Flavio Manenti, Gintaras V. Reklaitis (Eds.), Proceedings of the 34th European Symposium on Computer Aided Process Engineering / 15th International Symposium on Process Systems Engineering (ESCAPE34/PSE24), June 2-6, 2024, Florence, Italy

Modeling Fed-batch Cultures of Yeast for the Production of Heterologous Proteins - an Industrial Experimental Study

Micaela Benavides[a,b], Pascal Gerkens[b], Gaël de Lannoy[b], Laurent Dewasme[a] and Alain Vande Wouwer[a,*]

[a]*Systems, Estimation, Control and Optimization (SECO), University of Mons, 7000 Mons, Belgium* [b]*GlaxoSmithKline, 1330 Rixensart, Belgium*
**alain.vandewouwer@umons.ac.be*

Abstract

This work reports on the development of a dynamic model of yeast cultures based on a few industrial vaccine production data sets, hypothesizing on the structure of the kinetics, and testing various parametrizations. The proposed model describes the catabolism of a genetically modified strain of the yeast *Saccharomyces cerevisiae*. The main metabolic mechanisms are translated into simple multiplicative activation and inhibition kinetic factors, avoiding the use of nonlinear switching functions, and involving only a few measurable concentrations, i.e., biomass, glucose, and ethanol. A parameter identification study is carried out, based on the minimization of a non- linear least squares criterion, and taking measurement noise into account. A sensitivity analysis is performed, where the resulting Fisher Information Matrix is used to characterize the precision of the parameter estimates. A model extension, including dissolved oxygen, is also proposed as a promising alternative.

Keywords: Mathematical Modeling, Parameter identification, Estimation, Biotechnology, *Saccharomyces cerevisiae*.

1. Introduction

Antigen production is an essential step for vaccine development, usually achieved with the help of a vector such as a genetically modified yeast strain. The latter is cultivated in scaled-up bioreactors, where the prediction of critical compound concentration profiles can be achieved by dynamic process models. Following advanced genetic studies, the yeast *S. cerevisiae* has been successfully modified for the production of recombinant vaccines (Silva et al., 2022).

During culture scale-up, the bioreactor environmental conditions and the culture medium must meet some quality attributes, such as a high yield of recombinant protein expression and, in turn, vaccine efficiency (Vieira Gomes et al., 2018). One way to meet these conditions consists in controlling the input flow rate, based on a model-based feedback loop using the available measurements, i.e. biomass, substrate, ethanol, ammonium, and dissolved oxygen concentrations. However, all these state variables should be measurable online and, therefore, require expensive equipment. Software sensors are an interesting alternative, providing unmeasurable state estimates using the available mathematical models and measurement device configuration (Bogaerts and Vande Wouwer, 2003), (Dewasme et al., 2009).

The metabolism of *S. cerevisiae* is ruled by its respiratory capacity (Sonnleitner and Käppeli, 1986). When operated at low glucose levels, the yeast culture oxidizes all the available glucose into biomass and carbon dioxide (CO2). This metabolic regime is called respirative and only a part of the respiratory capacity is used. When the respiratory capacity becomes insufficient to oxidize all the available glucose, and acts as a bottleneck, the yeast metabolism switches to the fermentation pathway, where the glucose excess

fermentates into ethanol. This regime is also known as the Crabtree effect or overflow metabolism (Deken, 1966). It should be noticed that, in presence of ethanol, a switch back to the respirative regime leads to ethanol consumption, thanks to the remaining part of the respiratory capacity (Wills, 1990). Fed-batch applications often aim to avoid this "short-term Crabtree effect" by using exponential feeding to maintain a low glucose concentration during yeast cultivation (Patel and Thibault, 2009; Dewasme et al., 2011). Several other models have been suggested to expand the unstructured model of Sonnleitner and Käppeli (1986). For instance, Richelle et al. (2014) have considered the consumption of nitrogen as an evident limiting factor of biomass growth and a responsible factor for the production of trehalose. However, most of these models consider short-term cultures, which usually do not allow triggering several metabolic switches, for instance leading to a second ethanol production phase as depicted in (Grosfils, 2007).
The main motivation of this work is to present a mechanistic model of yeast culture experiments operated in fed-batch mode including several metabolic switches. In the next section, the experimental conditions are briefly described. Section 3 develops the model based on macroscopic mass balance equations and the selection of kinetic structures combining different modulation factors. Section 4 compares the most promising models and conclusions are drawn in section 5.

