



A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations

Spann, Robert; Roca, Christophe ; Kold, David ; Eliasson Lantz, Anna; Gernaey, Krist V.; Sin, Gürkan

Published in:
Biochemical Engineering Journal

Link to article, DOI:
[10.1016/j.bej.2018.03.016](https://doi.org/10.1016/j.bej.2018.03.016)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. *Biochemical Engineering Journal*, 135, 49-60.
<https://doi.org/10.1016/j.bej.2018.03.016>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Title: A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations

Authors: Robert Spann, Christophe Roca, David Kold, Anna Eliasson Lantz, Krist V. Gernaey, Gürkan Sin



PII: S1369-703X(18)30099-8
DOI: <https://doi.org/10.1016/j.bej.2018.03.016>
Reference: BEJ 6919

To appear in: *Biochemical Engineering Journal*

Received date: 21-11-2017
Revised date: 20-2-2018
Accepted date: 16-3-2018

Please cite this article as: Spann R, Roca C, Kold D, Lantz AE, Gernaey KV, Sin G, A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations, *Biochemical Engineering Journal* (2018), <https://doi.org/10.1016/j.bej.2018.03.016>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations

Robert Spann^a, Christophe Roca^b, David Kold^b, Anna Eliasson Lantz^c, Krist V. Gernaey^a,
Gürkan Sin^{a*}

^a Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Soeltofts Plads Building 229, 2800 Kgs. Lyngby, Denmark

^b Chr. Hansen A/S, Boege Allé 10-12, 2970 Hoersholm, Denmark

^c PILOT PLANT, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Soeltofts Plads Building 229, 2800 Kgs. Lyngby, Denmark

*** Corresponding author:**

Gürkan Sin

Technical University of Denmark

Søltofts Plads Building 229

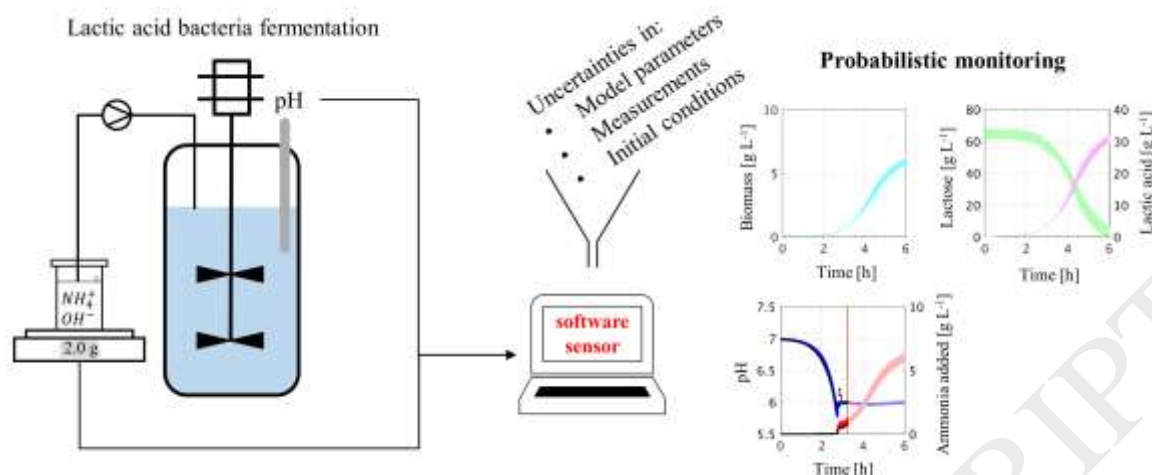
2800 Kgs. Lyngby, Denmark

Tel: +45 45252806

E-mail address: gsi@kt.dtu.dk

Declaration of interest: none

Graphical abstract



Highlights

- A model-based soft sensor for the monitoring of lactic acid bacteria is proposed
- Predictions based on limited available on-line measurements: base addition and pH
- Coupling of a biokinetic model and a mixed weak acid/base model
- Uncertainties are quantified and accounted for using Monte Carlo simulations
- Probabilistic prediction for on-line risk-based monitoring and control

Abstract

A probabilistic soft sensor based on a mechanistic model was designed to monitor *S. thermophilus* fermentations, and validated with experimental lab-scale data. It considered uncertainties in the initial conditions, on-line measurements, and model parameters by performing Monte Carlo simulations within the monitoring system. It predicted, therefore, the probability distributions of the unmeasured states, such as biomass, lactose, and lactic acid concentrations. To this end, a mechanistic model was developed first, and a statistical parameter estimation was performed in order to assess parameter sensitivities and

uncertainties. The model coupled a biokinetic and a mixed weak acid/base model to predict biological variables and chemical variables like the pH, respectively. In the soft sensor, the limited available on-line measurements, namely the quantity of added ammonia and pH, were used to update the model parameters that were then used as input to the mechanistic model. The soft sensor predicted both the current state variables, as well as the future course of the fermentation, e.g. with a relative mean error of the biomass concentration of 8 %. This successful implementation of a process analytical technology monitoring system opens up further opportunities, including for on-line risk-based monitoring and control applications.

Keywords: lactic acid bacteria (LAB) fermentation; modelling; soft sensor for monitoring; uncertainty analysis; process risk assessment; process analytical technology (PAT)

1. Introduction

Lactic acid bacteria (LAB) are used as starter cultures in the dairy industry, to produce probiotics, lactic acid, and exopolysaccharides [1,2]. *Streptococcus thermophilus* strains are aerotolerant, homofermentative LAB and traditionally used as fermentation starter cultures for yogurt and cheese production. The bacteria are produced in batch and fed-batch fermentations, and real-time monitoring of the process is needed in order to understand and optimize the production process. However, robust in-line sensors for key process variables, like biomass, substrate, and lactic acid concentrations, are not available in the required concentration range due to the high complexity of the fermentation system [3]. This makes the real-time quantification of key process variables challenging. The process analytical technology (PAT) guidance from the FDA [4] requested already the development of real-time monitoring and control tools. The tools are meant to enhance the on-line monitoring and control capabilities. Hence, process conditions could be adjusted in real time to assure quality requirements, instead of relying solely on the end product quality control. Although the guidelines were originally intended for the (bio-) pharmaceutical industry, they have also been applied in other life science industries like the food industry.

Soft sensors, which utilize the on-line available measurements, are applied to predict the unknown state variables and monitor the fermentation process in real time [5–7]. There are, generally spoken, data-based and model-based soft sensors, whereas also other approaches exist. Chemometric methods like principal component analysis (PCA) may be applied in data-based soft sensors [8]. Model-based soft sensors can for example be based on mechanistic understanding using first principles models (e.g. the mass balance), or empirical models, when the details of the process are not understood sufficiently. The development of first principles models is based on a fundamental process understanding and mechanistic

models may be implemented. Even though the development of mechanistic models might be time consuming, we prefer to use mechanistic models since they have many advantages over black-box models, e.g. that they can be reused and applied to multiple processes by updating the model parameters [7].

Soft sensors rely typically on available on-line and at-line measurements, such as pH, conductivity, dissolved oxygen, heat generation, acid/base addition for pH control, and exhaust gas analysis. Different spectroscopic measurements, e.g. near-infrared [9], multi-wavelength [10], and Raman [11] spectroscopy have also been used beside other methods in data-based soft sensor applications in fermentations. Mears et al. [12] applied a model-based soft sensor for on-line monitoring to a filamentous fungal fermentation at pilot scale using exhaust gas measurements and ammonia addition, and predicted various state variables (biomass, product, dissolved oxygen, k_La among others). However, especially aerotolerant, homofermentative LAB fermentations lack various on- and at-line measurements, such as exhaust gas measurements, and rely solely on conductivity, pH, and base addition measurements.

When developing and applying mechanistic models for bioprocesses it is good modelling practice for PAT applications to analyze the reliability of the model [13,14]. Unfortunately, models describing LAB fermentations rarely provide reported results of identifiability, sensitivity, or uncertainty analysis, e.g. confidence intervals of the estimated parameters. If a model with unreliable parameters is applied as a soft sensor, predictions will be doubtful and the results could lead to questionable interpretations. Furthermore, a deterministic model implementation may lead to a good fit, but does not take the imperfect knowledge, i.e. uncertainties of model parameters and measurements into account. Several studies

implemented soft sensors to monitor LAB [15–18], but they did not consider uncertainties in the model structure.

The aim of this study was therefore to design and evaluate a probabilistic model-based soft sensor in order to monitor *S. thermophilus* fermentations. To this end, a mechanistic model was first developed and validated, and then used as soft sensor for monitoring at lab scale. A statistical parameter estimation was performed to analyze parameter uncertainties. The soft sensor comprised a data reconciliation module, a parameter update module and the dynamic model. The data reconciliation and parameter update module updated model parameters based on the available on-line measurements. One major challenge of this study was that only pH and ammonia addition measurements were available on-line, whereas e.g. exhaust gas measurements were not available. The dynamic model consisted of a biokinetic and a chemical model. The biokinetic model described substrate consumption, biomass growth, and lactic acid secretion while the chemical model comprised a mixed weak acid/base system with the purpose to predict the pH. Monte Carlo simulations of the dynamic model were performed within the monitoring system to account for uncertainties in the lactose (substrate) concentration, ammonia addition rate, and model parameters. The output of the monitoring system was consequently a probability distribution of the state variables.

