Promega

Would you like to accelerate your time to results?

Find a manufacturing partner who can meet your current and future needs.

Let's TALK CUSTOM

40+ years CUSTOM EXPERIENCE

5 GLOBAL Mfg Sites

PRIMARY MANUFACTURER

Custom manufacturing often appears daunting, with so many options, standards and possibilities. There's no need to figure it out alone. We will help you find the right solution for your specifications and your timetable.

Learn more with our free webinar:
promega.com/CustomWebinar

Received: 15 January 2020 | Revised: 5 May 2020 | Accepted: 5 June 2020

DOI: 10.1002/bit.27454

ARTICLE

Biotechnology and Bioengineering WILEY

Biomass soft sensor for a *Pichia pastoris* fed-batch process based on phase detection and hybrid modeling

Vincent Brunner | **Manuel Siegl** | **Dominik Geier** | **Thomas Becker**

Chair of Brewing and Beverage Technology, Technical University of Munich, Freising, Germany

Correspondence
Dominik Geier, Chair of Brewing and Beverage Technology, Technical University of Munich, Weihenstephaner Steig 20, 85354
Freising, Germany.
Email: dominik.geier@tum.de

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: 031B0475E

Abstract

A common control strategy for the production of recombinant proteins in *Pichia pastoris* using the alcohol oxidase 1 (AOX1) promotor is to separate the bioprocess into two main phases: biomass generation on glycerol and protein production via methanol induction. This study reports the establishment of a soft sensor for the prediction of biomass concentration that adapts automatically to these distinct phases. A hybrid approach combining mechanistic (carbon balance) and data-driven modeling (multiple linear regression) is used for this purpose. The model parameters are dynamically adapted according to the current process phase using a multilevel phase detection algorithm. This algorithm is based on the online data of CO_2 in the off-gas (absolute value and first derivative) and cumulative base feed. The evaluation of the model resulted in a mean relative prediction error of 5.52% and R^2 of .96 for the entire process. The resulting model was implemented as a soft sensor for the online monitoring of the *P. pastoris* bioprocess. The soft sensor can be used for quality control and as input to process control systems, for example, for methanol control.

KEYWORDS

biomass, hybrid model, phase detection, *Pichia pastoris*, soft sensor

1 | INTRODUCTION

The methylotrophic yeast *Pichia pastoris* (now reclassified as *Komagataella phaffii*) is frequently used as a host for expressing heterologous proteins for both basic research and industrial production (Cereghino, Cereghino, Ilgen, & Cregg, 2002). When the methanol-inducible alcohol oxidase 1 (AOX1) promotor is used for controlling protein expression, the process is typically separated into two main phases with different objectives. In the first phase, the carbon source—typically glycerol—is converted to biomass. It aims to produce large amounts of biomass before methanol induction. This phase is often referred to as the glycerol or biomass phase and can optionally be extended by a glycerol feed to accumulate more biomass before methanol induction (Gao et al., 2012; Jahic, Veide, Charoenrat, Teeri, & Enfors, 2006). The second phase, also referred to as the induction or methanol phase, starts when methanol is added to the medium to induce protein expression via the genetically modified AOX1 promotor. This phase aims to reproducibly generate the highest product titers.

Besides product titer, biomass concentration can be seen as one of the most critical quality attributes in upstream bioprocessing due to its effect on all other quality attributes, which holds true for *P. pastoris* bioprocesses in both the glycerol and the methanol phases (J. Harms, Wang, Kim, Yang, & Rathore, 2008). Several techniques such as turbidimetry, infrared or fluorescence spectroscopy, and flow cytometry are available for monitoring biomass, as reviewed by P. Harms, Kostov, and Rao (2002), Luttmann et al. (2012), and Schügerl (2001). However, the use of online measurement systems for monitoring biomass in a technical context is still often problematic. The reasons for this include lack of reliability, the considerable dependence on the process and product matrix (isolated solutions), and high standards of operation and maintenance (Kano & Fujiwara, 2012). For these reasons, biomass is in many cases not measured online at all.

Because the direct measurement of biomass is often not feasible, soft sensors can be used for predicting it. Soft sensors consist of computational models or algorithms that allow the prediction of target values, such as biomass concentration, via continuously measured secondary variables, such as exhaust gas concentrations, dissolved oxygen (DO), and flow rates (Luttmann et al., 2012).

Various modeling techniques have been proposed for developing soft sensors, the majority of which are based on mechanistic or data-driven approaches. An overview of soft sensors and the selection of appropriate modeling techniques for online bioreactor state estimation has been presented elsewhere (Zhang, 2009). Mechanistic modeling approaches include, for example, differential balancing systems, which describe the material and energy conversions at the cellular level, as well as mass and energy balances (Jenzsch, Gnoth, Kleinschmidt, Simutis, & Lübbert, 2007). Data-driven approaches include, among others, artificial neural networks (ANN; Gonzaga, Meleiro, Kiang, & Maciel Filho, 2009) and methods from the field of multivariate statistical process control (Kadlec, Gabrys, & Strandt, 2009), such as principal component regression and partial least squares regression. In hybrid modeling, mechanistic and data-driven modeling approaches are combined, as reviewed by Kalos, Kordon, Smits, and Werkmeister (2003) and Solle et al. (2017).

The main challenges in the development of soft sensors are as follows: control of model complexity (overfitting vs. underfitting) (Kordon, Smits, Kalos, & Jordaan, 2003); limited amount of data sets or data points (Fortuna, Graziani, & Xibilia, 2009); outliers resulting from, for example, sensor faults (Zhang, 2009); adaption mechanisms for model maintenance (Bakirov, Gabrys, & Fay, 2017); input variable selection; reliability of soft sensors; and changes in process characteristics and operating conditions (Kano & Fujiwara, 2012). In addition, a specific challenge arises in soft sensor development for *P. pastoris* bioprocesses given its distinct process phases, as described previously: The underlying principles of prediction models for biomass are related to the inherent biological relationships between online measured variables and biomass (Chen, Nguang, Li, & Chen, 2004); thus, the soft sensor needs to be adaptive to the current process phase to give accurate prediction results throughout the entire process.

In this study, an adaptive soft sensor for biomass concentration was developed. The novelty of this study is that the soft sensor changes its model coefficients regarding the current process phase (batch, transition, or fed-batch phase) of the *P. pastoris* bioprocess. The soft sensor's underlying prediction model is based on a hybrid of mechanistic and data-driven approaches. The mechanistic part comprises mass balancing of carbon using methanol and CO_2 fluxes. The outcome of this mechanistic model—the generation rate of total organic carbon inside the bioreactor—is fed into a data-driven model that in turn leads to an online prediction of biomass concentration. The adaptability of the soft sensor to the distinct process phases is guaranteed by automatic and reliable detection of glycerol depletion based on online process variables, namely, CO_2 in the off-gas (absolute value and first derivative) and cumulative base feed. The soft sensor's model coefficients switch automatically depending on the current process phase and thus give accurate biomass predictions throughout the entire process. Finally, the soft sensor was implemented in a real-time capable system to enable online biomass monitoring.

