacol.2025.03.067

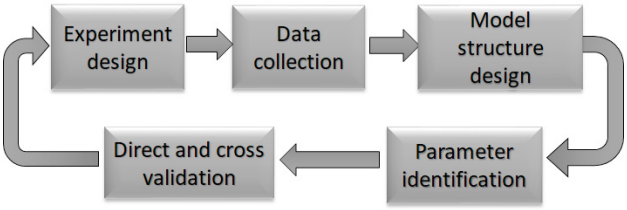

Fig. 1. Methodology to obtain the model

of yeast metabolism, explaining biomass growth, substrate consumption, and ethanol production by fermentation as well as its possible consumption. This cyclic procedure is illustrated in Figure 1. It starts with a model structure proposal based on culture data set analysis. Then, the model parameters are estimated and evaluated via their confidence intervals, ensuring their practical identifiability through *a posteriori* analysis. The latter analysis allows verifying the uniqueness of the solution of the corresponding optimization problem (Balsa-Canto et al., 2010).

The parameter identification procedure involves the use of optimizers to achieve the data fitting. The model, along with its estimated parameters, must not only be validated using the initial experiment (direct validation) but also through new experiments (cross-validation). These steps are usually assessed using a root mean square error (RMSE) criterion to characterize the fitting accuracy and the analysis of the FIM to obtain parameter confidence intervals that indicate the model precision. If the RMSE and the parameter confidence interval results do not meet some expected levels, new experiments are proposed to enrich the data set and better highlight the kinetic mechanisms that were not sufficiently well identified. Moreover, before suggesting a new experiment, a sensitivity analysis for each parameter will be conducted to evaluate whether the new experimental data will improve the identification of that parameter. The cycle is then repeated conformly to Figure 1. This paper presents only the results from the final cycle, which includes four experiments.

3. MODEL STRUCTURE DESIGN

The dynamic model is inspired by the bottleneck assumption of Sonnleitner and Käppeli (1986). This concept suggests that yeast metabolism is constrained by its limited respiratory capacity and that any substrate overflow leads to ethanol formation by fermentation considered as a "short-term Crabtree effect" (Crabtree, 1929). The proposed reaction scheme involves three macroreactions and follows the same scheme as described in (Benavides et al., 2024). The first reaction involves the oxidation of the substrate, the second reaction describes the respiro-fermentative pathway following substrate overflow, and the third reaction considers the potential consumption of ethanol, as follows:

$$\text{Substrate oxidation} : S \xrightarrow{r_1 X} k_1 X \tag{1a}$$

$$\text{Substrate overflow} : S \xrightarrow{r_2 X} k_2 E + k_4 X \tag{1b}$$

$$\text{Ethanol oxidation} : E \xrightarrow{r_3 X} k_3 X \tag{1c}$$

where S, X, and E, respectively, stand for substrate, biomass, and ethanol concentrations.

Following the bottleneck assumption, yeast metabolism is likely to take two main pathways depending on the available amount of substrate (glucose) and the corresponding respiratory capacity. When the respiratory capacity is not fully saturated, the yeast operates in a respirative regime, and the remaining capacity can be used to oxidize ethanol, activating reactions (1a) and (1c) while (1b) is assumed to be negligible. However, if the respiratory capacity becomes saturated, the excess of substrate which is not oxidized enters the fermentation pathway, producing ethanol. The corresponding fermentation regime considers the activations of only reactions (1a) and (1b) while (1c) is negligible.

In this study, we propose using a nonlinear continuous model based on Monod activation/saturation kinetics (Monod, 1949) complemented by inhibition factors (Jerusalimski and Engamberdiev, 1969). This approach differs from Sonnleitner and Käppeli (1986), who suggest discontinuous switching kinetics, for instance formulated in (Dewasme et al., 2011; Richelle et al., 2014; Retamal et al., 2018; Huet et al., 2022). Moreover, in contrast to the study outlined in (Benavides et al., 2024), we incorporate an additional inhibitory factor attributed to the presence of ethanol in the second reaction. This factor was included after restructuring the model based on data obtained from new experiments, following the cycling methodology as depicted in Figure 1.

