A compartment model for risk-based monitoring of lactic acid bacteria cultivations

Spann, Robert; Gernaey, Krist V.; Sin, Gürkan

Published in:
Biochemical Engineering Journal

Link to article, DOI:
10.1016/j.bej.2019.107293

Publication date:
2019

Document Version
Peer reviewed version

Title: A Compartment Model for Risk-Based Monitoring of Lactic Acid Bacteria Cultivations

Authors: Robert Spann, Krist V. Gernaey, Gürkan Sin

PII: S1369-703X(19)30229-3
DOI: https://doi.org/10.1016/j.bej.2019.107293
Article Number: 107293

Reference: BEJ 107293

To appear in: *Biochemical Engineering Journal*

Received date: 19 January 2019
Revised date: 15 May 2019
Accepted date: 7 July 2019

Please cite this article as: Spann R, Gernaey KV, Sin G, A Compartment Model for Risk-Based Monitoring of Lactic Acid Bacteria Cultivations, *Biochemical Engineering Journal* (2019), https://doi.org/10.1016/j.bej.2019.107293

A Compartment Model for Risk-Based Monitoring of Lactic Acid Bacteria Cultivations

Robert Spann[a], Krist V. Gernaey[a], and 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

*** Corresponding author:**
Gürkan Sin
gsi@kt.dtu.dk

Highlights

- Risk assessment tool for lactic acid bacteria cultivations
- Soft sensor implementation for dynamic model predictions
- Monte Carlo simulation for probabilistic model predictions
- Accurate prediction of pH gradients using a compartment model
- Scenario analysis of different base addition positions

Abstract

A soft sensor for on-line risk-based monitoring was applied for a 700-L *Streptococcus thermophilus* cultivation using a biochemical model that was coupled with a compartment model, the latter to account for mixing effects. The process risk, defined as the likelihood of not achieving the target biomass production per batch, was calculated continuously during the cultivation process. A Monte Carlo simulation accounted thereby for uncertainties in the model parameters. In the present cultivation, the estimated process risk was to lose ca. 3.5 % of the total production capacity. The compartment model allowed the prediction of the spatial distribution of the pH in the bioreactor. The compartment model was based on a computational fluid dynamics (CFD) simulation and its computational speed (< 2 s for one simulation) enables both on-line applications, e.g., as soft sensor, and rapid off-line process condition testing, in contrast to a CFD simulation that takes several hours/days to simulate. With the on-line soft sensor, pH gradients between pH 5.8 and 6.1 were predicted with an accuracy of ± 0.1 pH units in comparison to experimental measurements. This process analytical technology (PAT) tool was therefore further applied to test different scenarios with the aim to propose a better base addition position for pH control to reduce pH gradients in the bioreactor.

Keywords: Compartment model; Lactic acid bacteria fermentation; Monte Carlo simulation; Soft sensor; Risk assessment; Process analytical technologies (PAT)

1 Introduction

More and more scientific and risk-based methodologies have been implemented in pharmaceutical and related processes since the publications of the process analytical technology (PAT) guidance [1] and the quality by design (QbD) approach [2] [3]. These methodologies assist the industry to understand the manufacturing process and to control the process in a way that the quality of the product is assured by design. In the quality by design approach, the desired product attributes, such as purity, stability, and concentration, are defined, and critical quality attributes (CQAs) are identified. Critical process parameters (CPPs), i.e. process parameters that have an impact on the CQAs, are then determined based on process characterization studies [4]. CPPs may include temperature, pH, feed flow rate etc. The acceptable range of the CPPs is defined as the design space that leads to the desired product quality. During production, these parameters need to be controlled by the PAT system and maintained within the design space to ensure a robust process operation and to ensure product quality in bioprocesses [5].

Models are implemented to predict the CQAs by using the measured CPPs as model inputs in the framework of PAT [6]. Commonly, statistical models such as multivariate data analysis are applied to predict the effect of the CPPs on the CQAs [7,8]. Nevertheless, mechanistic models and hybrid models (a combination of mechanistic and data-driven modelling techniques) are used as well [9–12]. Since the CQAs can be hardly measured in real time, the models are especially beneficial in a soft sensor for on-line monitoring and control of industrial processes [13]. They enable to follow the dynamics of the CQAs in real time and to control the process accordingly.

Traditionally, risk assessment is conducted in the process design phase to identify process parameters with a high risk, which are then further investigated for process characterization [3,14]. Risk management methods such as Failure Mode Effects Analysis (FMEA) provide a method to evaluate these risks [15]. The risk is thereby weighted based on the severity, occurrence and detection of a process failure, i.e. a deviation from the parameter's nominal operating point. The severity is a measure for the seriousness of the consequences (with respect to the target product) if such a process failure happens; the occurrence is the expected probability of this event; and the detection indicates to which extent this process failure can or cannot be detected timely before the product is used [15].