2 | MATERIALS AND METHODS

2.1 | Strain and preculture conditions

The inoculum of a recombinant *P. pastoris* strain based on type strain DSMZ 70382 was prepared in three 150 ml shake flasks containing 50 ml of the mineral medium FM22 with glycerol as the carbon source: $(NH_4)_2SO_4$, $5\,g \cdot L^{-1}$; $CaSO_4 \cdot 2H_2O$, $1\,g \cdot L^{-1}$; K_2SO_4, $14.3\,g \cdot L^{-1}$; KH_2PO_4, $42.9\,g \cdot L^{-1}$; $MgSO_4 \cdot 7H_2O$, $11.7\,g \cdot L^{-1}$; glycerol, $40\,g \cdot L^{-1}$ (Stratton, Chiruvolu, & Meagher, 1998); and trace element stock solution (PTM4), $2.0\,ml \cdot L^{-1}$ of the culture medium. The PTM4 stock solution contained $CuSO_4 \cdot 5H_2O$, $2\,g \cdot L^{-1}$; KI, $0.08\,g \cdot L^{-1}$; $MnSO_4 \cdot H_2O$, $3\,g \cdot L^{-1}$; $Na_2MoO_4 \cdot 2H_2O$, $0.2\,g \cdot L^{-1}$; H_3BO_3, $0.02\,g \cdot L^{-1}$; $CaSO_4 \cdot 2H_2O$, $0.5\,g \cdot L^{-1}$; $CoCl_2$, $0.5\,g \cdot L^{-1}$; $ZnCl_2$, $7\,g \cdot L^{-1}$; $FeSO_4 \cdot H_2O$, $22\,g \cdot L^{-1}$; biotin, $0.2\,g \cdot L^{-1}$; and conc. H_2SO_4, 1 ml. Cells were grown for 70 hr at 30°C on a shaker at $150\,min^{-1}$.

2.2 | Fed-batch cultivation in bioreactor

The shake flask culture was used to inoculate the main culture in the bioreactor Biostat® Cplus (Sartorius AG, Goettingen, Germany) with working and total volumes of 15 and 42 L, respectively. The main culture medium was FM22. Pressure, pH, temperature, and dissolved oxygen were controlled to 500 mbar, 5, 30°C, and 40%, respectively. NH_4OH was used as nitrogen source and to set and maintain a pH of 5. A dissolved oxygen minimum of 40% was controlled by a cascade control using variable stirrer speed ($300–600\,min^{-1}$) and air flow rate ($20–40\,L \cdot min^{-1}$).

The end of the batch phase, that is, the depletion of glycerol, was indicated online by a characteristic peak in the off-gas CO_2 concentration. The complete depletion of glycerol was verified offline via HPLC analysis (data not shown). After a short transition phase, which prevents the potential repression of the AOX1 promotor by glycerol residues from the preceding batch phase, the culture was induced with methanol. The methanol feed was supplemented with $12\,ml \cdot L^{-1}$ PTM4 stock solution. Methanol concentration was controlled via a fuzzy logic controller to $4.5\,g \cdot L^{-1}$. This controller uses methanol concentration as the main input and the feed rate of methanol as output. The general concept of fuzzy logic controllers is described, for example, in Birle, Hussein, and Becker (2013).

The off-gas CO_2 concentration was measured with a BlueInOne Cell sensor (BlueSens gas sensors GmbH, Herten, Germany). Methanol concentration was measured with an inline Alcosens sensor (Heinrich Frings GmbH & Co. KG, Rheinbach, Germany).

2.3 | Determination of dry cell weight

Dry cell weight was determined in triplicate by centrifugation of 2 ml cell suspension in previously weighed centrifuge tubes, followed by discarding the supernatant and drying the cell pellet to a constant weight at 80°C. Samples for the determination of dry cell weight were taken using the BaychroMAT® autosampler (Bayer AG, Leverkusen, Germany) with a minimum sampling interval of 2 hr.

2.4 | Data management

The digital control unit (DCU) of the Biostat® bioreactor (Sartorius AG) was used for primary process control (pressure, pH, temperature, and dissolved oxygen) and signal recording. SIMATIC SIPAT (Siemens AG, Munich, Germany) was used for data management and to store the process (online) and laboratory (offline) data in a central database with a recording interval of 30 s. Offline data preprocessing and modeling were performed in MATLAB R2019b (The MathWorks, Inc., Natick, MA); signal processing, real-time prediction of the target quantity, biomass concentration, by means of the developed soft sensor as well as model-based control via a fuzzy logic controller were performed in SIMULINK R2019b (The MathWorks, Inc.). An interface capable of real-time communication between the DCU, the data management system (SIMATIC SIPAT), and the online modeling software (SIMULINK) was realized via a Sartorius OPC DA server (Sartorius AG).

3 | RESULTS AND DISCUSSION

This study aims to develop a soft sensor for the prediction of biomass concentration that provides accurate online predictions for a multiphase process (batch, transition, and fed-batch phase) with two different carbon sources (glycerol and methanol). The general concept of the hybrid-model-based soft sensor presented here consists of two main levels: The first level comprises a phase detection algorithm to differentiate online among batch, transition, and fed-batch phase; the second level consists of a hybrid-model-based prediction equation that automatically adjusts the model parameters based on the current process phase (batch, transition, or fed-batch phase). For the development of the first and second levels, nine and six data sets, respectively, were used. Only the latter six data sets had a fed-batch phase with control of methanol concentration and therefore can be compared with each other.

The hybrid model uses a carbon balance as the mechanistic part. The result of the carbon balance is fed into a data-driven part to provide accurate prediction of the biomass concentration. The information-bearing model inputs that were used in this study to predict biomass concentration are cumulative methanol and base feed as well as concentrations of off-gas CO_2 and methanol.

Figure 1 shows the time course of the relevant model inputs of the soft sensor for an exemplary process run. This process run is used as an illustrative example throughout the following sections. In this case, the batch phase ends at 39.6 hr, followed by a transition phase that lasts for 6.9 hr, and a fed-batch phase that starts at 46.5 hr. In the batch phase, glycerol is metabolized and biomass is generated. The presence of the transition phase prevents the potential repression of the AOX1 promotor by glycerol residues from the preceding batch phase. In the transition phase, no significant increase (due to the absence of carbon sources) or decrease of biomass concentration was observable. In the fed-batch phase, methanol is fed into the bioreactor via a pump for the first time. Subsequently, methanol concentration is controlled to a setpoint of 4.5 g · L^{-1} via a fuzzy logic controller. This process run shows control errors such as high initial overshoot and an increasing deviation of the measured methanol concentration to the setpoint in the subsequent time course. Base (NH_4OH, 5 M) is fed into the bioreactor via a pump and is used to maintain pH at 5.0. The cumulative base feed represents the degree of metabolic activity, that is, substrate depletion. This variable shows