2. Materials and Methods

2.1 Fermentation conditions

Streptococcus thermophilus batch fermentations were performed in 2 L stirred tank bioreactors (Biostat[®] B, Sartorius AG, Germany) at 300 rpm, 40 °C, and with nitrogen headspace gassing. The pH was controlled by adding 24 % ammonia solution. 10 batch fermentations were performed under different cultivation conditions (initial lactose concentration 20 or 65 g L⁻¹, pH 5.5 – 7.0) and used for the parameter estimation, model

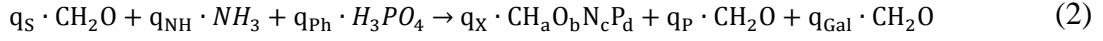
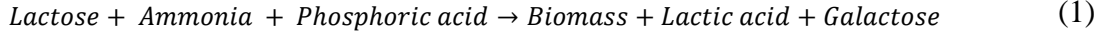
validation, and implementation of the monitoring system (see Table 4 in the Results and Discussion section). The pH (EasyFerm Bio VP 225, Hamilton Robotics, Reno, NV) and ammonia addition (balance value) were measured on-line. The fermentation medium contained 20 or 65 g L⁻¹ lactose, 10 g L⁻¹ casein hydrolysate, 12 g L⁻¹ yeast extract, 11.5 mM K₂HPO₄, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM MgSO₄, and 0.3 mM MnSO₄.

2.2 Off-line analysis

Sugars and organic acids were quantified from filtered samples (filter pore size: 0.2 µm) in an HPLC system (Dionex UltiMate 3000, Thermo Fisher Scientific, Waltham, MA). It was equipped with an Aminex[®] HPX-87H column (Bio-Rad Laboratories, Hercules, CA) and a refractive index detector (ERC RefractoMax 520), and run with 5 mM H₂SO₄ at a flow rate of 0.6 mL min⁻¹ at 50 °C according to suppliers instructions. Samples were diluted 1:4 with 5 mM H₂SO₄ prior to analysis. Dry cell weight was quantified with replicates of 2 mL cell broth, which were taken in sampling tubes, centrifuged, washed with 0.9 % (w/v) NaCl solution, dried at 70 °C for 24 h, and weighted. Ammonia and phosphate were quantified with the cuvette tests LCK302, LCK303, and LCK350 (Hach[®], Manchester, Great Britain).

2.3 Biological model

The dynamic biokinetic model described the evolution of the state variables such as biomass, lactose, and lactic acid of the *S. thermophilus* fermentation. The model was based on the global stoichiometric process equation [19] (Eq. 1 – 2). The biomass growth rate was modelled as a function depending on the lag-time (f_{lag}), lactose inhibition and limitation (f_s) [20], lactate inhibition (f_p) [21], and the pH (f_{pH}) [22] (Eq. 3 – 4). It was assumed that only the dissociated form of lactic acid was growth inhibiting under the investigated pH conditions according to the studies of Schepers et al. [22] and Amrane and Prigent [23]. A biomass composition of $CH_{1.95}O_{0.63}N_{0.22}P_{0.02}$ [24] was assumed in the present study.



$$\frac{dC_X}{dt} = \mu_{\text{max}} \cdot f_{\text{lag}} \cdot f_S \cdot f_P \cdot f_{\text{pH}} \cdot C_X \quad (3)$$

$$\frac{dC_X}{dt} = \mu_{\text{max}} \cdot \left(1 - e^{-\frac{t}{t_{\text{lag}}}}\right) \cdot \frac{C_S}{C_S + K_S + \frac{C_S^2}{K_I}} \cdot \frac{1}{1 + e^{K_{P,La}(C_{LA} - K_{La1})}} \cdot e^{-\left(\frac{(\text{pH}_{\text{opt}} - \text{pH})^2}{\sigma_{\text{pH}}^2}\right)} \cdot C_X \quad (4)$$

Where K_{La1} was dependent on the pH:

$$K_{La1} = K_{La} \cdot \frac{1}{1 + e^{K_{P,pH1} \cdot (\text{pH} - K_{P,pH2})}} \quad (5)$$

An amended Luedeking and Piret equation [25] that takes only the growth dependent lactic acid synthesis into account was used [26]:

$$\frac{dC_P}{dt} = \alpha \cdot \frac{dC_X}{dt} \quad (6)$$

The lactose consumption is the sum of the biomass growth and the lactic acid synthesis rate considering the secretion of galactose (Y_{gal}) since the studied strain metabolizes only glucose and secretes galactose under the present cultivation conditions:

$$\frac{dC_S}{dt} = -(1 + Y_{\text{gal}}) \cdot \left(\frac{dC_X}{dt} + \frac{dC_P}{dt}\right) \quad (7)$$

A P-controller with a controller gain (K_P) of 10 mol was applied to maintain the pH at the set point value by adding ammonia solution:

$$\text{NH}_4\text{OH}_{\text{add}} = K_P \cdot (\text{pH}_{\text{set}} - \text{pH}) \quad (8)$$

The model was implemented and solved in MATLAB[®] (The MathWorks[®], Natick, MA) using the solver ode15s. ode15s was chosen because the model contains slow (e.g. biomass growth) and fast time constants (mixed weak acid/base model, see below) resulting in a stiff system of differential equations.

2.4 Mixed weak acid/base model

The purpose of the mixed weak acid/base model was to predict the pH during the fermentation. It comprised the dissociation reactions of the charged compounds in the fermentation broth as described in Musvoto et al. [27] (Table 1). The dissociation reactions of ammonia, phosphoric acid, lactic acid, carbonic acid, water, and an unspecified compound Z were considered. Z accounted for the unknown compounds in the fermentation broth, such as amino acids. The pK_a values were derived from Dawson [28] and Loewenthal et al. [29] (Table 2). The activity coefficients were calculated by a modification of the Debye-Hückel theory from Davies [30]:

$$\log(f_i) = -1.825 \cdot 10^6 \cdot (78.3 \cdot T)^{-1.5} \cdot z_i^2 \cdot \left(\frac{\sqrt{I}}{1 + \sqrt{I}} - 0.3 \cdot I \right) \quad (9)$$

With the ionic strength (I):

$$I = \frac{1}{2} \sum_i z_i^2 C_i \quad (10)$$

The implemented stoichiometric matrix may be found in the Supplementary Material.

2.5 Parameter estimation

The parameter estimation was performed to fit the experimental lactose, biomass, and lactic acid concentration measurements using the maximum likelihood estimation method from Seber and Wild [31]. The model was fitted to five fermentations, which were controlled at different constant pH (1x pH 5.5, 2x pH 6.0, 1x pH 6.5, and 1x pH 7.0) and were performed with an initial lactose concentration of 65 g L^{-1} . For the parameter estimation, the pH was held constant at the set point in the simulation, and the mixed weak acid/base model was not considered in order to obtain parameter estimates that were independent of the mixed weak/acid base system. The parameter estimation followed the methodology from Sin and Gernaey [32] as described in Spann et al. [33]. Initial parameter estimates were taken from literature [20–22] (Supplementary Table S1). Sensitivity and identifiability analysis were

conducted to find an identifiable parameter subset for regression [32]. Once the regression was completed, the confidence intervals of the estimated parameters were derived from a linear approximation method using the Jacobian matrix of the parameter estimation [34]. The parameter estimation was conducted in MATLAB with the nonlinear least-squares solver lsqnonlin. In the objective function, the weighted error of the model predictions was calculated for the three concentrations lactose, biomass, and lactic acid at all measured time points i (Eq. 11). The residuals vector then contained the weighted error vectors of the three states j .

$$Error_i = \left| \frac{\hat{y}_i - y_{meas,i}}{w_j} \right| \quad (11)$$

where w_j is the maximum value of each specific component, here 65 g L^{-1} for lactose, 30 g L^{-1} for lactic acid, and 6 g L^{-1} for biomass. For model simplification purposes, the lag-time parameter, t_{lag} , was described as a pH dependent distribution, in order to account for the different lag-times observed for fermentations having a different pH set point (Eq. 12). This approach simplifies the model complexity significantly and requires the estimation of only 2 parameters, instead of 5 parameters that would have been needed, if t_{lag} was fitted for each fermentation separately.

$$t_{lag} = 2/e \frac{(pH_{opt,lag} - pH)^2}{\sigma_{lag}^2} \quad (12)$$

The uncertainty of the estimated parameters was quantified with the relative error (RE) between the standard deviation of the parameter estimate with respect to the estimated mean value:

$$RE_i = \frac{\sigma_{\hat{\theta}_i}}{\hat{\theta}_i} \quad (13)$$

2.6 Initial conditions

The initial conditions for the dynamic model are given in Table 2.

2.7 Assessment of the model fit

The goodness of fit for the model prediction in the model validation procedure and on-line monitoring application was assessed with the root mean sum of squared errors (RMSSE):

$$RMSSE = \sqrt{\frac{1}{n} \sum_i^n (y_{meas,i} - \hat{y}_i)^2} \quad (14)$$

3. Framework for the soft sensor

3.1 Design of the probabilistic model-based soft sensor

The objective of the probabilistic model-based soft sensor is to monitor the *S. thermophilus* fermentation. It predicts the probabilistic distribution of the states, such as biomass, lactose, lactic acid, and pH, in real time based on the on-line available ammonia addition and the pH measurements. The soft sensor consists of a data reconciliation module, a parameter update module, and a dynamic process model (Fig. 1). The model parameters μ_{max} and t_{lag} are updated every 5 minutes based on the latest on-line measurements, and the soft sensor predicts both the current value and the future course of the state variables. The current states are saved as initial conditions for the next interval. Monte Carlo simulations of the dynamic model are performed every interval using samples from the input uncertainty domain. To this end, the Latin hypercube sampling technique was used to generate 100 random samples from the input uncertainty domain in which we included uncertainties in the initial conditions, model parameters, and ammonia addition (Table 2). The outcome from the Monte Carlo simulations was a probability distribution of the state variables.