To date, risk is quantified statically during the process design phase [16] but often not quantified dynamically in real time while a process is running. To achieve on-line risk quantification, we applied a Monte Carlo simulation in a model-based soft sensor for on-line monitoring and risk quantification in a 700 L lactic acid bacteria cultivation. Lactic acid bacteria cultures are produced in large-scale bioreactors to be used subsequently in the dairy industry e.g. for yogurt or cheese production. The applied model comprised a compartment, chemical, and bio-kinetic model. The compartment model was based on a computational fluid dynamics (CFD) model to account for heterogeneous process conditions (especially pH) in the process. The pH value is an important critical process parameter in lactic acid bacteria cultivations as lactic acid bacteria are fast acidifiers. In the Monte Carlo simulation, several uncertainties were considered: model parameter uncertainties, process input variations, and on-line measurement errors. The soft sensor predicted therefore a probability distribution of the state variables on-line, including the critical quality attribute for the case study defined as the biomass yield. The probability of not achieving the target biomass production and the corresponding risk were quantified based on the predicted probability distribution and updated on-line.

There are several important applications where this study makes an original contribution to. Thanks to the computational speed of the compartment model (compared to a CFD model) it could be applied for on-line applications, which is out of the question for a CFD model. Both, on-line pH gradient monitoring and risk quantification are demonstrated, which could improve bioprocess control. In addition, this model was utilized for testing different base addition control strategies. A design change of the bioreactor was proposed to achieve a better distribution of the base in the bioreactor and hence to reduce the pH gradient.

2 Materials and Methods

2.1 Cultivation conditions and analysis

A 700-L batch cultivation of the homolactic *Streptococcus thermophilus* (provided by Chr. Hansen A/S, Hørsholm, Denmark) was carried out in a stirred tank bioreactor at a stirring speed of 130 rpm, 40 °C, and with N_2 headspace gassing. The stirred tank bioreactor (Chemap AG, Switzerland) was equipped with three 6-blade Rushton turbines, had four baffles, and was filled with approx. 700 L cultivation medium initially (Figure 1 A). The pH was controlled by adding ammonia solution (24 % (w/v) NH_4OH) at the bottom of the bioreactor with the pH set point at 6. The pH controlling sensor was located 0.3 m above the bottom of the bioreactor close to the bioreactor wall. In addition, pH sensors (CPS471D, Endress+Hauser AG, Switzerland) measured the pH at heights of 0.1 m and 1.6 m with a distance of 0.1 m to the reactor wall. The initial pH was 6.8. The cultivation medium contained 70 g L^{-1} lactose, 10 g L^{-1} casein hydrolysate, 12 g L^{-1} yeast extract, 11.5 mM K_2HPO_4, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM $MgSO_4$, and 0.3 mM $MnSO_4$.

The dry cell weight was determined from cultivation broth samples that were first centrifuged, then washed twice with 0.9 % NaCl, and finally dried at 70 °C for 24 h. Organic acids and sugars were quantified from filtered (0.2 µm) samples in an HPLC (Dionex UltiMate 3000, Thermo Fisher Scientific, Waltham, MA). It was equipped with a refractive index detector (ERC RefractoMax 520) and an Aminex® HPX-87H column (Bio-Rad Laboratories, Hercules, CA), and operated at a flow rate of 0.6 mL min^{-1} at 50 °C using 5 mM H_2SO_4 according to suppliers instructions.

2.2 Design and validation of the compartment model

A compartment model was designed based on the steady state velocity profiles that were obtained from a CFD simulation. The CFD simulations that were conducted in ANSYS CFX 17.1 (ANSYS, Inc., US-PA) are described in detail elsewhere [17]. The CFD model was validated at 240 rpm with NaOH tracer pulse experiments from the top of the bioreactor using multi-position pH measurements that were distributed vertically in the bioreactor. Both the dynamic pH change after the pulse addition and the mixing time have been accurately predicted by the CFD model. A biokinetic model was then coupled with the CFD model to simulate a lactic acid bacteria cultivation. Both the biokinetics and pH gradients could be predicted accurately, but the simulation time was 4 days on 20 CPU cores on the DTU High Performance Computing Cluster (https://www.hpc.dtu.dk/). In order to achieve a faster simulation time, which is required for on-line applications such as in a soft sensor, a compartment model was designed and applied in this study.