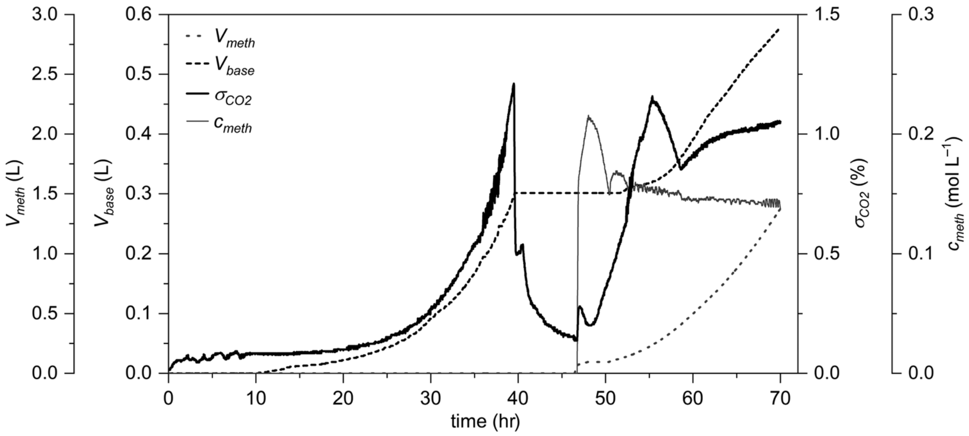

FIGURE 1 Time course of the relevant model inputs of the soft sensor for an exemplary process run, namely, cumulative feed volume of methanol, V_{meth}, and base, V_{base}, as well as concentrations of CO_2 in the off-gas, σ_{CO2}, and methanol, c_{meth}. For this exemplary process, the batch phase ends at 39.6 hr and the fed-batch phase starts at 46.5 hr

high collinearity to the biomass concentration (see later in Figure 6). In the batch phase, the off-gas CO_2 signal almost continuously increases until the end of this phase. Here, the signal drops abruptly and, except for minor fluctuations, begins to rise again only upon methanol induction. After methanol induction, the cells need to adapt to the metabolization of methanol.

3.1 | Multilevel process phase detection

3.1.1 | General concept of process phase detection

This algorithm step aims to differentiate among the three distinct process phases, which are listed in Table 1 together with its process data characteristics regarding process phase detection. The detection of the end of the batch phase is primarily based on the off-gas CO_2 signal. The metabolization of glycerol together with an increasing cell concentration leads to an almost continuous increase in CO_2 emission during the batch phase. When glycerol is depleted, the off-gas CO_2 signal drops abruptly, as shown in Figure 1 (here at 39.6 hr). The relationship between the CO_2 drop and substrate consumption is shown and discussed in detail in Munch et al. (2020). This abrupt drop is the main sign of the end of the batch phase and is hereinafter referred to as trigger 3. To increase robustness of the phase detection algorithm, two additional trigger conditions upstream of trigger 3 were implemented, namely the exceeding of absolute values for cumulative base feed (trigger 1) and off-gas CO_2 concentration (trigger 2).

The output of the algorithm for process phase detection is a binary value indicating whether the end of the batch phase has been reached (1 = true) or not (0 = false) together with the corresponding timestamp. Variable inputs to the algorithm consist of the signals for cumulative base feed (V_{base}) for trigger 1, the absolute off-gas CO_2 concentration (σ_{CO2}) for trigger 2, and the timewise derivative of the off-gas CO_2 concentration ($d\sigma_{CO2}/dt$) for trigger 3. Only when triggers 1 and 2 are initiated, that is, they are "true", trigger 3 is active and can be initiated. The end of the batch phase is indicated when all three triggers are "true."

The process variable V_{base} represents the cumulative metabolic activity regarding the consumption of the carbon source. Because the batch process starts with a glycerol concentration of $40\,g \cdot L^{-1}$, the total volume of base fed into the bioreactor at the end of the batch phase is restricted to the stoichiometry of glycerol metabolization. Trigger 1 is therefore initiated when a defined threshold for V_{base} is exceeded. In the transition phase, the variable V_{base} remains constant because cells do not grow. Similar to V_{base}, the process variable σ_{CO2} is strongly related to biomass growth and substrate consumption. During exponential growth, σ_{CO2} increases almost continuously until the end of the batch phase. Trigger 2 is therefore initiated when a defined threshold for σ_{CO2} is exceeded. This trigger is implemented to guarantee that natural fluctuations in σ_{CO2}, which can statistically occur in biological systems (see Figure 1), and sensor faults impede the functionality of the process phase detection as little as possible. Trigger 2 thus slightly increases robustness of the phase detection algorithm. Figure 2 shows the functioning of trigger 3 in terms of the time course of $d\sigma_{CO2}/dt$ for an exemplary process run. The value of $d\sigma_{CO2}/dt$ falls below the threshold uniquely at the end of the batch phase (here at 39.6 hr). A median filtering step was implemented before and after the derivation step to decrease noise of the variables σ_{CO2} and $d\sigma_{CO2}/dt$, respectively.

3.1.2 | Threshold definition

The thresholds for triggers 1, 2, and 3 were calculated as shown in (1), where $threshold_i$ is the threshold for the trigger variable used for process phase detection with $i = \{V_{base}, \sigma_{CO2}, d\sigma_{CO2}/dt\}$; $mean_i$ and std_i is the arithmetic mean and standard deviation, respectively, of the variable i at the end of the batch phase. The end of the batch phase was for this purpose defined as the time at the minimum of $d\sigma_{CO2}/dt$. SF is a constant safety factor of 3 that is implemented to avoid false positive detections of the end of the batch phase of the phase detection and thus to increase the robustness of the multilevel detection algorithm.

$$threshold_i = mean_i - std_i\ SF. \tag{1}$$

For illustration and comparison of the three triggers, Figure 3 shows the results (mean ± standard deviation) for the trigger variables normalized to the corresponding threshold. The resulting threshold values together with the mean and standard deviation are summarized in Table 2. These threshold values were implemented in SIMULINK to automatically detect the end of the batch phase and therefore to select the right model coefficients for the biomass soft sensor shown in the following.

3.2 | Mass balance for carbon

The underlying principle of the mechanistic modeling part is mass balancing of carbon. The boundary for the balancing is the bioreactor system: Carbon is fed into the bioreactor in the form of methanol (fed-batch phase) and leaves the boundary in the form of CO_2.

TABLE 1 Main characteristics of the three distinct process phases (batch, transition, and fed-batch phase) regarding process phase detection

Process phase	Main process objective	Carbon source	Main process data characteristic
Batch phase	Biomass generation	Glycerol	Abrupt drop in off-gas CO_2 signal at the end of batch phase
Transition phase	Derepression of the AOX1 promotor	None	No base feed due to absent cell growth
Fed-batch phase	Product formation	Methanol	Starting with methanol feed

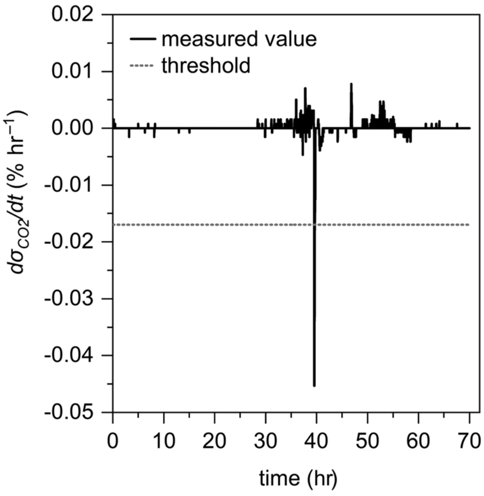

FIGURE 2 Timewise derivative of the off-gas CO_2 sensor reading, $d\sigma_{CO2}/dt$, for an exemplary process run. A median filter (window size = ten sensor readings) is implemented before and after the derivation step to handle noisy sensor readings. The characteristic negative peak (here at 39.6 hr) is the main indicator for the depletion of the batch phase substrate (glycerol) and thus the end of the batch phase. This landmark is used to initiate the start of the transition and fed-batch phase, respectively