Each reaction presents a rate of the following form:

$$r_1 = \mu_{m1} \cdot \frac{S}{K_{S1} + S} \cdot \frac{1}{1 + \frac{X}{K_{IX}}} \cdot \frac{1}{1 + \frac{E}{K_{IE}}} \tag{2a}$$

$$r_2 = \mu_{m2} \cdot \frac{S}{K_{S2} + S} \cdot \frac{1}{1 + \frac{X}{K_{IX}}} \cdot \frac{1}{1 + \frac{E}{K_{IE2}}} \tag{2b}$$

$$r_3 = \mu_{m3} \cdot \frac{E}{K_E + E} \cdot \frac{1}{1 + \frac{X}{K_{IX}}} \cdot \frac{1}{1 + \frac{S}{K_{IS}}} \tag{2c}$$

The first reaction rate, denoted r_1 (2a), is activated by a Monod factor related to glucose uptake which is limited by the available respiratory capacity. The main inhibitory components are biomass density and ethanol concentration (respectively in the second and third factors). The second reaction rate r_2 (2b) is also governed by a Monod factor related to glucose uptake and is similarly inhibited by biomass and ethanol. However, the kinetic coefficients are assumed to differ from (2a) since the rates do not operate through the same pathways. The third reaction rate r_3 (2c) is activated by the presence of ethanol under Monod kinetics. This reaction rate also includes a respiratory capacity inhibition factor, similar to the factors in equations (2a) and (2b), which are related to biomass accumulation. Additionally, another inhibition factor is included to address the preferential consumption of the substrate.

This rate is activated when glucose, the primary substrate, becomes exhausted. When applying mass balance to each macroreaction, the following ordinary differential equation system is obtained:

$$\frac{dX}{dt} = (k_1 \cdot r_1 + k_3 \cdot r_3 + k_4 \cdot r_2) \cdot X - D \cdot X \tag{3a}$$

$$\frac{dS}{dt} = -(r_1 + r_2) \cdot X - D \cdot S + D \cdot S_{in} \tag{3b}$$

$$\frac{dE}{dt} = (k_2 \cdot r_2 - r_3) \cdot X - D \cdot E \tag{3c}$$

$$\frac{dV}{dt} = D \cdot V = F_{in} \tag{3d}$$

where S_{in} represents the glucose concentration in the inlet feed, $D = F_{in}/V$ is the dilution rate, F_{in} is the inlet feed flow rate and V the bioreactor volume.

4. MATERIALS AND METHODS

A recombinant yeast strain of *Saccharomyces cerevisiae* was used in this work. Each culture starts with an initial bioreactor volume of 5.5 liters and a stirrer speed set to 260 rpm. The temperature is maintained at 30°C, while the pH is regulated at 5 using a base solution. The bioreactor is equipped with an in-line pO2 sensor that measures dissolved oxygen in percentage, a stirrer motor controlled to maintain the pO2 above 60%, and a peristaltic pump that controls the F_{in}. For offline measurements, approximately 5 mL samples were taken every hour using an automated cell culture sampling system, Numera® (Urdorf, Switzerland). These samples were analyzed for optical density, glucose, and ethanol. The optical density measurements provide an estimate of biomass concentration based on a dry weight calibration. The cultures were conducted at GSK, Rixensart, Belgium, and any other detail remains confidential.

To build the model, we are considering an initial experiment that consists of a 90-hour culture, with the first 20 hours being achieved in batch mode, followed by a fed-batch culture with an exponential feed flow rate (F_{in}). This input flow is shown in Figure 2, and is referred to as EXP1. All values have been normalized for the sake of data confidentiality.

Following the methodology outlined in Figure 1, experiments EXP2, EXP3, EXP4, and EXP5 were proposed based on the progressing validation results of each iteration. These latter include the sensitivity analysis of the model with respect to each parameter for each experiment reported in section 5. The resulting F_{in} profiles are presented in Figure 2.