3.1.1 Data reconciliation method

The data reconciliation module is based on elemental and bio-energetic conservation principles such as the charge balance (Eq. 15) and the lactic acid production rate expression (Eq. 16). It uses the amount of added ammonia, where one measurement value is available every minute, to update the volumetric biomass growth and lactic acid production rate. The

ammonia addition data points of each interval are fitted with a smoothing spline line and the average ammonia addition of the interval, $q_{NH,add}$, is estimated. Missing measurement points can also be handled due to the implementation of the fit. With the data reconciliation the growth rate (q_X) is obtained and used as input for the parameter update module.

$$NH_4^+ + C_3H_5O_3^- = q_{NH,add} + q_P = 0 \quad (15)$$

$$q_P = \alpha \cdot q_X \quad (16)$$

3.1.2 Parameter update

The updated q_X is used to update μ_{max} in every interval. μ_{max} is updated in an iterative procedure until the change of μ_{max} is less than 5 % compared to the previous iteration. In the first iteration ($k=1$) μ_{max} is calculated based on the updated q_X , the function values, and biomass concentration of the previous interval (Eq. 17). The subsequent iterations use the function values and biomass concentration based on the new μ_{max} value. The function values and biomass concentration derive from an evaluation of the dynamic model.

$$\mu_{max,k} = \frac{q_{X,updated}}{f_{lag,k-1} \cdot f_{S,k-1} \cdot f_{P,k-1} \cdot f_{pH,k-1} \cdot X_{k-1}} \quad (17)$$

It is not possible to use the updated rates (q_P and q_X) directly in the dynamic model, as they resemble only the conditions of the previous 5 minutes. However, inhibition and limitation effects, as well as pH variations, which will occur during a fermentation, influence the rates. It is therefore necessary to calculate the rates within the dynamic model according to Eq. 4 and Eq. 6 in order to predict the future course of the fermentation, as well.

The lag-time parameter t_{lag} is updated based on the measured pH value as soon as the measured pH reaches the controlling value (here $pH = 6$). The continuous pH measurement is saved every minute. t_{lag} is adjusted so that the modelled and measured pH reach the control value at the same time (Eq. 18). t_{lag} is updated in an iterative procedure until the change is less than 2 % compared to the previous iteration.

$$t_{lag,k} = t_{lag,k-1} + (t_{pH=6,measured} - t_{pH=6,predicted}) \quad (18)$$

Once t_{lag} is updated, the model is run from the beginning, because t_{lag} influences the whole prediction range. The current state is then saved as initial conditions for the next iteration.

The updated parameters μ_{max} and t_{lag} are used as input for the dynamic model.

3.1.3 Dynamic mechanistic process model

The dynamic process model comprises the biological model and the mixed weak acid/base model as described in the Materials and Methods section.

3.1.4 Monte Carlo simulations

The Monte Carlo method includes three main steps namely (1) identification and definition of input uncertainties, (2) sampling and (3) Monte Carlo simulation. For step 1, uncertainties in the biological model parameters, initial lactose concentration, initial biomass concentration, and the ammonia addition are considered (Table 2). The uncertainties of the model parameters are represented by the covariance matrix (includes the standard deviation and correlation matrix), which is derived from the parameter estimation. Uncertainties in the initial lactose and biomass concentration, and the measured ammonia addition are considered to be normally distributed with $3\sigma = 10\%$. The model parameters are assumed to be normally distributed as well because the measurement errors, on which the parameter estimation is founded, are assumed to be normally distributed. In order to account for the ammonia addition uncertainty, samples with a normal distribution, a mean value 1, and $3\sigma = 0.1$ are generated and will be multiplied with the measured ammonia addition rate in the Monte Carlo simulations. Uncertainties in the parameters of the mixed weak acid/base system are not considered since pH predictions were not necessary for the online monitoring application, as pH was directly measured and used as input for the data reconciliation module. The identification of uncertain input sources and the definition of the uncertainty ranges depend on the system studied. In general, this should therefore be systematically evaluated for each

studied system separately. In this study, uncertainties of 5 % were expected based on an expert discussion and considering the available data. To be on the safe side, we considered $3\sigma = 10\%$ for the uncertainties in the initial conditions and the ammonia measurement. For step 2, the Latin Hypercube Sampling (LHS) technique is used together with the Iman Conover rank correlation method to induce the correlation matrix in the input domain (see step 1) [35,36]. The sampling procedure features the following generic steps: First, LHS sampling for independent inputs is performed in the unit probability domain [0 1] for N sampling numbers (N =100 used in this study). Then, the correlation matrix is induced via the Iman Conover method [37] for the correlated parameters. Finally, the correlated parameter samples are inverted from the probability domain to real values considering the inverse cumulative distribution function for each input e.g. using the Matlab function `icdf`. In this study, we assumed both measurement errors as well as parameter estimation errors to be normally distributed hence we set the option “Probability distribution name” to “Normal” in the `icdf` function. In this step, the user can define any other distribution function deemed appropriate as well (e.g. uniform, gamma etc.).

In step 3, Monte Carlo simulations of the dynamic model are performed for each sample. The output of the Monte Carlo simulations consists of 100 model predictions, representing a probabilistic distribution of the predicted state variables.

4. Results and Discussion

4.1 Parameter estimation

A parameter estimation of the biological model was performed in order to assess the model reliability. Uncertainty and sensitivity analysis were conducted to find an identifiable parameter subset. It must hereby be considered that the estimated parameters depend among others on the nominal parameter values, the cultivation conditions, and the model structure

[34]. The first parameter estimation, fitting all biokinetic model parameters, revealed identifiability issues (Supplementary Table S2). $K_{P,pH1}$, $K_{P,pH2}$, K_S , and K_I could not be estimated and were therefore maintained at their initial values for the subsequent steps. The subsequent parameter estimation with the remaining 9 parameters revealed an identifiable parameter subset. The estimated parameter values were in the expected range and in the order of magnitude as known from previous studies (Table 3). It has to be noted that $\mu_{\max} = 2.06 \text{ h}^{-1}$ was higher than the actual biological value because it had to compensate for the functions in the growth rate expression. The relative errors of all parameters were lower than 10 %. In addition, all parameters had a significant effect on the model outputs (Supplementary Table S3). Some of the parameters met the criterion of a correlation coefficient smaller than 0.5 to be uniquely identifiable. However, this parameter subset should be considered as a whole due to the linear dependency between most of the parameters. The estimated value of one parameter is conditional on the value of another parameter. Therefore, the covariance matrix of the parameters should be used, e.g. when performing parameter sampling in Monte Carlo simulations, when performing model simulations, as done in this work.

The model showed an acceptable fit of the fermentation data (Fig. 2, Supplementary Fig. S1 – 4). To measure process performance, the focus was on the biomass concentration because the cells were the target fermentation product. The RMSSE for biomass was around 0.4 g L^{-1} for many of the fermentations, corresponding to a discrepancy of less than 10 %, giving evidence of a good fit (Table 4). A good model fit was achieved for the fermentations at pH 5.5, 6.0, and 6.5 but not at pH = 7.0, which had an error of 30 %. Furthermore, the secretion of galactose was underestimated in all fermentations. This could be attributed to an inconsistent carbon balance in the experimental fermentation data, where more carbon was produced than lactose consumed. The supplemented yeast extract, which was not taken into

account in the model, does contain approximately 6 g L⁻¹ carbon [19] when assuming the elemental composition of a *S. cerevisiae* cell for the yeast extract. Hence, amino acids that derived from the yeast extract and were taken up by the cells might have led to the inconsistency in the carbon balance. The parameter estimation aimed therefore not to fit the galactose concentration, but it was anyhow kept in case the model will be extended in future studies.

The evaluation of the pH function f_{pH} showed a clear maximum at pH = 6.4 (Fig. 3A).

Furthermore, growth was already reduced by 25 % at pH = 5.5 and 7. Similar trends of the influence of the pH on the growth of LAB were observed in other studies [21,38,39]. These studies found slightly different pH optimums in the range between 6 and 7 since different strains were studied. The growth inhibition by lactate was pH dependent, as well (Fig. 3B). 20 g L⁻¹ lactate inhibited growth by 50 % in the pH range from 5.5 to 6.5, whereas at pH = 7 already 10 g L⁻¹ lactate inhibited growth by 50 %. pH dependent inhibition of growth caused by lactate was also already observed for the lactic acid producing bacterium *Enterococcus faecalis* [40]. This underlines the necessity of the pH dependent lactate inhibition parameter K_{La1} (Eq. 5).