The compartment model was designed utilizing the information of the CFD model. In the compartment model, each recirculation loop that was revealed by the velocity profiles of the CFD simulation was considered as one compartment (Figure 1 B and C, and Supplementary Figure S1). This was based on the fact, that mixing is very fast in each recirculation loop and hence could be regarded as a well-mixed regime. To further define the boundaries and the inflow and outflow of fluids among the compartments, the following procedure was used: (i) Horizontal planes were set up every 1 cm in the CFD model (CFX-Post) and the axial velocities and node areas were exported

for each plane from the CFD results. (ii) The flow rates were calculated as the product of the velocity and the area, and the positive (upward) and negative (downward) axial flow rates were separated. (iii) The arithmetic means of both the positive and negative flows were calculated separately. (iv) The local minima (here five) of the mean positive axial flow rates defined the interfaces between the compartments (here six compartments) (Supplementary Figure S2). These positions matched with the local maxima of the mean negative flow rates over the bioreactor height. (v) The arithmetic mean of the mean positive flow and of the absolute mean negative flow of the interface planes were set as the flow between the compartments. In this way, a continuity is ensured, which avoids mass accumulation in compartments. The properties of the compartment model can be found in Table 1.

In addition, a 7th compartment was designed capturing the 10 cm at the top of the bioreactor, which was necessary to support the tracer pulse simulation (see the Supplementary Material). The resulting configuration of the compartmental model is shown in Figure 1 C.

The compartment model has been implemented in MATLAB (The MathWorks®, Natick, MA) as an ordinary differential equation (ODE) system. To simulate the lactic acid bacteria cultivation the biokinetic and the pH model (see below) were defined together with the compartment model that represented the stirrer speed of 130 rpm in the ODE system. For comparison, a one-compartment model with a volume of 700 L was simulated in MATLAB to model the cultivation without the effects of gradients (see the Supplemental Material for the results).

2.3 Biokinetic model and pH simulation

The dynamic model comprised a biological and a chemical model as described in detail in Spann et al. [18]. The biokinetic model predicted the evolution of the state variables, such as biomass, lactic acid, and lactose concentration. The chemical model was a mixed weak acid/base model describing the dissociation reactions of the charged components, such as ammonium and lactate.

2.3.1 The biokinetic model

The biokinetic model was based on the global stoichiometric process equation [19] (Eq. (1)-(2)).

$$\boldsymbol{Lactose + Ammonia + Phosphoric\ acid \rightarrow Biomass + Lactic\ acid + Galactose} \quad (1)$$

$$\mathbf{q_S \cdot CH_2O + q_{NH} \cdot NH_3 + q_{Ph} \cdot H_3PO_4 \rightarrow q_X \cdot CH_aO_bN_cP_d + q_P \cdot CH_2O + q_{Gal} \cdot CH_2O} \quad (2)$$

where q_S represents the volumetric substrate consumption rate, q_{NH} the volumetric ammonia consumption rate, q_{Ph} the volumetric phosphoric acid consumption rate, q_X the volumetric biomass growth rate, q_P the volumetric lactic acid secretion rate, and q_{Gal} the volumetric galactose secretion rate.

The biomass growth rate was modelled as a function that depended on the maximum specific growth rate (μ_{max}), the lag-time (f_{lag}), lactose inhibition and limitation (f_S) [20], lactate inhibition(f_P) [21], the pH in the cultivation broth (f_{pH}), and the biomass concentration (C_X) [22] (Eq.(3)-(4)). Thereby, the different cultivation conditions and their effects on biomass growth were considered as shown, for example, in Aghababaie et al. (2015).

$$\frac{dC_X}{dt} = \mu_{max} \cdot f_{lag} \cdot f_S \cdot f_P \cdot f_{pH} \cdot C_X \tag{3}$$

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

where t_{lag} is the lag-time coefficient, C_S the carbon source (lactose) concentration, K_S the Monod half-saturation coefficient, K_I the substrate inhibition parameter, $K_{P,La}$ the second lactate inhibition parameter, C_{LA} the lactate concentration, K_{La1} the pH dependent lactate inhibition parameter (Eq. (5)), pH_{opt} the optimal pH parameter in the pH function, and σ_{pH} the spread parameter in the Gaussian pH function.

The inhibition of the lactic acid bacteria growth caused by lactate was pH dependent [23], and described by Eq. (5):

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

where K_{La} represents the lactate inhibition parameter, and $K_{P,pH1}$ and $K_{P,pH2}$ are the first and second lactate inhibition pH parameter, respectively.

The lactic acid synthesis was considered to be growth dependent [24]:

$$\frac{dC_P}{dt} = \alpha \cdot \frac{dC_X}{dt} \tag{6}$$

where C_P is the lactic acid concentration, and α the growth related production coefficient of lactic acid.