5. PARAMETER IDENTIFICATION

4 experimental data sets were obtained from 4 distinct fed-batch cultures identified as EXP1, EXP2, EXP3, and EXP4, to estimate the model parameters. This procedure follows the methodology outlined in (Benavides et al., 2024), which employs a least-squares criterion representing the weighted distances between the experimental data and the model predictions as follows:

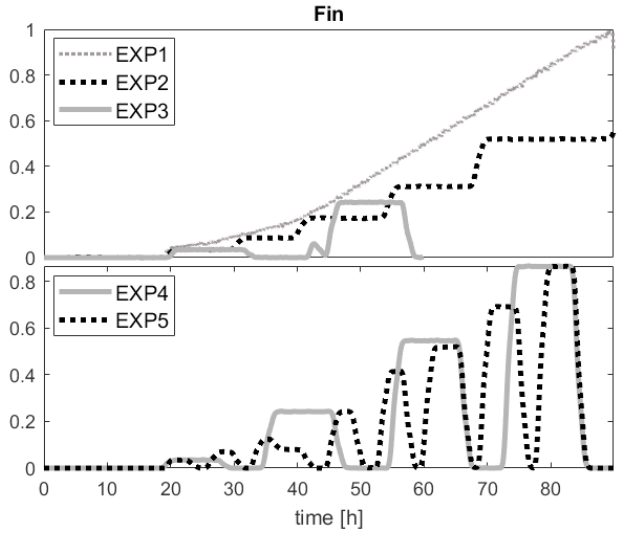

Fig. 2. Feed flow rates for the designed experiments

$$J(\theta) = \sum_{i=1}^{N} \left[(y_i(\theta) - y_{i,\text{meas}})^T \cdot W_i^{-1} \cdot (y_i(\theta) - y_{i,\text{meas}}) \right] \tag{4}$$

where $y_{i,\text{meas}}$ represents the measurement vector, y_i the model prediction vector, θ the parameter vector, N the number of measurement samples, and W_i the weighting matrix which is usually diagonal and contains the variance of the measurement errors. In most cases, this variance cannot be assessed precisely and is generally replaced by the square of each state variable maximum level taken from the data sets.

The minimization of the cost J is a nonlinear programming problem requiring a numerical solver or optimizer. Among the several available optimizers of the MATLAB libraries, we selected the "fminsearch" function, which applies the Nelder-Mead simplex algorithm. This method was chosen for its effectiveness in handling nonlinear optimization problems which are likely to present local minima.

The parameter values obtained from this optimization procedure are listed in Table 1. The coefficient of variation (CV), which measures the relative variability of the parameter estimates, is calculated using the FIM defined as follows (Walter and Pronzato, 1997):

$$FIM = \sum_{i=1}^{N} S_{\boldsymbol{\theta},i} W_i^{-1} S_{\boldsymbol{\theta},i}^T \tag{5}$$

where $S_{\boldsymbol{\theta}}$ is the vector of the model local sensitivities (i.e., the sensitivities of the state variables $\mathbf{y}$) to a specific set of parameters $\boldsymbol{\theta}$, which reads:

$$S_{\boldsymbol{\theta},i} = \frac{\partial \mathbf{y_i}}{\partial \boldsymbol{\theta}} \tag{6}$$

The variances of the parameter estimation errors are extracted from the diagonal of the inverse of the FIM, which is assumed to be an optimistic estimate (or lower bound) to the covariance matrix Cov as follows:

$$FIM^{-1} < Cov \tag{7}$$

The CVs, expressed in percentage, were computed as the relative standard deviations (square roots of the variances, normalized by the estimated value of the corresponding parameter), characterizing parameter precision, and therefore assessing model reliability.

Table 1. Parameter estimate values with their respective CVs using four experiments

Parameter Name	Parameter Value	Units	CV(%)
μ_{m1}	0,89	[gS/gX/h]	22,0
μ_{m2}	0,46	[gS/g(X+E)/h]	41,2
μ_{m3}	0,40	[gE/gX/h]	6,5
K_{IX}	45,77	[gX/L]	1,8
K_{S1}	0,01	[gS/L]	26,4
K_{S2}	0,00	[gS/L]	27,0
K_{IE}	31,79	[gE/L]	64,8
K_{IS}	0,00	[gS/L]	25,3
K_{E}	0,41	[gE/L]	12,1
k_1	0,12	[g/g]	30,2
k_2	1,05	[g/g]	42,2
k_3	1,54	[g/g]	1,3
k_4	0,22	[g/g]	33,8
K_{IE2}	5,30	[gE/L]	14,3

The corresponding relative sensitivities $\bar{S}_\theta$ are also computed as in:

$$\bar{S}_{\boldsymbol{\theta}_i} = S_{\boldsymbol{\theta}_i} \frac{\boldsymbol{\theta}_i}{\bar{\mathbf{y}}} \tag{8}$$

where $\bar{y}$ stands for the mean value.