4.2 Model validation

Following the statistical assessment of the quality of the parameter estimates above, the model was validated against two independent fermentation data sets, which were performed at pH = 6.0 and an initial lactose concentration of 20 g L⁻¹ (Fig. 4, Supplementary Fig. S5). The model predicted the measured lactose, biomass, lactic acid, and galactose concentrations (Fig. 4 A-D). The lag-time parameter t_{lag} was fitted for both fermentations because the lag time differs from batch to batch. The assessment of the validation model fit showed that the model gave an acceptable prediction accuracy with an RMSSE for biomass of 0.2 g L⁻¹

(Table 4). The pH prediction was also very accurate with a discrepancy of less than 0.1 pH units (Fig. 4 E). In the beginning of the fermentation, the pH dropped from 6.1 to 6.0 before the controller started to add ammonia solution. This drop was predicted to be faster than actually measured, which could be attributed to a slightly different buffer capacity of the medium in reality compared to the mixed weak acid/base model. However, a prediction accuracy within ± 0.1 pH units was deemed sufficiently accurate for monitoring purposes, as pH measurement errors were expected to be in the same range. The only exception of an accurate pH fit was at the moment when the pH controller started: too much ammonia was added in the experiment so that the pH showed an overshoot. The pH prediction is closely dependent on the predictions of the ammonia addition and lactic acid concentration. The validity of the mixed weak acid/base model was therefore demonstrated by a correct prediction of the added ammonia solution (Fig. 4 F), as the pH is held constant by adding ammonia solution. Nevertheless, the validity of the applied Davis equation to calculate the activity coefficients (Eq. 9) for $I \leq 0.5 \text{ mol L}^{-1}$ has to be noted, and could be improved in future studies in particular for fermentations with an ionic strength higher than 1 mol L^{-1} . Overall, these results indicate the validity of the model, which encourages its further application for monitoring of a fermentation process as presented below.

4.3 Application of the probabilistic model-based soft sensor

The probabilistic model-based soft sensor was applied to the data sets of three historical fermentations, where the historical on-line data was used as it would be available on-line. Here, the initial pH was around 7 and the pH dropped to the control value 6 due to acid secretion as by-product during the LAB fermentation (Fig. 5, Supplementary Fig. S6 and S7, while the Supplementary Movies show the virtual on-line implementation of the soft sensor). The on-line data, namely pH and quantity of added ammonia were used as an input to update the monitoring system (Fig. 5 left column). The Monte Carlo propagation of the error for the

biomass, lactose, and lactic acid concentration is then predicted by the monitoring system (Fig. 5 middle and right column). The predictions of the future states of the system are shown at different times, 2 h, 2 h 40 min, 3 h, 4 h, and 6 h (Fig. 5 rows). Since the pH was higher than the control value 6, no ammonia solution was added within the first 2 h and 35 min. Therefore, no data reconciliation and parameter update were conducted (Fig. 5, time = 2 h), and Monte Carlo simulations were performed in the defined input uncertainty space (Table 2 and Supplementary Fig. S8) considering uncertainties in the biological model parameters and initial conditions. The monitoring system was running without updating the parameters until the ammonia addition started to control the pH (Fig. 5, at time = 2 h 40 min). At this point, t_{lag} was updated ensuring that the pH controller in the experiment and simulation started at the same time. It is clear that there is a lag-time variation from batch to batch, which has to be taken into account. On the one hand, a dependency on the pH measurement is introduced by this procedure. On the other hand, it is the only possibility – given the limited amount of on-line measurements – to align the modelled and measured ammonia addition, which is crucial for the monitoring system. Once t_{lag} was updated, and the ammonia addition started, the data reconciliation and parameter update modules updated μ_{max} every 5 minutes, as described in the Framework description. With time, more measurement information was available and the prediction accuracy of the state variables increased (Fig.5, time = 3 h – 6 h).

The RMSSE for biomass was 0.8 g L^{-1} when the fermentation started, and improved to 0.5 g L^{-1} at the end of the fermentation (both with a standard deviation of 0.1 g L^{-1}) (Fig. 6). Mainly, the update of t_{lag} after 2 h and 40 minutes of the fermentation improved the prediction accuracy. The reproducibility of the Monte Carlo simulations was validated as the RMSSE for biomass varied less than 0.5 % in 10 repeated Monte Carlo simulations with 100 input samples in each simulation. Changing the tolerance limit to estimate μ_{max} in the

iterative update procedure (Eq. 17) to 1% and 0.1% did not improve the prediction accuracy for the presented fermentations (data not shown). However, this might be necessary for other applications. In summary, an accurate prediction of the state variables was achieved.

Several reports have implemented soft sensors to monitor LAB fermentations. Acuña et al. [16] and Peter and Röck [15] implemented a model-based monitoring system for LAB fermentations using the base addition and pH measurements, whereas the second implementation is limited to monitoring the lactic acid concentration. Fayolle et al. [17] and Payot et al. [18] designed a data-based soft sensor using mid-infrared spectroscopy and conductivity, respectively. However, all studies presented deterministic predictions and did not consider the imperfect knowledge of the process by taking uncertainties into account.

Contrary to the earlier published investigations, this study accounted for several sources of uncertainties in the probabilistic monitoring system and assessed the combined effect of system uncertainties on the predictions. The initial conditions, on-line measurements, and biological model parameters were considered as uncertainty sources. The concentrations of the medium components (initial conditions) vary from batch to batch as the medium preparation procedure underlies uncertainties. The biomass concentration depends on the cryo-stock and pre-culture quality. Since the monitoring system relies on the ammonia addition measurement, it is important to incorporate measurement uncertainties, as well. Mears et al. [12] pointed out that an error of 5 % of the carbon evolution rate or oxygen uptake rate, caused by measurement errors in the exhaust gas, led to errors of more than 50 % in the model prediction of the final biomass concentration in a filamentous fungi process at pilot scale. The exact extent of uncertainties of the initial conditions and measurements could not be determined in the present study because statistically relevant data was not available.

The implemented uncertainties were instead based on expert knowledge. However, the model parameter uncertainties were obtained in the parameter regression step that has been presented above.

The monitoring system predicted the current state variables and forecasted the future course of the fermentation and could therefore support a lean production. If this monitoring system is applied at production sites, it will provide plant operators with a PAT tool to monitor the biological variables in the fermentation process, such as biomass concentration, instead of on-line ammonia addition measurements, where the latter are difficult to comprehend (as shown in the virtual implementation in the Supplementary Movies). In addition, the tool could predict whether and when the target cell yield will be reached. This helps run the batch period optimally and schedule other unit operations: All downstream processing steps could be coordinated with the upstream fermentation batch time and therefore be prepared in time. Moreover, cleaning, sterilization, media preparation, and pre-culture steps affiliated with the start up of the batch process could be optimized to reduce the overall downtime of the fermentation unit. An optimized schedule with efficient utilization of the different operation units can contribute to a more economical operation of the production plant. The monitoring system could also cover the early diagnosis of process failures and warn the operators if biomass growth had stopped unexpectedly. The standstill of ammonia addition is a sign of interrupted biomass growth, as it happened in the shown fermentation after 3 h (Fig. 5, time = 3 h). However, a warning should only be given in case the ammonia addition stopped for a longer period of time – in contrast to the present fermentation, where the ammonia addition stopped only for a short while because the pH controller overshoot. Furthermore, the system could be extended to calculate the risk of not achieving the target biomass yield as a result of the outcome of the soft sensor, which provides the probability distribution of biomass

concentration at the end of a given batch. It could then be implemented for risk-based monitoring and be further developed for control purposes.

5. Conclusion

A probabilistic model-based soft sensor was proposed for the monitoring of *S. thermophilus* fermentations. State variables, such as biomass, substrate, and lactic acid, which were not possible to be measured on-line, could be successfully predicted. The predictions were based on very limited available on-line measurements, namely base addition and pH measurements since exhaust gas measurements were not available. The aim was achieved by coupling a biokinetic model and a mixed weak acid/base model (for the pH calculation), which were validated comprehensively. Uncertainties in the initial substrate concentration, base addition, and biological model parameters were quantified and accounted for using Monte Carlo simulations in the probabilistic monitoring system. The future objective of this study will be to implement the monitoring system for on-line risk-based monitoring and control in pilot- and large-scale LAB studies.

Competing interest

The authors declare that they have no competing interests.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643056 (Biorapid project). We are thankful for the cooperation with Chr. Hansen A/S, and the support from Klaus Pellicer Alborch (Technische Universität Berlin) during the experiments.