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

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

The kinetic parameters were estimated from the data obtained in five lab-scale cultivations, and validated with an independent data set. The experiments were conducted under different substrate (20 and 70 g L^{-1}) and pH conditions (5.5 <= pH <= 7.0) including identifiability and uncertainty analysis [18]. The derived parameters including the uncertainty of the estimated parameter values are listed in Table 2.

2.3.2 The mixed weak acid/base model

The objective of the mixed weak acid/base model was to predict the pH (as the negative logarithm of the hydrogen ion activity: $pH = -\log_{10}\{H^+\}$). To this end, this model part comprised the dissociation reactions of the charged components in the cultivation [25], such as ammonium,

lactate, phosphate, carbonate, etc. which are relevant in the investigated pH range (Table 3) [18]. The pK_a values were derived from Dawson (1969) [26] and [27] (Table 2), and the activity coefficients (f_i) were calculated by a modified Debye-Hückel model by Davies [28]:

$$\boldsymbol{log(f_i) = -1.825 \cdot 10^6 \cdot (78.3 \cdot T)^{-1.5} \cdot z_i^2 \cdot \left(\frac{\sqrt{I}}{1+\sqrt{I}} - 0.3 \cdot I\right)} \quad (8)$$

Where T represents the temperature in the cultivation broth, z_i the charge of the i-th ion, and I the ionic strength:

$$\boldsymbol{I = \frac{1}{2}\sum_i z_i^2 C_i} \quad (9)$$

A P-controller with a controller gain (K_P) of (5 mol L^{-1} · liquid volume [L]) was applied to maintain the pH at the set point (pH_{set}) value of 6 by adding ammonia solution:

$$\boldsymbol{NH_4OH_{add} = K_P \cdot (pH_{set} - pH)} \quad (10)$$

To summarize the model structure, the biokinetic model predicts the biomass growth and lactic acid production. The mixed weak acid/base model simulates the dissociation of the charged components in the cultivation broth, such as lactic acid/lactate or ammonia/ammonium with the objective to predict the pH. The changing lactic acid concentration is predicted by the biokinetic model and is then used in the mixed weak acid/base model. The pH can be predicted based on the simulated H^+ concentration, and ammonia solution is subsequently added to maintain the pH. The added amount of ammonia is also considered in the mixed weak acid/base model, hence there is a dynamic interplay of both model parts.

The model was implemented and solved in MATLAB. The numerical solver ode15s was used because the present model contains slow (e.g. the biomass growth rate) and fast time constants (e.g. the ammonia dissociation rate and the flow rates between the compartments) resulting in a stiff system of ordinary differential equations.

2.4 Probabilistic soft sensor for on-line monitoring

The aim of the probabilistic soft sensor is to predict the measurable and unmeasurable process variables, such as the biomass and substrate concentration, and the pH in real time. The algorithm for the probabilistic sensor is shown in Table 4 and the details of the soft sensor including a validation with 2 L lab-scale experiments can be found elsewhere [18].

Once the process is started, the soft sensor is updated in 5 min intervals (Table 4). The initial process conditions are defined as specified for the experiment (Table 4, step 1-2). The soft sensor uses the latest on-line measurements of the process, namely the added ammonia quantity and the pH (Table 4, step 3) to update the model parameters μ_{max} and t_{lag} (Table 4, step 4). The parameters are updated in 5 min intervals, and are then used as input to the dynamic model that predicts both the current value and the future course of the state variables (Table 4, step 5). In this study, the soft sensor was applied off-line once the cultivation was performed for demonstration purposes. The on-line measurements were hereby used as they would be available on-line. The off-line measurements were only used to assess the goodness of the model fit (see below) but not to update the soft sensor.

A Monte Carlo simulation of the dynamic model is performed every interval as explained in detail in Spann et al. [18] (Table 4, step 5). To this end, the input uncertainties are first identified and

defined. Second, random input samples are generated, and third, the Monte Carlo simulation is performed. In this study, uncertainties in the model parameters, initial conditions, and the ammonia addition are considered. The Latin hypercube sampling technique is used to generate N = 200 random samples (Supplementary Figure S3) from the input uncertainty domain [29,30]. 200 model simulations were therefore performed every interval that the soft sensor was updated providing a probability distribution of the model outputs. The model predictions of the biomass production were then assessed for the on-line risk quantification.

2.5 Assessment of the soft sensor predictions

The quality of the soft sensor predictions was assessed with the root mean sum of squared errors (RMSSE) with respect to the off-line measurements:

$$\boldsymbol{RMSSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}}^{\boldsymbol{n}} \left(\boldsymbol{y_{meas,i}} - \boldsymbol{\hat{y}_i}\right)^{\mathbf{2}}} \tag{11}$$

where n is the number of measurements, y_{meas} the off-line measurement at the corresponding time point, and ŷ the model predictions.