TABLE 2 Threshold, mean, and standard deviation for the three trigger variables V_{base}, σ_{CO2}, and $d\sigma_{CO2}/dt$ at the end of the batch phase (n = 9)

Trigger number	Variable	Mean	Standard deviation	Threshold
1	V_{base}	540 ml	117 ml	188 ml
2	σ_{CO2}	1.284%	0.177%	0.755%
3	$d\sigma_{CO2}/dt$	$-0.038\% \cdot hr^{-1}$	$0.007\% \cdot hr^{-1}$	$-0.017\% \cdot hr^{-1}$

The remaining carbon is in the form of either glycerol or methanol or is bound in cells as well as extracellular organic acids and proteins. The following sections show how the timewise rates of off-gas CO_2 and methanol are calculated. These rates are then balanced to enable calculation of the formation rate of total organic carbon (TOC) that remains bound in cells as well as extracellular organic acids and proteins. To determine the cumulative amount of TOC online, this rate needs to be multiplied by the total liquid volume and numerically integrated. This cumulative amount of TOC is used in the subsequent data-driven modeling part to predict biomass concentration (Figure 4).

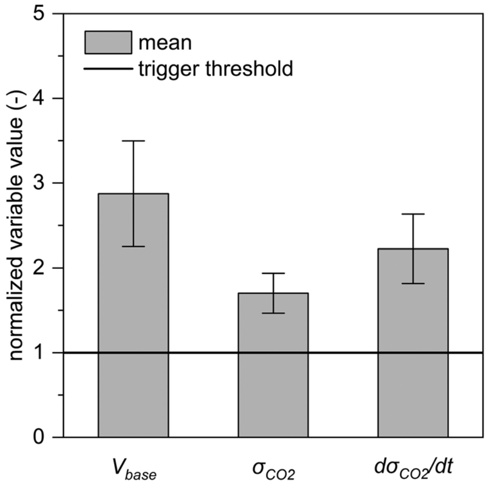

FIGURE 3 Triggers for the multilevel detection of the end of the batch phase (i.e., depletion of glycerol). Only when defined values for base, V_{base}, and absolute off-gas CO_2 concentration, σ_{CO2}, are reached for the first time, the last trigger—the timewise derivative of the off-gas CO_2 concentration, $d\sigma_{CO2}/dt$—is active. The three thresholds are defined based on the calculation of the mean and standard deviation for each of the three variables at the end of the batch phase as well as a safety factor. The diagram shows normalized absolute variable values; error bars correspond to the normalized standard deviation (n = 9)

3.2.1 | Calculation of liquid volume

To calculate the total liquid volume, all feeds and removals (sampling) need to be considered. The total reactor volume V_{total} is calculated as in (2), where V_{start} is the start volume after inoculation; V_{base}, V_{meth}, and V_{afoam} are the cumulative volumes of base, methanol, and antifoam, respectively, fed into the bioreactor; $V_{samples}$ is the cumulative volume of samples automatically taken via the BaychroMAT® autosampler:

$$V_{total} = V_{start} + V_{base} + V_{meth} + V_{afoam} - V_{samples}. \tag{2}$$

3.2.2 | Calculation of carbon dioxide emission rate

The calculation of the carbon dioxide emission rate r_{CO2} in (3) is adapted from Takors (2013), where Q_{air} is the air flow rate, p is the pressure, R is the universal gas constant (8.314×10^{-2} $L \cdot bar \cdot mol^{-1} \cdot K^{-1}$), T is the temperature, σ_{CO2} and σ_{O2} are the concentrations of carbon dioxide and oxygen, respectively, and the indices α and ω represent the gas inlet and outlet of the bioreactor, respectively:

$$r_{CO2} = \frac{Q_{air}\ p}{V_{total}\ RT}\left(\frac{1 - \sigma_{O2_\alpha} - \sigma_{CO2_\alpha}}{1 - \sigma_{O2_\omega} - \sigma_{CO2_\omega}}\sigma_{CO2_\omega} - \sigma_{CO2_\alpha}\right). \tag{3}$$

3.2.3 | Calculation of methanol reaction rate

As described above, errors in the methanol control, such as an initial overshoot or a deviation of the measured methanol concentration to the setpoint (Figure 1), can occur. The carbon balance is designed to compensate for disturbances of methanol control by incorporating the methanol accumulation rate $r_{meth,acc}$. Changes in $r_{meth,acc}$ result from the uptake of methanol by cells and methanol feeding (especially at the feed start when the methanol setpoint is reached for the first time). $r_{meth,acc}$ is determined by the timewise derivative of the

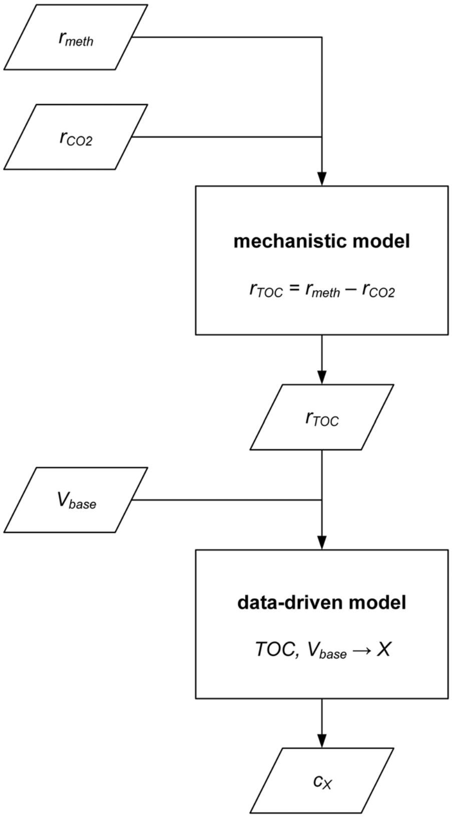

FIGURE 4 Simplified representation of the hybrid-model-based soft sensor for biomass concentration c_X. The methanol reaction rate r_{meth} and the carbon dioxide emission rate r_{CO2} are fed to the mechanistic model; carbon balancing is here used to calculate the formation rate of total organic carbon, r_{TOC}. The subsequent data-driven model uses the numerical integration of r_{TOC}, namely, TOC, together with the cumulative base feed, V_{base}, as inputs to calculate the amount of biomass X. Finally, X is divided by the total liquid volume inside the bioreactor, V_{total}, to calculate the biomass concentration c_X. Both the data-driven and the mechanistic parts can be carried out online

methanol concentration c_{meth} that is measured in the bioreactor (in-line), as follows:

$$r_{meth,acc} = \frac{dc_{meth}}{dt}. \tag{4}$$

The methanol reaction rate r_{meth} describes the net rate at which methanol accumulates in or is withdrawn from the liquid phase of the bioreactor and is calculated as follows, where $r_{meth,in}$ is the feed rate of methanol into the bioreactor related to the total liquid volume V_{total}:

$$r_{meth} = r_{meth,in} - r_{meth,acc}. \tag{5}$$

3.2.4 | Calculation of formation rate of total organic carbon

TOC refers to all carbon inside the bioreactor system that is bound in the substrate (glycerol or methanol) and cells as well as extracellular organic acids and proteins. The formation rate of TOC, r_{TOC}, is not directly measured by reference analysis but calculated as follows by balancing the methanol reaction rate r_{meth} and the carbon dioxide emission rate r_{CO2}:

$$r_{TOC} = r_{meth} - r_{CO2}. \tag{6}$$

In the batch phase $r_{meth} = 0$ and no glycerol is fed into the bioreactor; therefore, the carbon balance in this phase is $r_{TOC} = -r_{CO2}$.