The proposed experiments seek to highlight the effects of the kinetic factors from (2), by actuating the feed rate profile in such a way that the state variables span the desired ranges and the sensitivities reach significant values.

The proposed design of experiments is exemplified by the parameter K_{IE}, which exhibits the highest coefficient of variation (CV) value in Table 1. Figure 3 shows the normalized sensitivity of ethanol to the parameter K_{IE} for different inputs in EXP1, EXP2, EXP3, and EXP4. A high sensitivity of ethanol is located between 40 and 57 hours during EXP1 but, unfortunately, not at any other moment, which may induce a poor practical identifiability (i.e., large CV). Practically, the only way to activate this sensitivity (or growth inhibition by ethanol) would be to force high ethanol concentrations all along the culture, which could deteriorate the cell viability and prematurely end the culture. A trade-off is chosen in EXP2, EXP3, and EXP4, activating the sensitivity at different times (typically with feed rate steps, see Figure 2), enhancing K_{IE} practical identifiability, and reducing the corresponding CV.

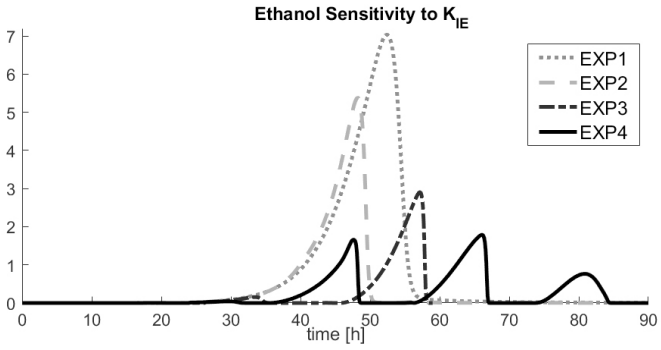

Fig. 3. Sensitivity of the ethanol to the parameter K_{IE} for EXP1, EXP2, EXP3 and EXP4

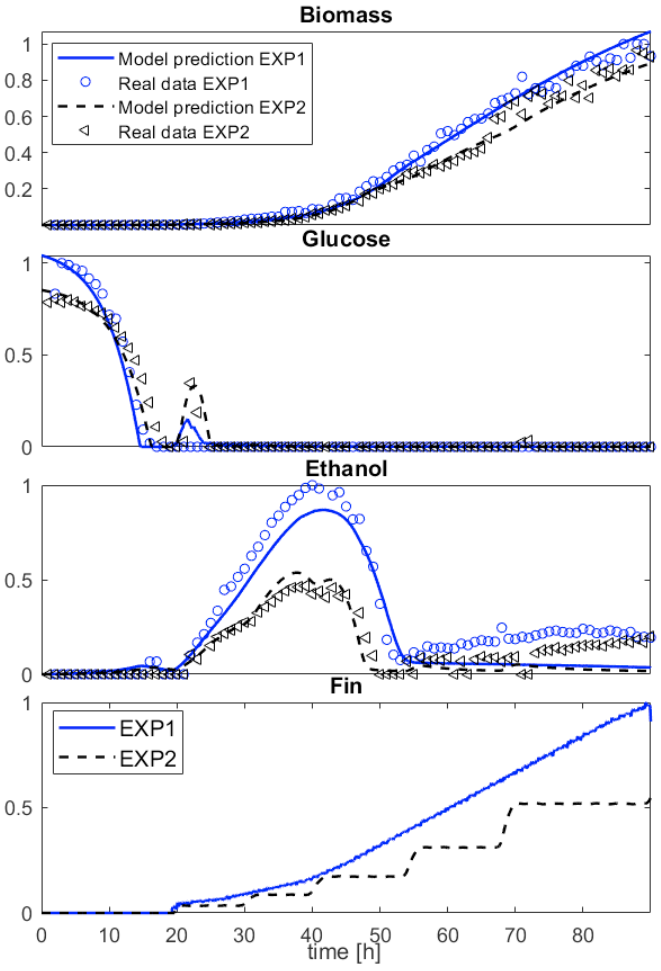

Fig. 4. Model direct validations for EXP1 and EXP2 using the four available experimental data sets