2.6 Process risk quantification

The risk of not achieving the target production of biomass was calculated on-line as a result of the soft sensor predictions (Table 4, step 6). The biomass was selected because the lactic acid bacteria were the desired product of this process. The target biomass production was defined as 4410 g biomass per batch that was based on previous 2 L lab-scale experiments (see the Supplementary Material for the detailed calculation). The loss/surplus (here named consequence) for each of the j Monte Carlo simulation predictions was then calculated as the difference between the model prediction (ŷ) and the target:

$$\boldsymbol{consequence_j} = \boldsymbol{\hat{y}_j} - \boldsymbol{target} \tag{12}$$

Risk is generally defined in the process industries as the likelihood of an undesirable event (u.e.) times the consequence of that event [31]. The risk of several undesirable events is consequently the sum of their individual risks (Eq. (13)). In this study, the consequence of an undesired event was the loss of the biomass in terms of total biomass amount per batch (Eq. (12)). The likelihood ($\boldsymbol{\mathcal{L}}$) of this event was the probability of this event that was predicted by the Monte Carlo simulation.

$$\boldsymbol{process\ risk} = \sum_{\boldsymbol{m}} \boldsymbol{consequence_m} \cdot \boldsymbol{\mathcal{L}}(\boldsymbol{\hat{y}_m} \mid \boldsymbol{MC}) \tag{13}$$

where m ϵ j is the number of undesirable events (u.e.) and MC represents the given Monte Carlo results.

3 Results and Discussion

A model-based soft sensor was applied to predict unmeasurable attributes such as the biomass concentration in a lactic acid bacteria cultivation and to quantify the risk of not achieving the target biomass production. To this end, a CFD-based compartment model was used to provide a reliable risk quantification since there exist pH gradients in the 700 L bioreactor that occur during the

cultivation due to insufficient mixing. In addition, the compartment model was utilized to test different scenarios with the aim to propose a better base addition position for pH control in order to reduce pH gradients in the reactor. The computation time for the compartment model was less than 2 s on an Intel® Core™ i7-5600U CPU @2.6 GHz (1000 repetitions showed this performance), which was considerably faster than the 4 days on 20 CPU cores on the DTU High Performance Computing Cluster (https://www.hpc.dtu.dk/) that was required for the CFD simulation. Thanks to this computational speed of the compartment model, the presented on-line applications and off-line scenario tests are feasible in a reasonable time. The compartment model has been validated and benchmarked against the CFD model and experimental data as described in detail in the Supplementary Material.

3.1 On-line pH gradient monitoring

The probabilistic soft sensor was applied to a historical cultivation data set of a 700 L *S. thermophilus* cultivation, whereas the historical on-line data (pH and balance readout of the ammonia addition) were used as they would be available on-line. The soft sensor used the on-line data to update the model parameters μ_{max} and t_{lag} in 5 min intervals, as described in the Materials and Methods section and in Spann et al. [18]. A Monte Carlo simulation of the dynamic model was performed within the soft sensor to account for uncertainties in the model parameters, the on-line measurement of the ammonia addition quantity, and the initial biomass inoculation and lactose concentration. The Monte Carlo simulation with 200 input samples accounting for the listed uncertainties propagated the error to the model outputs, such as the pH (Figure 2 left column) and the biomass, lactose, and lactic acid concentration (Figure 2 right column). The output of the soft sensor was therefore a probability distribution of the state variables, and the 95 % confidence intervals of the model predictions are shown. The predictions of the earlier, current, and future states of the system are shown as an example at different times: 2 h, 4 h, and 6 h (Figure 2 rows). This means that given the on-line information after 2 h of cultivation time, the soft sensor predicted the output that is shown in Figure 2 A. After 4 h of cultivation time, the soft sensor had more information available and produced the predictions shown in Figure 2 B. At the end of the cultivation after 6 h of cultivation time, the soft sensor output is depicted in Figure 3 C. Each time the soft sensor is updated with new information, i.e., every 5 minutes, it predicted the entire cultivation time, from the start to the end (0-6 h). The virtual implementation of the soft sensor with updates in 5 min intervals may be found in the Supplementary Movie.

The initial pH was ca. 6.8 and then dropped due to lactic acid secretion until the controlling pH value 6 was reached, which is the point when the base addition started (Figure 2 left column). In the first 2.5 h, no ammonia was added since the pH > 6 at the controlling position (for the position of the pH controlling sensor see Figure 1). There was consequently also no pH gradient predicted until the ammonia solution was added, as the cell density was maintained homogeneously distributed in the liquid phase.