2. Materials and Methods

In this study, a recombinant yeast strain of *S. cerevisiae* is cultivated in fed-batch mode. A total of three cultures have been conducted with an initial volume of 5.5 L, and a stirrer speed set to 260 rpm. The temperature was maintained at 30°C throughout all the experimental sessions, while the pH was regulated at 5, using a base solution. The bioreactor was equipped with an in-line pO2 sensor delivering dissolved oxygen in percents, a stirrer motor controlled to maintain this pO2 above 60%, and a peristaltic pump controlling the feed flow rate. The culture duration was set to 95 hours, and offline measurements of optical density, glucose, ethanol, and ammonium were taken every 2 hours. The optical density provides the concentration of biomass based on a dry weight calibration. The cultures were conducted in GSK laboratories, and any other detail remains confidential.

3. Model development

Two candidate models are proposed. The first model (Model 1), considers three reactions with variables including biomass (X), glucose (S), and ethanol (E). This model counts 13 parameters to be estimated. Model 2 simplifies the process description by using only two reactions and incorporates oxygen information.

3.1. Model 1

Model 1 focuses on the macroscopic description of both respirative and fermentative regimes, as shown in Table 1. The first reaction considers glucose oxidation, the second reaction describes the respiro-fermentative pathway while the third reaction considers the possible ethanol consumption. The first and third reactions occur in the respirative regime while the first and second ones correspond to the fermentative regime.

The first reaction rate r_1 considers Monod kinetics to describe glucose uptake, limited by the available respiratory capacity either inhibited by the biomass density (second factor) or the presence of ethanol (third factor). The second reaction rate r_2 is also ruled by a Monod factor related to glucose uptake and limited by the respiratory capacity depending on the biomass density. The third reaction rate r_3 is driven by the presence of ethanol using Monod kinetics, a respiratory capacity limitation factor comparable to reactions 1 and 2, and an inhibition factor by substrate, explaining the preferential selection, as main substrate, of glucose over ethanol.

Mass balance is applied to each macro-reaction, yielding a differential equation system eqs. (4) to (6). $D = F_{in}/V$ is the dilution rate where F_{in} is the inlet feed flow rate and V the bioreactor volume.

Table 1: Model 1

Stoichiometric equations	Reactions and differential equations	
$S \xrightarrow{r_1 X} k_1 X$	$r_1 = \mu_{m1} \cdot \frac{S}{K_{S1}+S} \cdot \frac{1}{1+\frac{X}{K_{IX}}} \cdot \frac{1}{1+\frac{E}{K_{IE}}}$	(1)
$S \xrightarrow{r_2 X} k_2 E + k_4 X$	$r_2 = \mu_{m2} \cdot \frac{S}{K_{S2}+S} \cdot \frac{1}{1+\frac{X}{K_{IX}}}$	(2)
$E \xrightarrow{r_3 X} k_3 X$	$r_3 = \mu_{m3} \cdot \frac{E}{K_E+E} \cdot \frac{1}{1+\frac{X}{K_{IX}}} \cdot \frac{1}{1+\frac{S}{K_{IS}}}$	(3)
	$\frac{dX}{dt} = (k_1 r_1 + k_3 r_3 + k_4 r_2) \cdot X - D \cdot X$	(4)
	$\frac{dS}{dt} = -(r_1 + r_2) \cdot X - D \cdot S + D \cdot S_{in}$	(5)
	$\frac{dE}{dt} = (k_2 r_2 - r_3) \cdot X - D \cdot E$	(6)
	$\frac{dV}{dt} = D \cdot V = F_{in}$	(7)

X :Biomass concentration (optical density)[g/L]; S:Substrate concentration [g/L]; E:Ethanol concentration [g/L]; S_{in}:Inlet feed substrate concentration [g/L]; F_{in}:Feed flow rate[L/h]; V :Culture volume[L], k_i:Yield coefficients [g/g]; K_{S1}, K_{S2}:Substrate half-saturation constants[g/L]; K_E:Ethanol half-saturation constant[g/L]; K_{IX}, $K_{IE}$and K_{IS}: Biomass, ethanol and substrate inhibition constants [g/L]; μ_{mi}: Maximum specific rate constants [g/gX/h].

3.2. Model 2

Model 2 merges the first two reactions of Model 1 into one reaction, assuming that a clear separation of glucose oxidation and fermentation mechanisms is difficult to observe. The second reaction of Model 2 still describes ethanol oxidation but also considers the additional dissolved oxygen variable *O*, assumed to drive ethanol consumption. Applying mass balance to all the involved macroscopic species leads to the new differential equations eqs. (10) to (14).