3.3 | Development of hybrid-model-based soft sensor

3.3.1 | Combination of mechanistic and data-driven parts in a hybrid model

Figure 4 shows the soft sensor algorithm and how process variables are passed through the mechanistic and data-driven modeling parts to finally result in the online prediction of biomass concentration c_X. The output of the mechanistic part (mass balance for carbon), r_{TOC}, is together with V_{base} fed into the data-driven part. The data-driven part comprises a numerical integration step for r_{TOC} to obtain the cumulative amount of total organic carbon, TOC, and a multiple linear regression (MLR) step. MLR was chosen as regression method because the prediction model uses only the two inputs TOC and V_{base}.

Using TOC only for biomass prediction leads to acceptable prediction results (data not shown). However, the concentrations of dissolved carbon dioxide (H_2CO_3) as well as extracellular proteins (c_P) and organic acids, which can in most cases not be measured online, distort the biomass prediction. The prediction model for biomass is therefore complemented by adding information about acids in the medium. The process variable with most information about acids in the medium is the cumulative base feed, V_{base}. Because $c_P \ll c_X$, the extracellular protein concentration is neglected for biomass prediction.

TOC is calculated as follows by multiplication with V_{total} and numeric integration from the beginning of the process run (t_0) to the current time (t):

$$TOC = \int_{t_0}^{t} r_{TOC}\, V_{total}\, dt. \tag{7}$$

When in sum (up to t) more carbon passed the bioreactor boundary to the outside than to the inside, TOC has a negative value. The time course of TOC is together with r_{meth} and r_{CO2} illustrated in Figure 5 for an exemplary process run. In the batch phase (Figure 5a), the only carbon passing through the bioreactor boundary is CO_2. Therefore, TOC has a negative gradient. In the fed-batch phase (Figure 5b), methanol is fed to the bioreactor, resulting in a net positive gradient for TOC.

The soft sensor uses three distinct sets of model coefficients for each the batch, transition, and fed-batch phase. For model calibration

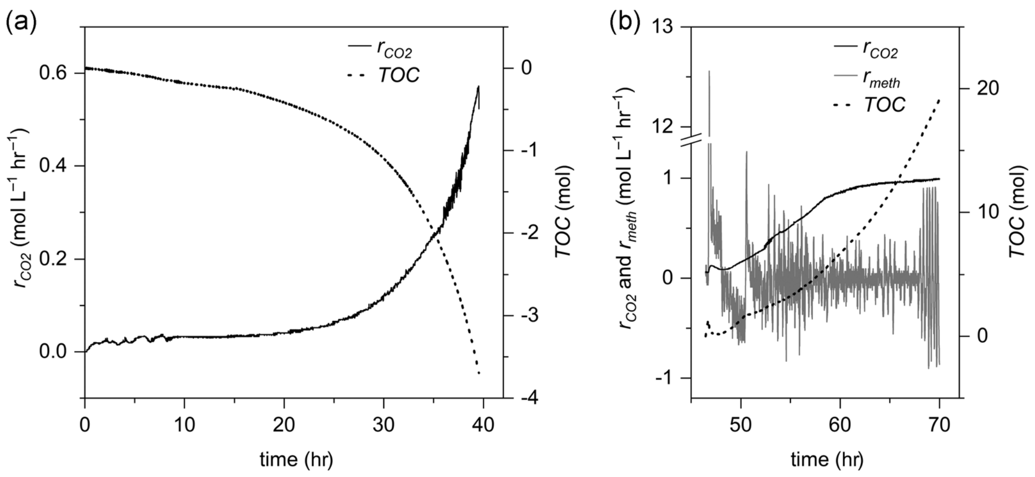

FIGURE 5 Illustration of the carbon balance for (a) the batch and (b) fed-batch phase for an exemplary process run. The carbon dioxide emission rate, r_{CO2}, and—in the fed-batch phase, additionally—the methanol reaction rate, r_{meth}, are used to calculate the formation rate of total organic carbon, r_{TOC}, as in (6). Multiplication of r_{TOC} with V_{total} and numeric integration as in (7) result in the cumulative amount of total organic carbon, *TOC*

via MLR in the batch phase, *TOC* and V_{base} are used as inputs and the biomass amount *X* (determined offline as dry cell weight) as output. The prediction equation is formulated as follows, where b_0, b_1, and b_2 are the model coefficients:

$$X = b_0 + b_1\, TOC + b_2\, V_{base}. \tag{8}$$

In the transition phase, no significant cell growth or decline was observed, so b_0 was set to the value of *X* at the end of the batch phase ($X_{batchend}$) and b_1 and b_2 were set to 0. In the fed-batch phase, b_0 was set to $X_{batchend}$ and b_1 and b_2 were determined analogously to the methods used in the batch phase.

The regression step in (8) is related to the total liquid volume inside the bioreactor, V_{total}. To determine the biomass concentration c_X, the biomass amount *X* is divided by V_{total}, as in the following equation:

$$c_X = \frac{X}{V_{total}}. \tag{9}$$

3.3.2 | Cross-validation approach for model calibration and validation

The model was calibrated and validated by a batch-wise cross-validation approach. The six data sets used for developing the biomass soft sensor were iteratively partitioned into two-thirds of calibration and one-third of validation data sets. This resulted in a total of 6!/(2!4!) = 15 different combinations of complementary subsets for cross-validation. For each iteration step, R^2, root mean squared error (*RMSE*), and normalized root mean squared error (*NRMSE*) of cross-validation were calculated separately for the batch and fed-batch phase as well as for the entire process (including the transition phase). R^2 is calculated for the four calibration data sets. The *NRMSE* in the following equation is the normalized version of *RMSE* and is calculated from reference measurements *y* and predictions $\hat{y}$ of the two validation data sets. y_{max} and y_{min} are the maximum and minimum values of *y*, respectively, and *m* is the number of data points in *y*:

$$NRMSE = \frac{1}{y_{max} - y_{min}} \sqrt{\frac{1}{m} \sum\nolimits_{i=1}^{m} (\hat{y}_i - y_i)^2}. \tag{10}$$

The use of separate subsets for internal and external (holdout) validation (OECD, 2014) does not appear to be practicable because

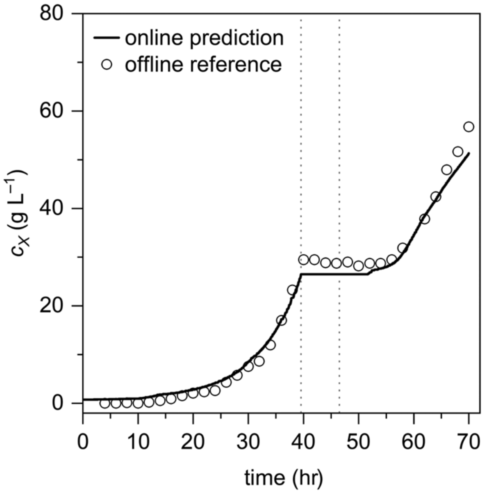

FIGURE 6 Online prediction of biomass concentration c_X during batch, transition, and fed-batch phase for an exemplary process run using the hybrid-model-based soft sensor. Both the batch and the fed-batch phase start with a lag phase after which cells grow exponentially (batch phase) or linearly (fed-batch phase). The two dashed, gray lines indicate the switches from batch to transition phase (39.6 hr) and from transition to fed-batch phase (46.5 hr), respectively