As soon as the base addition started, a pH gradient was formed (Figure 2 B and C): In the bottom compartment of the bioreactor where the ammonia solution was added, a pH of up to 6.1 ± 0.02 was predicted due to the base addition. In the compartment 6 (second compartment from the top of the bioreactor), the pH dropped down to 5.8 ± 0.02 during the exponential growth phase due to lactic acid production. The pH gradients were qualitatively predicted with a deviation of 0.1 pH units with the mixed weak mixed acid/base model (Figure 3). The small mismatch of the pH gradient predictions is likely related to the mixed weak acid/base model because it did not represent the medium components and reactions entirely. The accuracy of the predictions could be improved in several ways: (1) Additional components such as acetic acid and amino acids that are present in the cultivation could be added to the mixed weak acid/base model. (2) Further chemical mechanisms such as precipitation could be included in the model. (3) A finer compartment model design (i.e. increasing the number of compartments) can help to represent the fluid dynamics better

quantitatively (approximating that of the CFD simulation). However, this would lead to longer mixing times in the compartment model, and hence the predicted pH gradient might increase. The deviation of 0.1 pH units could also have resulted from calibration and measurement errors (e.g. drift of the pH sensors), which can be improved by the use of several pH sensors to double-check the results. Nonetheless, the difference of 0.1 pH unit between the prediction of the compartmental model and the measurements is deemed acceptable for the process design and monitoring purposes, as this difference is not likely to cause significant errors on the microbial growth kinetics. In addition, further experiments are needed to statistically verify the goodness of the model predictions at the 700-L scale.

Cells, which are exposed to the lower pH of ca. 5.9 (as measured at the top of the bioreactor), might grow 10 % slower than at pH = 6.2 (as measured at the bottom of the bioreactor). The pH function f_{pH} = 0.89 at pH 5.9, while f_{pH} = 0.98 at pH 6.2. Nevertheless, the question remains how fast *S. thermophilus* responds to the extracellular pH changes [32], and whether these fluctuations influence the microbial activity and productivity [33]. The applied biokinetic model does not include these dynamic response effects yet, which is clearly of interest in future studies. In addition to the investigated pH gradients, the model has also been applied to predict the gradients in the biological state variables, such as lactose, lactic acid, and biomass (see the Supplemental Material). These gradients were statistically not significant.

The results of the on-line prediction of the pH gradient could be used at the production scale to minimize the risk of faulty batches for example by (i) monitoring the extent of the gradients; (ii) controlling the process; and (iii) rethinking about an improved bioreactor, impeller, or base addition design. In case the soft sensor is implemented as a monitoring tool – as shown in this study – plant operators could manually supervise the process and take actions in case the pH gradients reach a critical level. They could take risk-based decisions as they have a measure for whether the mixing is sufficient with respect to the pH. The soft sensor could also be applied for automated on-line control. In order to avoid extensive pH gradients, the impeller speed could be increased, for example. Apart from this, the cultivation temperature could also be altered, in order to regulate the biomass growth rate (which is not included in the presented model). A decreased biomass growth rate would indeed decrease the lactic acid production, and hence the pH gradient might decrease. However, this might also result in a longer cultivation time. A model-based control algorithm could be implemented to predict the best control strategy [34,35]. In case severe gradients occur frequently, results of such a model might also be an incentive for the production department to re-evaluate the bioreactor design, especially parts like the impeller or base addition inlet that could be modified more easily than the bioreactor itself. In case the impeller speed is changed, the current model is no longer sufficient to predict the mixing. Instead of a fixed compartment model representing a fixed impeller speed (and viscosity, etc.), a dynamically changing compartment model is needed. This could adjust its features depending on the process conditions. However, this is not the objective of the presented study.

The probabilistic soft sensor predicted in addition to the pH gradients the biological state variables, such as the biomass, lactose (substrate), and lactic acid concentration (Figure 2 right column). With the cultivation time, the update of the parameters μ_{max} and t_{lag} improved the prediction. After 3 h, the RMSSE for the biomass concentration prediction – the target product – was smaller than 0.4 ± 0.1 g L^{-1} that corresponds to an error of less than 10 % with respect to the final biomass concentration. This soft sensor, if implemented at a production site, provides the plant operators with a PAT tool to monitor the course of the cultivation with biological variables instead of the base addition profiles that have little direct meaning. A further strength of the soft sensor is that it could be applied to predict the end time of the cultivation, i.e. when the target cell mass will be achieved [36]. Downstream capacities, including primed machines and workforce, could be scheduled and prepared accordingly. Furthermore, also subsequent steps in the upstream process,

such as cleaning or pre-cultures for subsequent cultivations could be optimally planned. Overall, it might reduce the downtime of the plant equipment leading to a more economical operation.