Table 2: Model 2

Stoichiometric equations	Reactions and differential equations	
$S \xrightarrow{r_1 X} k_1 X + k_2 E$	$r_1 = \mu_{m1} \cdot \frac{S}{K_{S1}+S} \cdot \frac{1}{1+\frac{X}{K_{IX}}} \cdot \frac{1}{1+\frac{E}{K_{IE}}}$	(8)
$k_4 E + O \xrightarrow{r_2 X} k_3 X$	$r_2 = \mu_{m2} \cdot \frac{O}{K_O+O} \cdot \frac{E}{K_E+E} \cdot \frac{1}{1+\frac{X}{K_{IX}}} \cdot \frac{1}{1+\frac{S}{K_{IS}}}$	(9)
	$\frac{dX}{dt} = (k_1 r_1 + k_3 r_2) \cdot X - D \cdot X$	(10)
	$\frac{dS}{dt} = -r_1 \cdot X - D \cdot S + D \cdot S_{in}$	(11)
	$\frac{dE}{dt} = (k_2 r_1 - k_4 r_2) \cdot X - D \cdot E$	(12)
	$\frac{dO}{dt} = -r_2 \cdot X - D \cdot O + OTR$	(13)
	$\frac{dV}{dt} = D \cdot V = F_{in}$	(14)

O : concentration of dissolved oxygen [g/L]; OTR: Oxygen transfer rate [g/Lh]; K_o: Oxygen half-saturation constant[g/L].

OTR is the oxygen transfer rate from the gas to the liquid phase (eq. 13) and it is a variable that is challenging to measure, requiring a gas analyzer. An interesting measurement approach is proposed in (Rocha, 2003) where both oxygen and nitrogen gas fractions and flow rates are measured at process input and output. An alternative way to approximate OTR is also to use the volumetric transfer coefficient of oxygen, $k_L a$, as follows (Papapostolou et al., 2019): $OTR = k_L a(O_{sat} - O)$, where the difference between the

dissolved oxygen concentration at saturation (O_{sat}) and the actual concentration (O) in the liquid phase is assumed to be proportional to the OTR in a specific bioreactor environment modeled by $k_L a$ and assumed constant. However, even approximating the constant $k_L a$ value is not trivial and the latter depends on several factors, such as the agitation speed, the viscosity of the culture medium, and the stirrer/tank geometry.

To overcome these difficulties, and since the dissolved oxygen tension pO2 measurements are available, the OTR is replaced by the derivative of pO2, assuming that this signal tracks the OTR trajectory. Consequently, Equation (13) is modified as follows:

$$\frac{dm}{dt} = -r_2 \cdot X - D \cdot m + \frac{dpO2}{dt} \tag{15}$$

where m represents a metabolic variable that is related to the concentration of dissolved oxygen. Since the concentration of pO2 is given in percentage, the value of m is also expressed in percentage. Additionally, O is replaced by m in Equation (9), and the parameter to be identified, K_O, is also expressed in percentage.

4. Identification of the proposed models

Two experimental data sets were considered for parameter identification, while a third data set was used for model cross-validation. A weighted least-squares criterion representing the weighted distances between the data and the model predictions is minimized and reads as follows:

$$J(\theta) = \sum_{i=1}^{N} \left[\left(y_i(\theta) - y_{i,meas}\right)^T \cdot W_i^{-1} \cdot \left(y_i(\theta) - y_{i,meas}\right) \right] \tag{16}$$

where $y_{i,meas}$ is the measurement vector, y_i is the model prediction vector, θ is the parameter vector, N is the number of measurement samples, and W_i^{-1} is the weighting matrix. The minimization of J is considered in the following nonlinear programming problem:

$$\hat{\theta} = \arg\min_{\theta} J(\theta) \tag{17}$$

Among the several available optimizers present in the MATLAB libraries, the Nelder-Mead function, achieving the optimization by application of the simplex algorithm, and used in the routine 'fminsearch', is selected. To assess the accuracy of the parameter estimates, the Fisher Information Matrix (FIM) (Walter and Pronzato, 1997) is built:

$$F = \sum_{i=1}^{N} \left(\frac{\partial y}{\partial \theta}(t_i) \right)^T \Sigma_i^{-1} \left(\frac{\partial y}{\partial \theta}(t_i) \right) \tag{18}$$