Process phase	$b_0 \pm CI_{.95,b_0}$ (g)	$b_1 \pm CI_{.95,b_1}$ (g·mol^{-1})	$b_2 \pm CI_{.95,b_2}$ (g·L^{-1})
Batch phase	4.60 ± 2.54	−13.69 ± 9.81	701.96 ± 79.62
Transition phase	Replaced by $X_{batchend}$	0	0
Fed-batch phase	Replaced by $X_{batchend}$	2.63 ± 1.57	1074.05 ± 94.28

TABLE 3 Results for model coefficients b_0, b_1, and b_2 in (8) and the corresponding 95% confidence intervals, $CI_{.95}$. In the transition and fed-batch phase, b_0 is set to the value of X at the end of the batch phase ($X_{batchend}$)

the total number of data sets that are available for model calibration and validation is too small (n = 6).

3.4 | Online prediction of biomass using the multilevel phase detection

The multilevel phase detection algorithm resulted in a 100% correct hit rate for the detection of the end of the batch phase. On average, the phase end was detected 2.56 measurements (corresponding to 77 s) before the minimum $d\sigma_{CO2}/dt$ was reached—which was defined as the end of the batch phase.

The arithmetic means for R^2, $RMSE$, and $NRMSE$ are calculated using the abovementioned 15 combinations of n = 6 data sets. The mean R^2 for the batch and fed-batch phases is .97 and .95, respectively; the mean R^2 for the entire process is .96. The mean $RMSE$ for the batch and fed-batch phase is 1.14 and 5.05 g · L^{-1}, respectively; the mean $RMSE$ for the entire process is 3.57 g · L^{-1}, which results in a mean $NRMSE$ of 5.52%.

Figure 6 shows the results for the online prediction of biomass concentration based on the hybrid-model-based soft sensor. The figure shows validation data of one iteration of the cross-validation for an exemplary process run. The underestimation of the online prediction at 40–52 hr and after 64 hr is due to an error in biomass prediction at the end of the batch phase that entails prediction errors in the transition phase.

The results for the model coefficients b_0, b_1, and b_2 in (8) are listed in Table 3. As described above, these model coefficients are used to determine the biomass amount X, which needs to be divided by V_{total} to calculate the biomass concentration c_X. V_{total} varies between $V_{total} = V_{start} = 10.00$ L and on average $V_{total} = 12.56$ L (n = 6) at the end of the cultivation. In the batch phase, the intercept b_0 describes the initial biomass from inoculation. As mentioned above, b_0 was in the transition and fed-batch phase replaced by $X_{batchend}$, which has a mean of 253.47 g (n = 6). The model coefficient for TOC, b_1, is negative in the batch phase because here the carbon balance in (6) is simplified to $r_{TOC} = -r_{CO2}$ (boundary for the balancing is the bioreactor system) and thus TOC in (7) decreases with increasing CO$_2$ emission and biomass, respectively. In the fed-batch phase, in which methanol is fed to the bioreactor, TOC correlates positively with X. The model coefficient for V_{base}, b_2, is positive for both the batch and fed-batch phase. In the fed-batch phase, b_2 is more than 50% higher than in the batch phase, which means that more than 50% base is necessary to maintain the pH setpoint on glycerol compared to methanol. The soft sensor's model coefficients switch automatically depending on the current process phase. The differences in the model coefficients b_1 and b_2 between the individual process phases indicate the necessity for the adaption of model coefficients with changing process phases.

The accuracy of the estimates of the model coefficients is given by the corresponding 95% confidence intervals, $CI_{.95}$ (Table 3). None of the $CI_{.95}$ contains the value zero, which is considered to be a primary indication that the model inputs are to a certain degree significant to the model output, biomass. The width of $CI_{.95}$ relative to the absolute value of the model coefficient is a further indicator for the quality of the regression and hence for the uncertainty of the soft sensor model (Fernandes et al., 2012). For b_0, the ratio of the width of $CI_{.95}$ to the absolute value of the model coefficient is 55%; for b_1, the ratio is 72% and 60% for the batch and fed-batch phase, respectively; for b_2, the ratio is 11% and 9% for the batch and fed-batch phase, respectively.

The contribution of the model coefficients b_0, b_1, and b_2 to the prediction of X is illustrated in Figure 7 for an exemplary process run. Here, each model coefficient's contribution was determined by disassembling the linear combination in (8) and dividing each model coefficient's prediction by the total model prediction. As expected, the contribution of b_0 starts with an initial value of 100% at the process start and decreases relative to the contribution increases of b_1 and b_2. Until the end of the batch and fed-batch phase, the

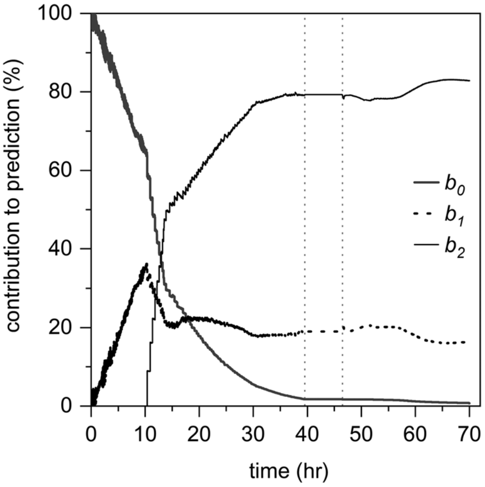

FIGURE 7 Contribution of model coefficients b_0, b_1, and b_2 to the prediction of biomass amount X during batch, transition, and fed-batch phase for an exemplary process run. The two dashed, gray lines indicate the switches from batch to transition phase (39.6 hr) and from transition to fed-batch phase (46.5 hr), respectively. The soft sensor updates its model coefficients automatically for the three distinct process phases

contribution of b_0 falls to values of 1.72% and 0.57%, respectively. Since b_0 is in the transition and fed-batch phase replaced by $X_{batchend}$, the contributions to $X_{batchend}$ (18.99% for b_1 and 79.29% for b_2) are used as offset for the contributions of b_1 and b_2 throughout the latter process phases. The contribution of b_1 initially rises to a maximum of 36.41% approximately at the end of the lag phase and reaches contributions of 18.99% and 18.26%, respectively, at the end of the batch and fed-batch phase. The contribution of b_2 starts to rise when base is first fed to the bioreactor (see Figure 1) and reaches values of 79.29% and 81.18%, respectively, at the end of the batch and fed-batch phase. It can be concluded from these results that, approximately after the end of the lag phase, V_{base} has a higher impact on biomass prediction than *TOC*. This result is consistent with the apparent high collinearity of V_{base} (see Figure 1) and c_X (see Figure 6).