3.2 On-line process risk quantification

A frequently asked question during the production process is, "What is the risk of not achieving the target yield (titer, productivity, etc.)?" In order to demonstrate the capability of the probabilistic soft sensor to quantify and update this risk while the process is running, we selected the biomass yield and total biomass production per batch as an example.

The target yield was defined to be 0.09 ± 0.003 g biomass per g lactose based on previous 2 L lab-scale experiments. The target was to achieve at least the same yield when the process was scaled up to the 700 L bioreactor. The undesired event was therefore to achieve less than the target yield. The risk was considered as the loss of product (biomass) per batch. It is quantified as the sum of the likelihood of the undesirable events times the amount of lost product (see Materials and Methods).

In our case, we got the likelihood from the output of the Monte Carlo simulation that considered uncertainties in the model parameters, initial process conditions, and the ammonia solution addition balance readout. The output of the probabilistic soft sensor were 200 model predictions. We considered the probability distribution of the biomass concentration prediction. The risk quantification method will be first presented with the biomass concentration of the final model prediction after 6 h of the cultivation as an example. Subsequently, the results of the on-line risk quantification considering the dynamic model updates will be shown.

First, the biomass yield was calculated and a histogram of the predicted probability distribution is shown (Figure 4 A). The biomass yield distribution ranged from 0.076 to 0.096 g biomass (g lactose)$^{-1}$. Some simulations did not reach the target yield. Second, the cumulative distribution function of the predicted yield minus the target yield was calculated (Figure 4 B). In this example, the probability of not achieving the target yield was 75 %. In other words, 75 % of the 200 simulations predicted that the final biomass yield was smaller than the desired target yield. 25 % of the Monte Carlo predictions were accordingly equal or larger than the target yield. Third, the biomass production of the entire batch was calculated considering the bioreactor volume (Figure 4 C). The total production amount might be of higher interest for a company than the yield as the obtained total mass is crucial for sale. Product quality aspects were not considered in this work but could of course be included in the model. The risk is equivalent with the area under the cumulative distribution function that corresponds to the undesired events (Figure 4 C, grey shaded area). In this example, the risk was the loss of 140 g biomass per batch.

As a result of the probabilistic soft sensor, the risk could be predicted on-line considering the model updates every 5 minutes (Figure 5). In this study, the predicted biomass concentration at the end of the cultivation (after 6 h) was considered. However, the time point and the desired product or other process attributes could be amended for other studies.

In the beginning of the cultivation, the risk could not be properly predicted as no or only little information from the on-line measurements was available (Figure 5, Initialization phase). As more on-line data was provided, the soft sensor could be updated, and hence the model predictions became more accurate. The on-line risk calculation needed therefore an initialization phase, waiting for enough on-line data (ammonia addition and pH measurements) to update the lag-time parameter and the maximum specific growth rate parameter. Once enough on-line data was available, a proper on-line risk quantification was achieved during the process operation. In this study, the initialization phase was set until t_{lag} was finally updated after 2 h and 25 min, when the base addition started.

Nevertheless, the boundaries for the initialization phase need to be adapted in case the system would be applied for a different cultivation system.

The risk was low, i.e. close to zero, when all Monte Carlo simulations achieved the target. In the present case, the risk is low after 2.5 h (Figure 5, Risk prediction phase). Later, the risk became higher between 3-4 h reaching a predicted risk of 140 g biomass that could be lost per batch. Next, the risk was predicted to be low again after 4 h and 30 min, and finally, when the soft sensor was updated after 6 h cultivation, the risk was 140 g biomass per batch. This oscillating risk prediction could be attributed to the on-line update of the μ_{max} parameter that was dependent on the base addition. The base addition reflects the lactic acid secretion by the lactic acid bacteria and hence the biomass growth can be predicted. The on-line risk quantification captured therefore effects of the biomass growth rate. The growth rate changed in the presented cultivation. This resulted in the oscillations of the predicted biomass yield that led to the oscillating risk prediction (Supplementary Figure S4). The predicted risk is nevertheless within the 1σ range of the target yield (0.09 ± 0.003 g biomass (g lactose)$^{-1}$) and regarded as natural variability of the process. This indicated that there was no yield decrease in the presented 700 L cultivation compared to the 2 L lab-scale experiments. However, replicates of the 700 L cultivations would be needed to validate the results statistically.