The FIM contains information about the measurement noise (via the inverse of the measurement error covariance matrix Σ_i^{-1}), and the output sensitivity function $\partial y/\partial \theta$ at each point i in time. The inverse of the FIM provides an optimistic estimate (i.e., a lower bound) of the parameter estimation error covariance matrix $\widehat{Cov}$. In practice, the parameter estimation error standard deviations can be inferred from the square roots of the diagonal elements of $\widehat{Cov}$, as follows:

$$\hat{\sigma}_i(\hat{\theta}_i) = \sqrt{F_{ii}^{-1}} \tag{19}$$

The coefficients of variations (CV) are therefore calculated as $CV_i = \frac{\hat{\sigma}_i}{\hat{\theta}_i}$. The parameter values along with their respective CV for the two models previously described are presented in Figure 1. Model 1 has the best curve fitting as highlighted by the root-mean-square error (RMSE) values in Figure 1, which can be explained by the higher number of parameters providing more degrees of freedom to fit the data.

Model 1			Model 2		
Parameter Name	Parameter Value	CV(%)	Parameter Name	Parameter Value	CV(%)
μ_{m1}	1,06	69,05	μ_{m1}	1,197	1,776
μ_{m2}	0,26	285,62	μ_{m2}	0,821	5,843
μ_{m3}	0,77	5,55	K_{IX}	50,479	2,049
K_{IX}	26,33	1,33	K_{S1}	0,007	22,392
K_{S1}	0,01	38,25	K_{IE}	20,051	6,956
K_{S2}	0,02	9,63	K_{IS}	0,001	19,998
K_{IE}	27,11	57,21	K_E	0,210	7,900
K_{IS}	0,00	13,31	k_1	0,343	2,057
K_E	0,26	10,85	k_2	0,461	0,797
k_1	0,05	58,49	k_3	0,435	3,073
k_2	2,09	287,04	k_4	0,608	2,307
k_3	1,73	3,54	K_O	1581,173	6,510
k_4	0,62	276,11			

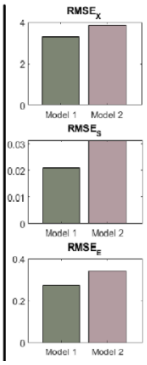

Figure 1: Parameter estimate values with their respective CVs and RMSEs of the biomass (X), glucose (S), and ethanol (E) predictions of the 2 models.

However, despite higher RMSEs, it can also be observed that the maximum CV values of Model 2 are lower than those of Model 1 which presents high CVs for all parameters involved in the second reaction. This suggests that the simplification done in Model 1, where reactions 1 and 2 are merged, is consistent with the inclusion of dissolved oxygen. The curve fitting of both models is shown in Figures 2 and 3 in direct validation using one of the two available data sets. For the sake of confidentiality, all values are normalized. It is worth mentioning that the quality of the model fitting on the other data set, used in direct validation, is similar. Based on these results, both models seem to properly reflect the process behavior. The results indicate that despite a higher RMSE value, Model 2 presents the advantage of better reproducing the data trends, more specifically when ethanol enters its second production phase at the end of the culture. Additionally, the cross-validation of Model 1 and Model 2 using the third experimental data set is presented in Figure 4, suggesting the good prediction capacity of both models. The same conclusions from the direct validations apply to the cross-validation since it can be observed that Model 2 provides a better prediction of the end of the experiment while Model 1 still presents a slightly better precision (the corresponding RMSE value is ±6% less). It is worth noting that the ethanol measurement device has a limited sensitivity assimilating too small non-zero values to a small constant. All initial conditions are however identified during the cross-validation, allowing the model to start below the device sensitivity threshold.

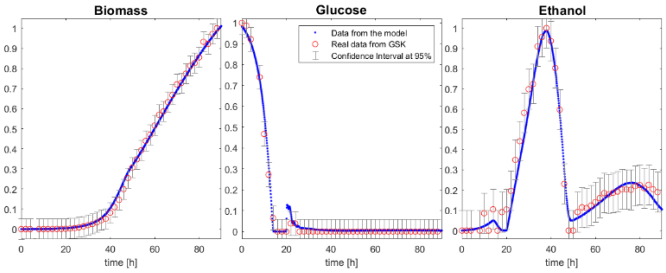

Figure 2: Estimation of biomass, glucose and ethanol using the proposed Model 1.