Nomenclature

C_{Gal}	galactose concentration (g L^{-1})
C_{Glc}	glucose concentration (g L^{-1})
C_{H^+}	H^+ concentration (mol L^{-1})
C_{LA}	lactate concentration (g L^{-1})
C_{OH^-}	OH^- concentration (mol L^{-1})
C_P	total lactic acid (lactate and lactic acid) concentration (g L^{-1})
C_S	lactose (substrate) concentration (g L^{-1})
C_{iCO}	total carbonic acid ($H_2CO_3^*$ and HCO_3^-) concentration (mol L^{-1})
C_{iNH}	total concentration of NH_4^+ and NH_3 (g L^{-1})
C_{iPh}	total concentration of H_3PO_4 , $H_2PO_4^-$, and HPO_4^{2-} (g L^{-1})
C_{iZ}	total concentration of the unknown compound (dissociated and undissociated form) (mol L^{-1})
C_X	biomass concentration (g L^{-1})
Error _i	Weighted model prediction error of a component at time point i
f_d	divalent activity coefficients (-)
f_{lag}	lag-time function (-)
f_m	monovalent activity coefficients (-)
f_p	lactic acid inhibition function (-)
f_{pH}	pH dependency function (-)
f_s	substrate limitation and inhibition function (-)
$H_2CO_3^*$	dissolved CO_2 and H_2CO_3
I	ionic strength (g L^{-1})
K'_{C1}	apparent equilibrium constant for the carbonic acid system (-)
K_I	substrate inhibition parameter (g L^{-1})
K_{La}	lactate inhibition parameter (g L^{-1})
K_{La1}	pH dependent lactate inhibition parameter (g L^{-1})
K'_{LA}	apparent equilibrium constant for the lactic acid system (-)
K'_{NH}	apparent equilibrium constant for the ammonia system (-)
K_P	P-controller controller gain
$K_{P,La}$	2. lactate inhibition parameter (L g^{-1})
$K_{P,pH1}$	lactate inhibition pH parameter (-)
$K_{P,pH2}$	2. lactate inhibition pH parameter (-)
K'_{P1}	apparent equilibrium constant for the phosphoric acid system (-)
K'_{P2}	apparent equilibrium constant for the dihydrogen phosphate system (-)
$K'_{r,C1}$	apparent reverse rate constant for carbonic acid dissociation (s^{-1})
$K'_{r,LA}$	apparent reverse rate constant for lactic acid dissociation (s^{-1})
$K'_{r,NH}$	apparent reverse rate constant for NH_4 dissociation (s^{-1})
$K'_{r,P1}$	apparent reverse rate constant for H_3PO_4 dissociation (s^{-1})
$K'_{r,P2}$	apparent reverse rate constant for $H_2PO_4^-$ dissociation (s^{-1})
$K'_{r,W}$	apparent reverse rate constant for water dissociation (s^{-1})
K_S	substrate limitation parameter (g L^{-1})
K'_W	apparent equilibrium constant for the water system (-)
K'_Z	apparent equilibrium constant for the unspecified compound system (-)
n	number of measurement points
pH_{opt}	optimal pH parameter in the pH function (-)
pH_{opt_lag}	optimal pH for the lag-time fit (-)
pH_{set}	pH control set point (-)
pK_{C1}	pK_a constant for carbonic acid dissociation
pK_{LA}	pK_a constant for lactic acid dissociation
pK_{NH}	pK_a constant for NH_4 dissociation
pK_{P1}	pK_a constant for H_3PO_4 dissociation
pK_{P2}	pK_a constant for $H_2PO_4^-$ dissociation
pK_W	pK_a constant for water dissociation
pK_Z	pK_a constant for the unspecified compound dissociation
q_{Gal}	volumetric galactose secretion rate ($\text{C-mol L}^{-1} \text{h}^{-1}$)
q_{NH}	volumetric ammonia consumption rate ($\text{mol L}^{-1} \text{h}^{-1}$)
$q_{NH,add}$	volumetric ammonia addition rate ($\text{mol L}^{-1} \text{h}^{-1}$)

q_p	volumetric lactic acid secretion rate (C-mol L ⁻¹ h ⁻¹)
q_{Ph}	volumetric phosphoric acid consumption rate (mol L ⁻¹ h ⁻¹)
q_s	volumetric substrate consumption rate (C-mol L ⁻¹ h ⁻¹)
q_x	volumetric biomass growth rate (C-mol L ⁻¹ h ⁻¹)
RE	relative error (-)
RMSSE	root mean sum of squared errors (g L ⁻¹)
T	temperature in the bioreactor (K)
t	time variable (h)
t_{lag}	lag-time coefficient (h)
w_j	maximum value of the state j for the weighted error in the objective function
Y_{gal}	galactose yield (g g ⁻¹)
z_i	charge number of the i-th ion
\hat{y}_i	i-th model value of one output (g L ⁻¹)
$y_{meas,i}$	i-th measurement value of one output (g L ⁻¹)

Greek Letters

α	growth related production coefficient of lactic acid (g g ⁻¹)
$\hat{\theta}_i$	estimated parameter value
μ_{max}	maximum specific growth rate (h ⁻¹)
σ	standard deviation
σ_{pH}	spread parameter is the gaussian pH function
σ_{lag}	standard deviation of the lag-time fit
$\sigma_{\hat{\theta}_i}$	standard deviation of the estimated parameter

References

- [1] A. Von Wright, L. Axelsson, Lactic Acid Bacteria: An Introduction, in: S. Lahtinen, A.C. Ouwehand, S. Salminen, A. VonWright (Eds.), *Lact. Acid Bact. Microbiol. Funct. Asp.*, 4th ed., CRC Press, Taylor & Francis, Boca Raton, 2012: pp. 1–16.
- [2] S. Patel, A. Majumder, A. Goyal, Potentials of Exopolysaccharides from Lactic Acid Bacteria, *Indian J. Microbiol.* 52 (2012) 3–12.
- [3] B. Sonnleitner, Automated Measurement and Monitoring of Bioprocesses: Key Elements of the M3C Strategy, in: *Adv. Biochem. Eng. Biotechnol.*, 2012: pp. 1–33.
- [4] FDA, Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, U.S. Food and Drug Administration, U.S. Department of Health and Human Services, 2004.
- [5] R. Luttmann, D.G. Bracewell, G. Cornelissen, K. V. Gernaey, J. Glassey, V.C. Hass, C. Kaiser, C. Preusse, G. Striedner, C.F. Mandenius, Soft sensors in bioprocessing: A status report and recommendations, *Biotechnol. J.* 7 (2012) 1040–1048.
- [6] K. V Gernaey, A.E. Lantz, P. Tufvesson, J.M. Woodley, G. Sin, Application of mechanistic models to fermentation and biocatalysis for next-generation processes., *Trends Biotechnol.* 28 (2010) 346–54.
- [7] L. Mears, S.M. Stocks, M.O. Albaek, G. Sin, K. V. Gernaey, Mechanistic Fermentation Models for Process Design, Monitoring, and Control, *Trends Biotechnol.* 35 (2017) 914–924.
- [8] P. Kadlec, B. Gabrys, S. Strandt, Data-driven Soft Sensors in the process industry, *Comput. Chem. Eng.* 33 (2009) 795–814.
- [9] J. Alves-Rausch, R. Bienert, C. Grimm, D. Bergmaier, Real time in-line monitoring of large scale Bacillus fermentations with near-infrared spectroscopy, *J. Biotechnol.* 189 (2014) 120–128.
- [10] P. Ödman, C.L. Johansen, L. Olsson, K. V. Gernaey, A.E. Lantz, On-line estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using in-situ multi-wavelength fluorescence and software sensors, *J. Biotechnol.* 144 (2009) 102–112.
- [11] A. Golabgir, C. Herwig, Combining Mechanistic Modeling and Raman Spectroscopy for Real-Time Monitoring of Fed-Batch Penicillin Production, *Chemie-Ingenieur-Technik.* 88 (2016) 764–776.
- [12] L. Mears, S.M. Stocks, M.O. Albaek, G. Sin, K. V. Gernaey, Application of a mechanistic model as a tool for on-line monitoring of pilot scale filamentous fungal fermentation processes—The importance of evaporation effects, *Biotechnol. Bioeng.* 114 (2017) 589–599.
- [13] R. Lencastre Fernandes, V.K. Bodla, M. Carlquist, A.-L. Heins, A. Eliasson Lantz, G. Sin, K. V. Gernaey, Applying Mechanistic Models in Bioprocess Development, in: *Adv. Biochem. Eng. Biotechnol.*, 2013: pp. 137–166.
- [14] G. Sin, K. V Gernaey, A.E. Lantz, Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis., *Biotechnol. Prog.* 25 (2009) 1043–1053.
- [15] E. Peter, H. Röck, On-line estimation of lactic acid concentration during batch fermentations of *Streptococcus thermophilus* based on pH measurement, *12th Int. Conf. Control. Autom. Syst.* (2012) 851–855.
- [16] G. Acuña, E. Latrille, C. Béal, G. Corrieu, A. Chéry, On-line estimation of biological variables during pH controlled lactic acid fermentations, *Biotechnol. Bioeng.* 44 (1994) 1168–1176.
- [17] P. Fayolle, D. Picque, G. Corrieu, Monitoring of fermentation processes producing

- lactic acid bacteria by mid-infrared spectroscopy, *Vib. Spectrosc.* 14 (1997) 247–252.
- [18] T. Payot, M. Fick, G. Chimique, On-line estimation of lactic acid concentration by conductivity measurement in fermentation broth, *Biotechnol. Tech.* 11 (1997) 17–20.
- [19] J. Villadsen, J. Nielsen, G. Lidén, *Bioreaction Engineering Principles*, Springer US, Boston, MA, 2011.
- [20] C. Åkerberg, K. Hofvendahl, B. Hahn-Hägerdal, G. Zacchi, Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by *Lactococcus lactis* ssp. *lactis* ATCC 19435 in whole-wheat flour, *Appl. Microbiol. Biotechnol.* 49 (1998) 682–690.
- [21] M. Aghababaie, M. Khanahmadi, M. Beheshti, Developing a detailed kinetic model for the production of yogurt starter bacteria in single strain cultures, *Food Bioprod. Process.* 94 (2015) 657–667.
- [22] A.W. Schepers, J. Thibault, C. Lacroix, *Lactobacillus helveticus* growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation, *Enzyme Microb. Technol.* 30 (2002) 187–194.
- [23] A. Amrane, Y. Prigent, Influence of an initial addition of lactic acid on growth, acid production and their coupling for batch cultures of *Lactobacillus helveticus*., *Bioprocess Eng.* 19 (1998) 307–312.
- [24] A.P. Oliveira, J. Nielsen, J. Förster, Modeling *Lactococcus lactis* using a genome-scale flux model., *BMC Microbiol.* 5 (2005) 39.
- [25] R. Luedeking, E.L. Piret, A Kinetic Study of the Lactic Acid Fermentation. Batch Process at Controlled pH, *J. Biochem. Microbiol. Technol. Eng.* 1 (1959) 393–412.
- [26] R.Y. Peng, T.C.K. Yang, H. Wang, Y. Lin, C. Cheng, Modelling of Lactic Acid Fermentation - An Improvement of Leudeking's Model, *J. Chinese Agric. Chem. Soc.* 35 (1997) 485–494.
- [27] E.V. Musvoto, M.C. Wentzel, R.E. Loewenthal, G.A. Ekama, Integrated Chemical-Physical Processes Modelling - I. Development of a Kinetic-Based Model for Mixed Weak Acid/Base Systems, *Water Res.* 34 (2000) 1857–1867.
- [28] R.M.C. Dawson, *Data for Biochemical Research*, Clarendon Press, Oxford, 1969.
- [29] R.E. Loewenthal, G.A. Ekama, G.R. Marais, Mixed weak acid/base systems Part I - Mixture characterisation., *Water SA.* 15 (1989) 3–24.
- [30] C.W. Davies, *Ion Association*, Butterworth, Londond, 1962.
- [31] G.A.F. Seber, C.J. Wild, *Nonlinear Regression*, (1989) 792.
- [32] G. Sin, K. V. Gernaey, *Data Handling and Parameter Estimation*, in: M.C.M. van Loosdrecht, P.H. Nielsen, C.M. Lopez-Vazquez, D. Brdjanovic (Eds.), *Exp. Methods Wastewater Treat.*, IWA Publishing Company, 2016: pp. 201–234.
- [33] R. Spann, C. Roca, D. Kold, A. Eliasson Lantz, K. V. Gernaey, G. Sin, A Consistent Methodology Based Parameter Estimation for a Lactic Acid Bacteria Fermentation Model, in: A. Espuña, M. Graelss, L. Puigjaner (Eds.), *Proc. 27th Eur. Symp. Comput. Aided Process Eng. – ESCAPE 27*, Elsevier, 2017: p. 3042.
- [34] G. Sin, A.S. Meyer, K. V. Gernaey, Assessing reliability of cellulose hydrolysis models to support biofuel process design—Identifiability and uncertainty analysis, *Comput. Chem. Eng.* 34 (2010) 1385–1392.
- [35] M.D. McKay, R.J. Beckman, W.J. Conover, Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics.* 21 (1979) 239–245.
- [36] G. Sin, K. V. Gernaey, M.B. Neumann, M.C.M. van Loosdrecht, W. Gujer, Uncertainty analysis in WWTP model applications: A critical discussion using an example from design, *Water Res.* 43 (2009) 2894–2906.