6. DIRECT AND CROSS-VALIDATION

The experimental data of biomass, glucose, and ethanol, as well as the feed rate F_{in} profiles applied in experiments EXP1 and EXP2, are visually reported in Figure 4. This figure also includes the predictions obtained with the model considering the respective state variable initial conditions as unknown parameters. Similarly, Figure 5 illustrates the corresponding information for cultures EXP3 and EXP4. These four experiments were employed for direct validation. In these experiments, biomass, glucose,

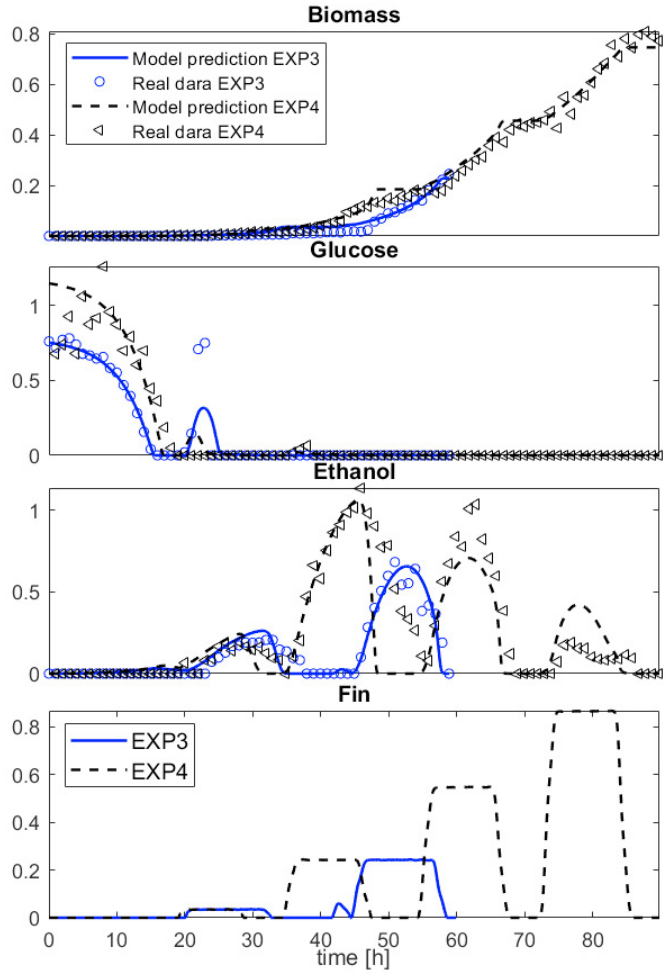

Fig. 5. Model direct validations for EXP3 and EXP4 using the four available experimental data sets

and ethanol data were measured offline every hour, while the actual applied feed rate was indirectly measured by the bioreactor weight variation every 5 minutes.

Model fitting quality is assessed using a normalized root mean square error (NRMSE) criterion defined as follows:

$$NRMSE_y = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(y_i\left(\theta^*\right) - y_{i,\text{meas}}\right)^2\right]}}{\overline{y}_{i,\text{meas}}} \tag{9}$$

where $y_{i,\text{meas}}$ is the measurement output at time i, $(y_i(\theta^*)$ is the model output predicted with the estimated parameter set θ^* at sampling time i, N is the total number of measurement samples, and $\overline{y}_{i,\text{meas}}$ is the measurement mean.

The NRMSE values for each variable can be found in Figure 6. It is worth noting that despite the diverse input profiles across the several experiments, the model fitting and the corresponding NRMSE are satisfactory. The NRMSE is also particularly low in EXP1, indicating that the model prediction reliability is maximum for an exponential feed profile.

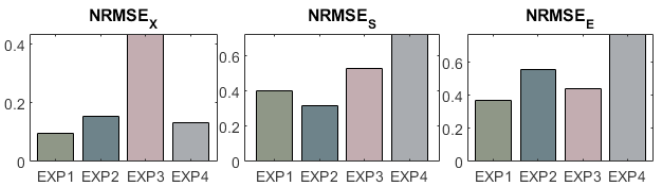

Fig. 6. Normalized RMSE of each variable and experiment.