5. Conclusions

In this study, a mathematical model has been proposed, that predicts the time evolution of biomass, glucose, and ethanol concentrations in cultures of the yeast *S. cerevisiae*. The model takes into account the reactions of glucose oxidation, glucose fermentation, and ethanol oxidation. Additionally, the model is based on the structuration of the kinetic laws, which are based on the product of activation and inhibition factors. A second model

has been developed which simplifies the oxidation and fermentation of the glucose into one reaction, thereby reducing the number of parameters involved. In this model, the inclusion of the dynamics of the dissolved oxygen has shown to significantly improve the model formulation and its prediction. The proposed models are confirmed to have good prediction capability through both direct and cross-validation results.

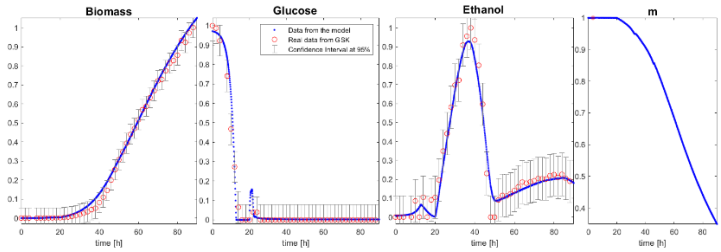

Figure 3: Estimation of biomass, glucose, ethanol and O^* using the proposed Model 2.

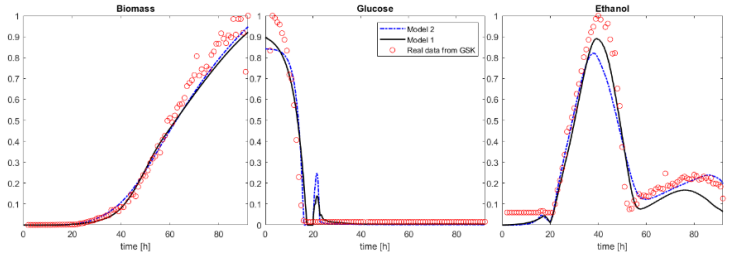

Figure 4: Cross-validation of the proposed Model 1 and 2, with normalized RMSE values of 0.8725 and 0.9304, respectively.

References

P. Bogaerts, A. Vande Wouwer, 2003. Software sensors for bioprocesses. ISA Transactions 42.

R. H. D. Deken, 1966. The crabtree effect: A regulatory system in yeast. J . gen. Microbiol. 44.

L. Dewasme, P. Bogaerts, A. Vande Wouwer, 2009. Monitoring of bioprocesses: mechanistic and data-driven approaches. Studies in Computational Intelligence, (Computational Intelligent Techniques for Bioprocess Modelling, Supervision and Control, Maria do Carmo Nicoletti, Lakhmi C. Jain, eds.). Springer Verlag, pp. 57–97.

L. Dewasme, B. Srinivasan, M. Perrier, A. V. Wouwer, 2011. Extremum-seeking algorithm design for fed-batch cultures of microorganisms with overflow metabolism. Journal of Process Control.

A. Grosfils, 2007. First principles and black box modelling of biological systems. Ph.D. thesis, Université libre de Bruxelles.

A. Papapostolou, E. Karasavvas, C. Chatzidoukas, 2019. Oxygen mass transfer limitations set the performance boundaries of microbial pha production processes–a model-based problem investigation supporting scale-up studies. Biochemical engineering journal 148, 224–238.

N. Patel, J. Thibault, 2009. Data reconciliation using neural networks for the determination of kLa. Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control.

A. Richelle, P. Fickers, P. Bogaerts, 2014. Macroscopic modelling of baker's yeast production in fed-batch cultures and its link with trehalose production. Computers & chemical engineering 6.

I. Rocha, 2003. Model-based strategies for computer-aided operation of recombinant e. coli fermentation. Ph.D. thesis, University of Minho, Portugal.

A. J. D. Silva, C. K. d. S. Rocha, A. C. de Freitas, 2022. Standardization and key aspects of the development of whole yeast cell vaccines. Pharmaceutics 14 (12), 2792.

B. Sonnleitner, O. Käppeli, 1986. Growth of saccharomyces cerevisiae is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. Biotechnology and bioengineering 28 (6), 927–937.

A. M. Vieira Gomes, T. Souza Carmo, L. Silva Carvalho, F. Mendonça Bahia, N. S. Parachin, 2018. Comparison of yeasts as hosts for recombinant protein production. Microorganisms 6 (2), 38.

E. Walter, L. Pronzato, 1997. Identification of parametric models. Springer Verlag New-York.

C. Wills, 1990. Regulation of sugar and ethanol metabolism in saccharomyces cerevisiae. Critical reviews in biochemistry and molecular biology 25 (4), 245–280.