- [37] R.L. Iman, W.J. Conover, A distribution-free approach to inducing rank correlation among input variables, *Commun. Stat. - Simul. Comput.* 11 (1982) 311–334.
- [38] M.-P.P. Zacharof, R.W. Lovitt, Modelling and simulation of cell growth dynamics, substrate consumption, and lactic acid production kinetics of *Lactococcus lactis*, *Biotechnol. Bioprocess Eng.* 18 (2013) 52–64.
- [39] H. Ohara, K. Hiyama, T. Yoshida, Kinetic study on pH dependence of growth and death of *Streptococcus faecalis*, *Appl. Microbiol. Biotechnol.* 38 (1992) 403–407.
- [40] A.D. Nandasana, S. Kumar, Kinetic modeling of lactic acid production from molasses using *Enterococcus faecalis* RKY1, *Biochem. Eng. J.* 38 (2008) 277–284.

ACCEPTED MANUSCRIPT

Table Captions

Table 1. Kinetics for the mixed weak acid/base model. f_m and f_d are mono- and divalent activity coefficients, respectively.

Table 2. Parameters of the dynamic model of the *S. thermophilus* fermentations.

Table 3. Estimated model parameters including the relative error and correlation matrix.

Table 4. Fermentation conditions and RMSSE of the biomass prediction for all used data sets: parameter estimation, validation, and the monitoring system.

Figure Captions

Fig. 1. Block diagram of the probabilistic model-base monitoring system.

Fig. 2. Model predictions for a *S. thermophilus* lab-scale batch fermentation.

Fig. 3. Growth affecting functions of pH and lactate inhibition.

Fig. 4. Model prediction for a validation lab-scale batch fermentation.

Fig. 5. Probabilistic monitoring system applied to lab-scale batch data of a *S. thermophilus* fermentation.

Fig. 6. 95% confidence interval of the RMSSE for the biomass prediction during the probabilistic monitoring of a *S. thermophilus* fermentation.

Table 1. Kinetics for the mixed weak acid/base model. f_m and f_d are mono- and divalent activity coefficients, respectively; see Loewenthal et al. (1989) and Musvoto et al. (2000).

Reaction	reaction rate vector	apparent equilibrium constant	
		symbol	value
$NH_4^+ \leftrightarrow NH_3 + H^+$	$K'_{r,NH} \cdot K'_{NH} \cdot [NH_4^+] - K'_{r,NH} \cdot [NH_3] \cdot [H^+]$	K'_{NH}	$10^{-pK_{NH}}$
$H_3PO_4 \leftrightarrow H_2PO_4^- + H^+$	$K'_{r,P1} \cdot K'_{P1} \cdot [H_3PO_4] - K'_{r,P1} \cdot [H_2PO_4^-] \cdot [H^+]$	K'_{P1}	$10^{-pK_{P1}} / f_m^2$
$H_2PO_4^- \leftrightarrow HPO_4^{2-} + H^+$	$K'_{r,P2} \cdot K'_{P2} \cdot [H_2PO_4^-] - K'_{r,P2} \cdot [HPO_4^{2-}] \cdot [H^+]$	K'_{P2}	$10^{-pK_{P2}} / f_d$
$H_2CO_3^* \leftrightarrow HCO_3^- + H^+$	$K'_{r,C1} \cdot K'_{C1} \cdot [H_2CO_3^*] - K'_{r,C1} \cdot [HCO_3^-] \cdot [H^+]$	K'_{C1}	$10^{-pK_{C1}} / f_m^2$
$C_3H_6O_3 \leftrightarrow C_3H_5O_3^- + H^+$	$K'_{r,LA} \cdot K'_{LA} \cdot [C_3H_6O_3] - K'_{r,LA} \cdot [C_3H_5O_3^-] \cdot [H^+]$	K'_{LA}	$10^{-pK_{LA}} / f_m^2$
$H_2O \leftrightarrow OH^- + H^+$	$K'_{r,W} \cdot K'_W - K'_{r,W} \cdot [OH^-] \cdot [H^+]$	K'_W	$10^{-pK_W} / f_m^2$
$ZH^+ \leftrightarrow Z + H^+$	$K'_{r,Z} \cdot K'_Z \cdot [ZH^+] - K'_{r,Z} \cdot [Z] \cdot [H^+]$	K'_Z	$10^{-pK_Z} / f_m^2$

Table 2. Parameters of the dynamic model of the *S. thermophilus* fermentations.

Symbol	Value	Reference	Uncertainty classification
Biological model			
K_I	164 g L ⁻¹	[20]	
K_{La}	19.80 g L ⁻¹	parameter estimation	see Table 3
$K_{P,La}$	0.24 L g ⁻¹	parameter estimation	see Table 3
$K_{P,pH1}$	20	expert knowledge	
$K_{P,pH2}$	7	expert knowledge	
K_S	0.79 g L ⁻¹	[20]	
pH_{opt}	6.39	parameter estimation	see Table 3
t_{lag}	individual parameter estimation		
Y_{gal}	0.69 g g ⁻¹	parameter estimation	see Table 3
α	5.19 g g ⁻¹	parameter estimation	see Table 3
μ_{max}	2.06 h ⁻¹	parameter estimation	
σ_{pH}	1.42	parameter estimation	see Table 3
Mixed weak acid/base model			
$K'_{r,C1}$	10 ⁷ s ⁻¹	[27]	uncertainties in the mixed weak acid/base model are not considered because the pH is measured and used as input for the data reconciliation and parameter update module
$K'_{r,LA}$	10 ⁷ s ⁻¹	[27]	
$K'_{r,NH}$	10 ¹² s ⁻¹	[27]	
$K'_{r,P1}$	10 ⁸ s ⁻¹	[27]	
$K'_{r,P2}$	10 ¹² s ⁻¹	[27]	
$K'_{r,W}$	10 ¹⁰ s ⁻¹	[27]	
$K'_{r,Z}$	10 ⁷ s ⁻¹	[27]	
pK_{C1}	3404.7/($T - 14.8435 + 0.03279 \cdot T$)	[29]	
pK_{LA}	3.86	[28]	
pK_{NH}	2835.8/($T - 0.6322 + 0.00123 \cdot T$)	[29]	
pK_{P1}	799.3/($T - 4.5535 + 0.01349 \cdot T$)	[29]	
pK_{P2}	1979.5/($T - 5.3541 + 0.01984 \cdot T$)	[29]	
pK_W	14	[29]	
pK_Z	9.4	expert knowledge (amino acid mix)	
T	313.16 K	Measurement	
Initial Conditions			
$C_{Gal,t=0}$	0.0 g L ⁻¹		
$C_{Glc,t=0}$	0.0 g L ⁻¹		
$C_{H^+,t=0}$	dependent on the pH and ionic strength		
$C_{OH^-,t=0}$	dependent on the pH and ionic strength		
$C_{P,t=0}$	0.0 g L ⁻¹		
$C_{S,t=0}$	off-line measurements for the parameter estimation and validation		
$C_{S,t=0}$	65 g L ⁻¹ for the monitoring system		normal distribution $\sigma = 2.2$ g L ⁻¹
$C_{iCO,t=0}$	1.002 · 10 ⁻⁵ mol L ⁻¹		
$C_{iNH,t=0}$	off-line measurements for the parameter estimation and validation		
$C_{iNH,t=0}$	0.005 g L ⁻¹ for the monitoring system		
$C_{iPh,t=0}$	off-line measurements for the parameter estimation and validation		
$C_{iPh,t=0}$	2 g L ⁻¹ for the monitoring system		
$C_{iZ,t=0}$	2 mol L ⁻¹		
$C_{X,t=0}$	0.025 g L ⁻¹ for the parameter estimation and validation		
$C_{X,t=0}$	0.025 g L ⁻¹ for the monitoring system		normal distribution $\sigma = 0.0008$ g L ⁻¹