4 | CONCLUSIONS

As mentioned at the beginning of this paper, several challenges can arise when attempting to develop soft sensors. One of these is specific to *P. pastoris* bioprocesses with distinct process phases such as batch, transition, and fed-batch phase. The underlying principles of prediction models for biomass are related to the inherent biological relations (Chen et al., 2004), which differ depending on the substrate used in the specific process phase. The fundamental differences in the metabolism of different carbon sources have a visible impact on CO_2 emission and the consumption of pH correction agent (see Figure 1), which are two of the main model inputs used in this study. For multiphase processes with more than one substrate, this means that the probability of finding a single model that captures the information necessary for prediction of biomass is rather low.

This study demonstrates the application of a multilevel phase detection algorithm to determine the end of the batch phase (glycerol depletion) online. In every tested case, the algorithm provided the correct end time of the batch phase. The detection of this end time was used to trigger the transition phase and the subsequent methanol induction. The knowledge about the significantly reduced CO_2 emission that comes with glycerol depletion was effectively utilized. Specifically, the stoichiometric restrictions concerning the cumulative amount of supplied base (trigger 1) and emitted CO_2 (trigger 2) were used to increase robustness of the third trigger (timewise derivative of the off-gas CO_2 signal). The usage of purely data-driven approaches for process phase detection (e.g., Abonyi, Feil, Nemeth, & Arva, 2005; Ye, Wang, & Yang, 2017) did not appear practicable in this case because only a relatively small number of data sets (nine) were available for the development of the phase detection algorithm.

The output of the phase detection algorithm was used to switch the parameters of the prediction model online. The prediction model was calibrated offline using a hybrid-model-based approach. The output of the mechanistic part (carbon balance) is fed to the data-driven part (MLR) to provide an accurate prediction of the biomass concentration. The process runs were conducted under the same operating conditions (initial glycerol concentration, constant setpoints for methanol, pH, dissolved oxygen, temperature, and pressure). However, the process runs and corresponding data sets used in this study were subject to variance of initial biomass concentration, which in turn resulted from the variability of the preculture. Further, errors in the methanol control, such as an initial overshoot or a deviation of the measured methanol concentration to the setpoint (Figure 1), occurred and additionally increased the variance between the data sets. Despite this variance between the used data sets, model evaluation results in a mean relative prediction error of 5.52% and R^2 of .96 for the entire process. These two evaluation criteria are of similar magnitude to those of other biomass soft sensors for *P. pastoris* fed-batch processes (Beiroti, Aghasadeghi, Hosseini, & Norouzian, 2019; Crowley, Arnold, Wood, Harvey, & McNeil, 2005; Fazenda et al., 2013; Surribas, Geissler, et al., 2006; Surribas, Montesinos, & Valero, 2006). In the approach presented here, however, the soft sensor is adaptable online to the different process phases, and no cost-intensive spectroscopic measurement system is necessary. The robustness of the soft sensor with regard to different process conditions (e.g., variation of methanol, pH, and temperature setpoint) was not in the scope of this study. These investigations are subject of future research.

The main constraint of the presented soft sensor is that the prediction in the transition and fed-batch phase is directly dependent on the prediction result in the batch phase. This is due to the passing on of the biomass prediction at the end of the batch phase ($X_{batchend}$) as a start value for the prediction models of the subsequent phases. The effect of error propagation can be visualized by considering the slight decrease of R^2 and increase of prediction error between batch and fed-batch phase. It should further be noted that the carbon balance in the individual phases depends on constant ratios of biomass formation, CO_2 emission, and—in the fed-batch phase—methanol metabolization. Longer periods of substrate limitation or metabolite inhibition would impede an accurate biomass prediction if these scenarios are not included in the data sets used for model calibration.

Knowledge-based relationships were combined with data-driven methodology in this study. No general statement can be made here about whether mechanistic, data-driven, or hybrid approaches are superior because the choice is strongly dependent on the available process knowledge and measurement systems (offline/online) as well as the number of data sets and data points (Solle et al., 2017). However, in this study, the usage of a hybrid approach appears to be suitable because of the benefits from both components of it. This is due to the availability of the necessary online measurement systems for capturing the information relevant for modeling biomass (off-gas CO_2, methanol, cumulative base feed) and, on the other hand, the relatively small number of data sets (six) for model calibration and validation.

The transferability of the developed phase-dependent soft sensor to other fed-batch cultivations with different *P. pastoris* strains, control strategies, media, and process parameters must be investigated in future research. It is supposed that the presented approaches for process phase detection and hybrid-model-based prediction are transferable to any methanol-induced *P. pastoris* process provided that the carbon source used for initially generating biomass (glycerol, glucose, etc.) is not co-fed to methanol.

The developed algorithm for process phase detection and the prediction model were implemented as a soft sensor for the online monitoring of biomass. The soft sensor can be used for quality control and as input to the process control system, for example, for methanol control.

NOMENCLATURE

b_0, b_1, b_2	model coefficients (g, g · mol^{-1}, g · L^{-1})
c	molar or mass concentration
$CI_{.95}$	95% confidence interval for model coefficients (g, g · mol^{-1}, g · L^{-1})
c_{meth}	methanol concentration (mol · L^{-1})
c_P	extracellular protein concentration (g · L^{-1})
c_X	biomass concentration (g · L^{-1})
m	number of data points in y (-)
mean	arithmetic mean of trigger variable (L or % or % · hr^{-1})
n	number of data sets (-)
NRMSE	normalized root mean squared error (%)
p	pressure (bar)
Q_{air}	air flow rate (L · hr^{-1})
r	timewise rate
R	universal gas constant (L · bar · mol^{-1} · K^{-1})
R^2	coefficient of determination (-)
RMSE	root mean squared error (g · L^{-1})
r_{CO2}	carbon dioxide emission rate (mol · L^{-1} · hr^{-1})
r_{meth}	methanol reaction rate (mol · L^{-1} · hr^{-1})
$r_{meth,acc}$	methanol accumulation rate (mol · L^{-1} · hr^{-1})
$r_{meth,in}$	methanol feed rate (mol · L^{-1} · hr^{-1})
r_{TOC}	formation rate of total organic carbon (mol · L^{-1} · hr^{-1})
SF	constant safety factor (-)
std	standard deviation of trigger variable (L or % or % · hr^{-1})
T	temperature (K)
t	time (hr)
threshold	threshold for trigger variable (L or % or % · hr^{-1})
TOC	cumulative amount of total organic carbon (mol)
V_{afoam}	cumulative volume of antifoam (L)
V_{base}	cumulative volume of base (L)
V_{meth}	cumulative volume of methanol (L)
$V_{samples}$	cumulative volume of samples (L)
V_{start}	start liquid volume inside bioreactor after inoculation (L)
V_{total}	total liquid volume inside bioreactor (L)
X	biomass amount (g)
$X_{batchend}$	biomass amount at the end of the batch phase (g)
y	reference measurement (g · L^{-1})
$\hat{y}$	prediction (g · L^{-1})
y_{max}/y_{min}	maximum/minimum values of reference measurements y (g · L^{-1})
α	index for gas inlet of the bioreactor (-)
σ	volume concentration
σ_{CO2}	off-gas CO_2 concentration (%)
$d\sigma_{CO2}/dt$	timewise derivative of the off-gas CO_2 concentration (% · hr^{-1})
σ_{O2}	off-gas O_2 concentration (%)
ω	index for gas outlet of the bioreactor (-)

ACKNOWLEDGMENTS

This study was supported by the German Federal Ministry of Education and Research (project 031B0475E). The authors would like to thank Jens Traenkle's group (Bayer AG) for providing the BaychroMAT® autosampler.