Cross-validation has been achieved with an additional dataset denoted EXP5. The experimental data and corresponding model predictions are shown in Figure 7. The NRMSE values of biomass, glucose, and ethanol are 0.49, 1.08, and 0.84 respectively. The relatively higher NRMSE value for glucose may be attributed to discrepancies located between 20 and 30 hours of culture, where an unexpected glucose peak occurs. This phenomenon is also observed at a lower level during direct validations and could correspond to an overestimation of the cell dynamics (i.e., its capacity to consume glucose during this period) which is easier to catch at a higher biomass concentration order of magnitude. This suggests a possible and unmodeled evolving metabolism from the first hours to the late instants of the culture. From a control point of view, a robust framework should be envisaged to alleviate model uncertainties (Dewasme et al., 2024). Hybrid modeling could also be considered to adapt model parameters with the help of physicochemical or other metabolic information.

7. CONCLUSION

In this study, a practical method was used to develop a mathematical model for predicting the time evolution of biomass, glucose, and ethanol concentrations in a 90-hour fed-batch culture of the yeast *S. cerevisiae*. The method is based on a cyclic procedure that involves multiple steps. First, a model structure is proposed after analyzing data from a primary culture. Following this, parameter identification is carried out, direct and cross-validation are performed, and the CV and sensitivity curves of each parameter are thoroughly analyzed. Based on this analysis, new experiments are proposed, new data is obtained, and a new model structure is proposed, thereby repeating the entire cycle procedure. This iterative process was repeated several times, ultimately involving four experiments for direct validation and one experiment for cross-validation. The final model structure was presented along with the estimated parameter values. The proposed model demonstrates good prediction capability through both direct and cross-validation results. Further research entails using hybrid modeling to detect possible metabolic changes inducing parameter variations or monitoring the cultures under a robust control framework that takes account of possible parameter variations.

ACKNOWLEDGEMENTS

The European researchers were funded by the Service Public de Wallonie (SPW), Belgium, under BElgian WAllonia REsearch (BEWARE) fellowships and European Union (EU) framework program for research and innovation, Marie Skłodowska-Curie Actions (MSCA).

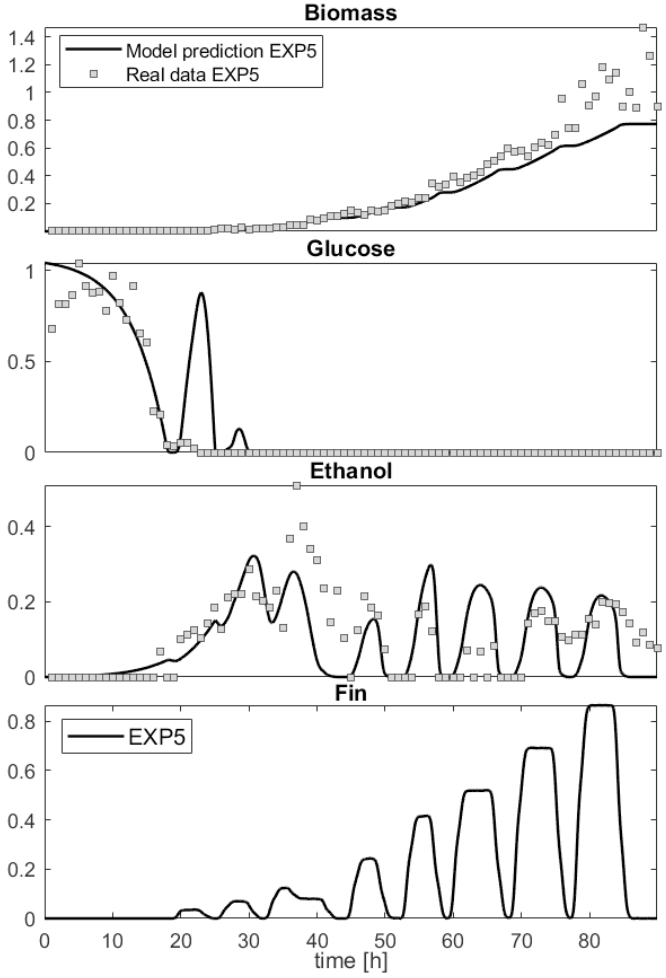

Fig. 7. Cross-validation of the proposed Model.