Table 3. Estimated model parameters including the relative error and correlation matrix.

kinetic parameters	estimated parameter value	relative error [%]	correlation matrix								
			μ_{\max}	K_P	K_{La}	pH_{opt}	σ_{pH}	α	$pH_{opt-lag}$	σ_{lag}	Y_{gal}
μ_{\max}	2.06	1	1	-0.74	-0.83	-0.17	0.5	-0.08	-0.52	0.53	0
$K_{P,La}$	0.24	13		1	0.77	-0.2	-0.58	-0.05	0.39	-0.28	0.04
K_{La}	19.80	0			1	-0.07	-0.54	0.31	0.44	-0.35	-0.28
pH_{opt}	6.39	1				1	-0.52	-0.13	0.76	-0.86	0.02
σ_{pH}	1.42	3					1	0.08	-0.93	0.85	0.06
α	5.19	0						1	-0.1	0.12	-0.4
$pH_{opt-lag}$	5.70	1							1	-0.97	-0.03
σ_{lag}	0.3	9								1	0.03
Y_{gal}	0.69	5									1

Table 4. Fermentation conditions and RMSSE of the biomass prediction for all used data sets: parameter estimation, validation, and the monitoring system.

batch data	pH	initial lactose conc. [g L ⁻¹]	RMSSE [g L ⁻¹]	final biomass [g L ⁻¹]*
parameter estimation	5.5	65	0.3	2.45 ± 0.1
	6	65	0.2	6.0 ± 0.2
	6	65	0.6	6.0 ± 0.1
	6.5	65	0.3	5.7 ± 0.1
	7	65	0.6	2.2 ± 0.2
validation	6	20	0.2	2.1 ± 0.2
	6	20	0.2	1.8 ± 0.1
monitoring system	6**	65	0.5 ± 0.1	5.9 ± 0.1
	6**	65	0.5 ± 0.1	5.8 ± 0.2
	6**	65	0.5 ± 0.1	5.9 ± 0.2

* with standard deviation of the last measurement at ca. 6 h fermentation time

** the initial pH was the pH of the medium (around pH = 7). The fermentation was controlled at pH = 6.

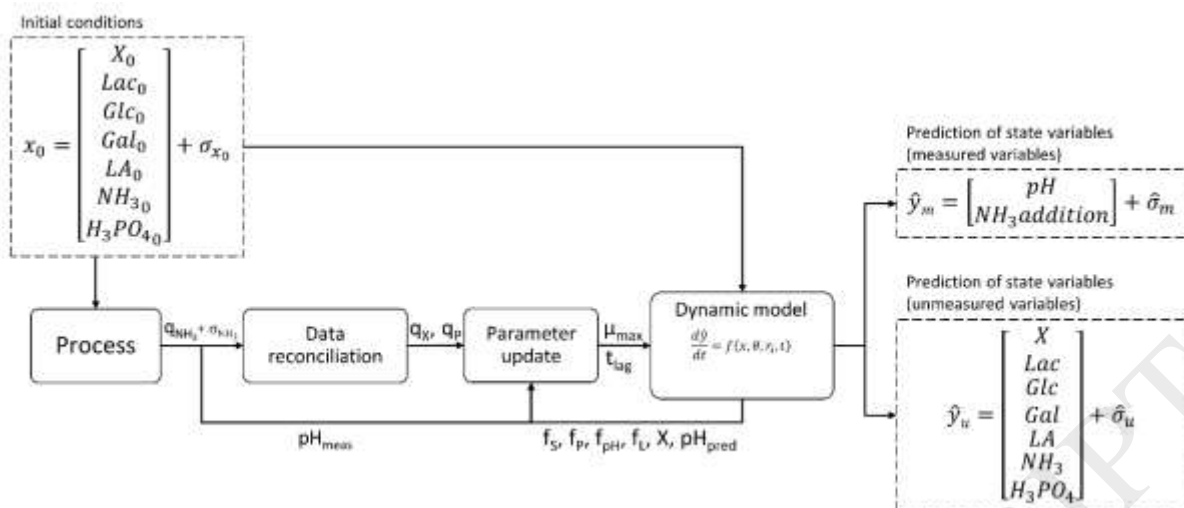


Fig. 1. Block diagram of the probabilistic model-based monitoring system. The initial conditions for the model were defined according to the process specifications including 10 % uncertainties in the lactose and biomass concentration. The on-line measured ammonia addition rate q_{NH_3} was used as input for the data reconciliation module to update the biomass growth and lactic acid production rate based on the charge balance and the lactic acid production rate expression. The parameter update module used the updated rates and the pH as input to update the model parameters μ_{max} and t_{lag} for the dynamic model. Monte Carlo simulations of the dynamic model were performed considering uncertainties in the initial lactose concentration, measured ammonia addition, and model parameters. The output of the dynamic model was a probability distribution of the state variables.

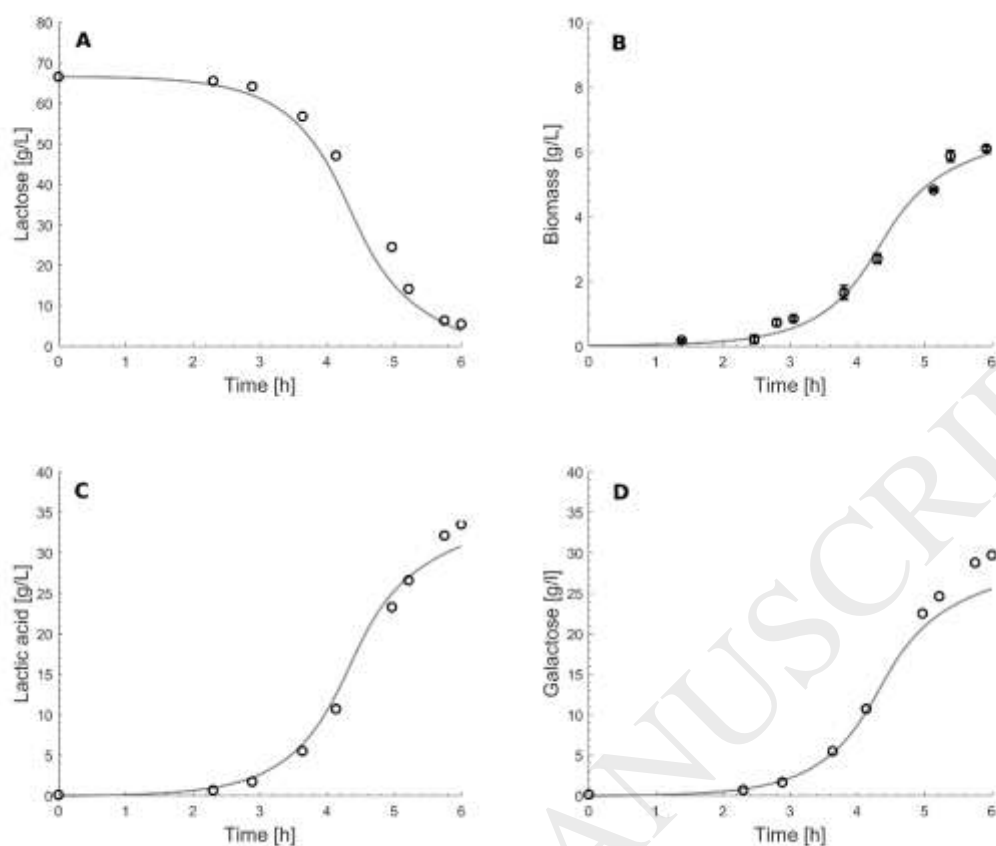


Fig. 2. Model predictions for a *S. thermophilus* lab-scale batch fermentation. . Lactose (A), biomass with standard deviation (B), lactic acid (C), and galactose (D) concentrations. The fermentation was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. The model prediction (solid line) for the measurements (circles) of one of the five lab-scale batches that were used for the parameter estimation is shown. The biomass measurement is shown with the standard deviation.

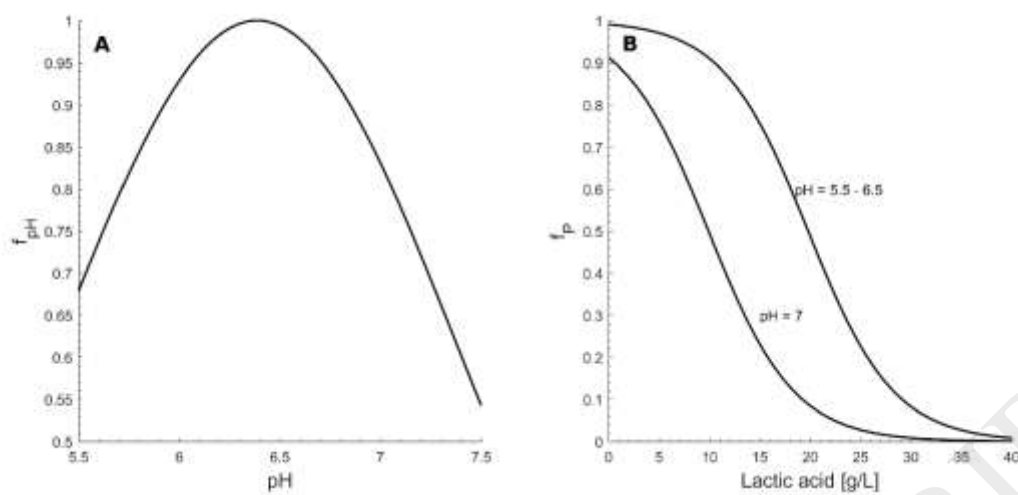


Fig. 3. Growth affecting functions of pH and lactate inhibition. pH function vs. pH (A) and lactate inhibition function vs. lactic acid concentration (B).