ORCID

Vincent Brunner http://orcid.org/0000-0002-3310-2236

REFERENCES

Abonyi, J., Feil, B., Nemeth, S., & Arva, P. (2005). Modified Gath–Geva clustering for fuzzy segmentation of multivariate time-series. *Fuzzy Sets and Systems*, *149*, 39–56.

Bakirov, R., Gabrys, B., & Fay, D. (2017). Multiple adaptive mechanisms for data-driven soft sensors. *Computers & Chemical Engineering*, *96*, 42–54.

Beiroti, A., Aghasadeghi, M. R., Hosseini, S. N., & Norouzian, D. (2019). Application of recurrent neural network for online prediction of cell density of recombinant *Pichia pastoris* producing HBsAg. *Preparative Biochemistry and Biotechnology*, *49*, 352–359.

Birle, S., Hussein, M., & Becker, T. (2013). Fuzzy logic control and soft sensing applications in food and beverage processes. *Food Control*, *29*, 254–269.

Cereghino, G. P. L., Cereghino, J. L., Ilgen, C., & Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Current Opinion in Biotechnology*, *13*, 329–332.

Chen, L. Z., Nguang, S. K., Li, X. M., & Chen, X. D. (2004). Soft sensors for on-line biomass measurements. *Bioprocess and Biosystems Engineering*, *26*, 191–195.

Crowley, J., Arnold, S. A., Wood, N., Harvey, L. M., & McNeil, B. (2005). Monitoring a high cell density recombinant *Pichia pastoris* fed-batch bioprocess using transmission and reflectance near infrared spectroscopy. *Enzyme and Microbial Technology*, *36*, 621–628.

Fazenda, M. L., Dias, J. M., Harvey, L. M., Nordon, A., Edrada-Ebel, R., LittleJohn, D., & McNeil, B. (2013). Towards better understanding of an industrial cell factory: Investigating the feasibility of real-time metabolic flux analysis in *Pichia pastoris*. *Microbial Cell Factories*, *12*, 51.

Fernandes, R. L., Bodla, V. K., Carlquist, M., Heins, A. L., Lantz, A. E., Sin, G., & Gernaey, K. V. (2012). Applying mechanistic models in bioprocess development. *Measurement, monitoring, modelling, and control of bioprocesses* (pp. 137–166). Berlin: Springer.

Fortuna, L., Graziani, S., & Xibilia, M. G. (2009). Comparison of soft-sensor design methods for industrial plants using small data sets. *IEEE Transactions on Instrumentation and Measurement*, *58*, 2444–2451.

Gao, M.-J., Zheng, Z.-Y., Wu, J.-R., Dong, S.-J., Li, Z., Jin, H., … Lin, C.-C. (2012). Improvement of specific growth rate of *Pichia pastoris* for effective porcine interferon-α production with an on-line model-based glycerol feeding strategy. *Applied Microbiology and Biotechnology*, *93*, 1437–1445.

Gonzaga, J., Meleiro, L. A. C., Kiang, C., & Maciel Filho, R. (2009). ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Computers & Chemical Engineering*, *33*, 43–49.

Harms, J., Wang, X., Kim, T., Yang, X., & Rathore, A. S. (2008). Defining process design space for biotech products: Case study of *Pichia pastoris* fermentation. *Biotechnology Progress*, *24*, 655–662.

Harms, P., Kostov, Y., & Rao, G. (2002). Bioprocess monitoring. *Current Opinion in Biotechnology*, *13*, 124–127.

Jahic, M., Veide, A., Charoenrat, T., Teeri, T., & Enfors, S. O. (2006). Process technology for production and recovery of heterologous proteins with *Pichia pastoris*. *Biotechnology Progress*, *22*, 1465–1473.

Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., & Lübbert, A. (2007). Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production. *Journal of Biotechnology*, *128*, 858–867.

Kadlec, P., Gabrys, B., & Strandt, S. (2009). Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, *33*, 795–814.

Kalos, A., Kordon, A., Smits, G., & Werkmeister, S. (2003). Hybrid model development methodology for industrial soft sensors. In *Proceedings of the 2003 American control conference* (pp. 5417-5422). IEEE.

Kano, M., & Fujiwara, K. (2012). Virtual sensing technology in process industries: Trends and challenges revealed by recent industrial applications. *Journal of Chemical Engineering of Japan*, *46*(1), 1–17.

Kordon, A., Smits, G., Kalos, A., & Jordaan, E. (2003). Robust soft sensor development using genetic programming. *Nature-Inspired Methods in Chemometrics*, *23*, 69–108.

Luttmann, R., Bracewell, D. G., Cornelissen, G., Gernaey, K. V., Glassey, J., Hass, V. C., … Mandenius, C. F. (2012). Soft sensors in bioprocessing: A status report and recommendations. *Biotechnology Journal*, *7*, 1040–1048.

Munch, G., Schulte, A., Mann, M., Dinger, R., Regestein, L., Rehmann, L., & Büchs, J. (2020). Online measurement of CO2 and total gas production in parallel anaerobic shake flask cultivations. *Biochemical Engineering Journal*, *153*, 107418.

OECD. (2014). *Guidance document on the validation of (quantitative) structure-activity relationship [(Q) SAR] models*. OECD Publishing.

Schügerl, K. (2001). Progress in monitoring, modeling and control of bioprocesses during the last 20 years. *Journal of Biotechnology*, *85*, 149–173.

Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., & Steckenreiter, T. (2017). Between the poles of data-driven and mechanistic modeling for process operation. *Chemie Ingenieur Technik*, *89*, 542–561.

Stratton, J., Chiruvolu, V., & Meagher, M. (1998). High cell-density fermentation. In D. R. Higgins & J. M. Cregg (Eds.), *Pichia protocols* (pp. 107–120). Totowa, NJ: Humana Press.

Surribas, A., Geissler, D., Gierse, A., Scheper, T., Hitzmann, B., Montesinos, J. L., & Valero, F. (2006). State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris*. *Journal of Biotechnology*, *124*, 412–419.

Surribas, A., Montesinos, J. L., & Valero, F. F. (2006). Biomass estimation using fluorescence measurements in *Pichia pastoris* bioprocess. *Journal of Chemical Technology & Biotechnology: International Research in Process, Environmental & Clean Technology*, *81*, 23–28.

Takors, R. (2013). Industrielle Mikrobiologie, *Bioverfahrenstechnik. Industrielle mikrobiologie* (pp. 19–41). Berlin: Springer.

Ye, A. X., Wang, B. P., & Yang, C. Z. (2017). Time sequential phase partition and modeling method for fault detection of batch processes. *IEEE Access*, *6*, 1249–1260.

Zhang, H. (2009). Software sensors and their applications in bioprocess. In *Computational intelligence techniques for bioprocess modelling, supervision and control* (pp. 25–56). Berlin: Springer.

How to cite this article: Brunner V, Siegl M, Geier D, Becker T. Biomass soft sensor for a *Pichia pastoris* fed-batch process based on phase detection and hybrid modeling. *Biotechnology and Bioengineering*. 2020;117:2749–2759. https://doi.org/10.1002/bit.27454