CONFLICT OF INTEREST

P. Gerkens and G. de Lannoy are employees of the GSK group of companies. M. Benavides is a postdoc researcher at UMONS and participates in a postgraduate fellowship program partly achieved at GSK.

REFERENCES

Abadli, M., Dewasme, L., Tebbani, S., Dumur, D., and Vande Wouwer, A. (2022). Experimental validation of a nonlinear model predictive controller regulating the acetate concentration in fed-batch escherichia coli BL21(DE3) cultures. *Advanced Control for Applications: Engineering and Industrial Systems*, 4(1), e95.

Amribt, Z., Niu, H., and Bogaerts, P. (2013). Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures. *Biochemical Engineering Journal*, 70, 196–209.

Balsa-Canto, E., Alonso, A.A., and Banga, J.R. (2010). An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Systems Biology*, 4, 1–18.

Banga, J.R. and Balsa-Canto, E. (2008). Parameter estimation and optimal experimental design. *Essays in Biochemistry*, 45, 195–210.

Benavides, M., Gerkens, P., de Lannoy, G., Dewasme, L., and Vande Wouwer, A. (2024). Modeling fed-batch cultures of yeast for the production of heterologous proteins-an industrial experimental study. In *Computer Aided Chemical Engineering*, volume 53, 109–114. Elsevier.

Bhonsale, S., Nimmegeers, P., Akkermans, S., Telen, D., Stamati, I., Logist, F., and Van Impe, J.F. (2022). Optimal experiment design for dynamic processes. In *Simulation and Optimization in Process Engineering*, 243–271. Elsevier.

Bogaerts, P. and Vande Wouwer, A. (2003). Software sensors for bioprocesses. *ISA transactions*, 42(4), 547–558.

Crabtree, H. (1929). Observations on the carbohydrate metabolism of tumors. *Biochemical Journal*, 23, 536–545.

Dewasme, L., Mäkinen, M., and Chotteau, V. (2024). Multivariable robust tube-based nonlinear model predictive control of mammalian cell cultures. *Computers & Chemical Engineering*, 183, 108592.

Dewasme, L., Srinivasan, B., Perrier, M., and Vande Wouwer, A. (2011). Extremum-seeking algorithm design for fed-batch cultures of microorganisms with overflow metabolism. *Journal of Process Control*, 21(7), 1092–1104.

Huet, A., Sbarciog, M., and Bogaerts, P. (2022). Macroscopic modeling of intracellular trehalose concentration in Saccharomyces cerevisiae fed-batch cultures. *IFAC-PapersOnLine*, 55(20), 391–396.

Jerusalimski, N. and Engamberdiev, N. (1969). *Continuous cultivation of microorganisms*, volume 517. Academic Press, New York.

Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1), 371–394.

Retamal, C., Dewasme, L., Hantson, A.L., and Vande Wouwer, A. (2018). Parameter estimation of a dynamic model of Escherichia coli fed-batch cultures. *Biochemical Engineering Journal*, 135, 22–35.

Richelle, A., Fickers, P., and Bogaerts, P. (2014). Macroscopic modelling of baker's yeast production in fed-batch cultures and its link with trehalose production. *Computers & Chemical Engineering*, 61, 220–233.

Sonnleitner, B. and Käppeli, O. (1986). Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity : Formulation and verification of a hypothesis. *Biotech. Bioeng.*, 28, 927–937.

Telen, D., Logist, F., Van Derlinden, E., Tack, I., and Van Impe, J. (2012). Optimal experiment design for dynamic bioprocesses: A multi-objective approach. *Chemical Engineering Science*, 78, 82–97.

von Stosch, M., Davy, S., Francois, K., Galvanauskas, V., Hamelink, J.M., Luebbert, A., Mayer, M., Oliveira, R., O'Kennedy, R., Rice, P., and Glassey, J. (2014). Hybrid modeling for quality by design and PAT-benefits and challenges of applications in biopharmaceutical industry. *Biotechnology Journal*, 9(6), 719–726.

Walter, E. and Pronzato, L. (1997). *Identification of parametric models: from experimental data*. Springer Verlag New-York.

Yousefi-Darani, A., Paquet-Durand, O., Hinrichs, J., and Hitzmann, B. (2021). Parameter and state estimation of backers yeast cultivation with a gas sensor array and unscented Kalman filter. *Eng Life Sci.*, 21, 170–180.