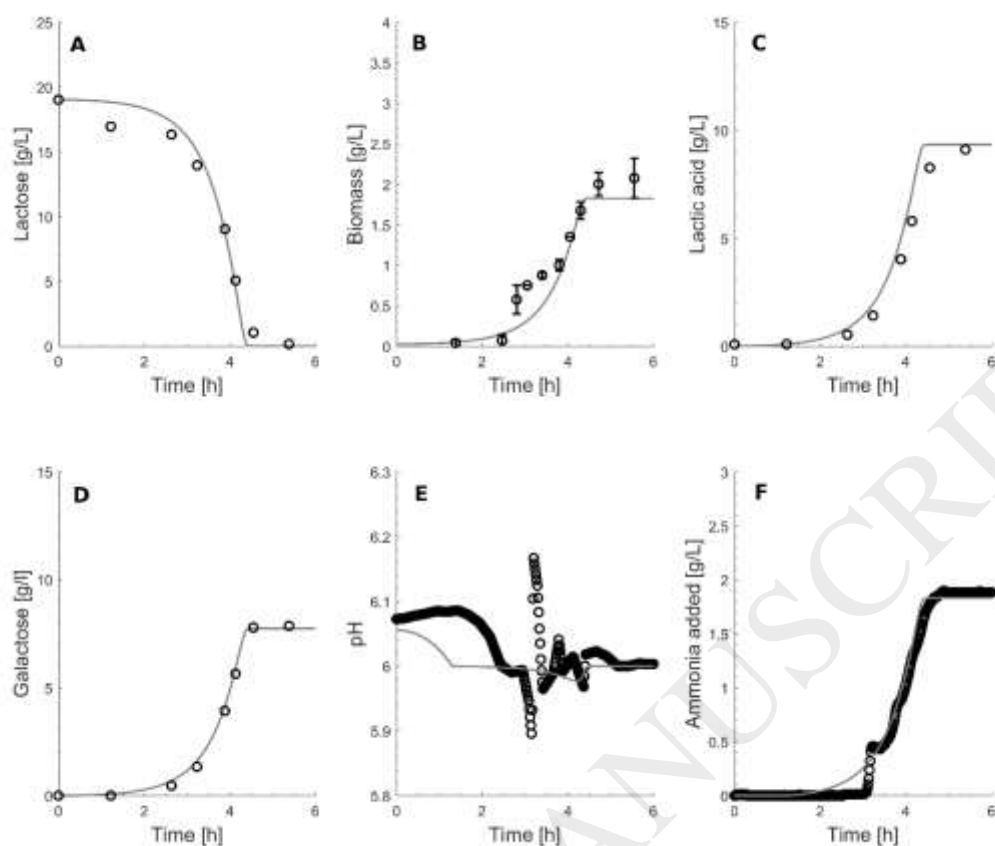


Fig. 4. Model prediction for a validation lab-scale batch fermentation. Lactose (A), biomass with standard deviation (B), lactic acid (C), galactose (D) concentrations, pH (E), and the added ammonia amount (F). The *S. thermophilus* fermentation was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. Model prediction (solid line) for the measurements (circles) of one of the two validation lab-scale batches.

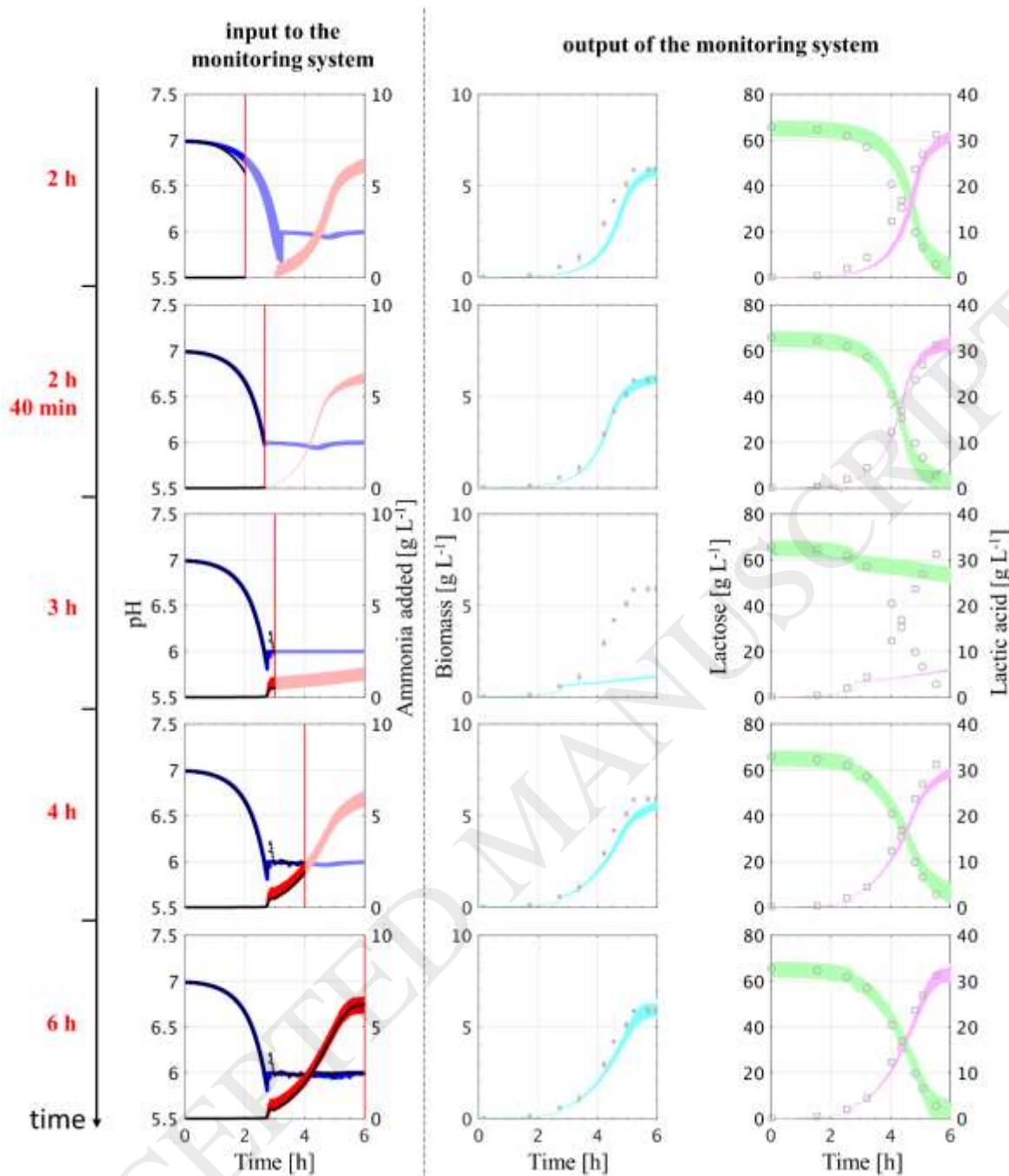


Fig. 5. Probabilistic monitoring system applied to lab-scale batch data of a *S. thermophilus* fermentation. The monitoring system reads in the on-line available data (left column, black dots), ammonia addition and pH, and predicts the state variables every 5 minutes (middle and right column). 100 Monte Carlo simulations of the dynamic model were performed within the monitoring system considering uncertainties in the initial conditions, ammonia addition, and model parameters. The 95 % confidence intervals of the predictions are shown at five time

points during the fermentation (2 h, 2 h 40 min, 3 h, 4 h, 6 h). Predictions of the pH (blue), ammonia addition (red), biomass (cyan), lactose (green), and lactic acid (magenta) concentrations are shown. The off-line measurements for biomass (gray dot with standard deviation), lactose (gray circle), and lactic acid (gray square) are shown for comparison only, but were not used for the data reconciliation and parameter update (see Fig. 1).

ACCEPTED MANUSCRIPT

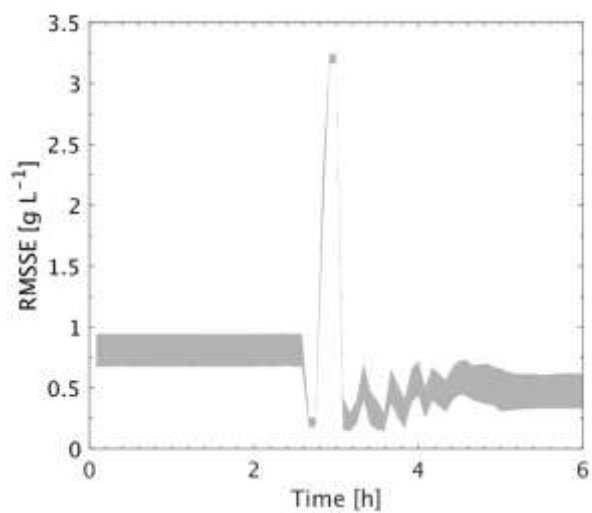


Fig. 6. 95% confidence interval of the RMSSE for the biomass prediction during the probabilistic monitoring of a *S. thermophilus* fermentation.