For an industrial application, the risk could be calculated as profit loss, i.e. an economic risk assessment [37,38], as the economic aspect is the driving force for the production. The risk could then be quantified in e.g. $ per batch. Furthermore, possible loss of product quantity during the downstream operations could also be considered. The benefit from the monitoring system is that one can reflect and take action either by automated on-line control or manually, i.e. the action by a process operator. The operators could obtain an on-line measure to assess the risk of faulty batches and react accordingly, e.g. by increasing the stirrer speed to decrease pH gradients. In future, the soft sensor could be applied for on-line control, and hence controlling process parameters in such a way that the risk of losing product or profit remains as low as possible. To this end, it might also be necessary to include further uncertainties in Monte Carlo simulation, such as stochastic variabilities, e.g. process equipment failures [39].

3.3 Scenario Tests with Different Base Addition Positions

The compartment model could also be applied for process design besides the above mentioned on-line monitoring and control applications. It could be applied to simulate different scenarios, e.g. to test different base addition strategies and pH sensor locations in order to decrease the pH gradient as performed in this study.

In the studied system, the base was added below the bottom impeller and the controlled pH was measured in compartment 2. Here, a pH gradient between 5.8 and 6.1 was predicted in the exponential growth phase as discussed above (Figure 3). In case the base addition would be placed underneath the middle impeller in the bioreactor, a pH gradient between 5.9 and 6.05 was predicted using the compartment model (Figure 6 A). If the base addition was placed below the top impeller while the controlled pH was still measured in compartment 2, a pH gradient between 5.95 and 6.3 was predicted (Figure 6 B). If the base would be added to the top of the bioreactor, a pH gradient between 5.95 and 7.0 was predicted (Figure 6 C).

According to these results, the pH gradient could be significantly reduced if the base inlet would be placed below the middle impeller. In the worst-case scenario, with the base addition from the top together with the measurement of the pH at the bottom (the measurement input to the controller is in compartment 2, Figure 6 C) the pH gradient would increase drastically. In general, also other combinations of the position of both the base addition and the controlling pH measurement could be assessed. Experimental validation of the proposed design for the sensor location using the simulation of the compartmental model iteratively would help to improve the optimization of the

process. Another alternative would also be multiple point injection of the base at the large scale as suggested for the chemical industry in order to reduce the gradients [40].

Indeed, the compartment model could complement the process design phase at an early stage; in particular, it allows exploring and testing different scenarios with a short simulation time. Using a CFD simulation instead requires much longer simulation times and more computational resources [41]. However, we believe that a detailed analysis of the best candidates would be needed subsequently with a high resolution CFD simulation to substantiate the results. As an example, the gradients in the area in the vicinity of the base addition point could not be simulated with the compartment model. A higher pH is expected here as the base concentration is very high [42], which could be predicted by the CFD simulation with a higher spatial resolution [17]. Compartment models offer an opportunity as a compromise between the computational complexity for describing mixing and the simulation speed that is needed for various applications such as on-line monitoring and control applications, as well as fast off-line simulations, e.g. to test different scenarios for reactor geometry, mixing equipment and sensor locations [43].

Moreover, thanks to the promising results obtained in this study with a 700 L bioreactor, it is now intended to apply the CFD-based compartment model to larger (production-scale) bioreactors, e.g. > 50 m^3, and to support industrial production processes. The presented tools can for example be applied to investigate to which extent pH gradients exist at production scale, whether they could have an influence on the metabolic activity and especially the biomass growth and product quality, and to design scale-down experiments at the lab scale that mimic large-scale conditions [33,44,45]. We expect that the risk-based process monitoring methodology and the underlying modeling can be used for a number of application in process scale-up and optimization studies such as (i) to calculate profit risks for evaluating different automation or control strategies, (ii) scenario testing and evaluation as part of design of experiments to find out the process condition space for experimentally study, among others.

4 Conclusion

A soft sensor was applied as a PAT tool that was based on a mechanistic model and a CFD-based compartment model. These applications were feasible thanks to the computational speed of the compartment model that could not be achieved by a CFD model. An on-line risk assessment tool was proposed to quantify both pH gradients and the risk of not achieving the target production in a lactic acid bacteria cultivation. It provided, on the one hand, an on-line prediction of the pH gradient in the bioreactor, which is a critical process parameter. This would enable plant operators to assess the mixing and the base addition strategy. On the other hand, the soft sensor quantified the risk of not achieving the target biomass production. The likelihood of the undesired event, i.e. the target biomass production could not be achieved, was calculated based on the probabilistic model predictions that were obtained from the Monte Carlo simulation of the soft sensor model. The Monte Carlo simulation was performed to consider uncertainties in the model parameters, on-line measurements, and initial process conditions. In the investigated 700 L cultivation, the risk was to lose max. 140 g biomass per batch. The compartment model could also be applied to test different scenarios by simulating the effect of different base addition positions. The model suggested for the studied system that the pH gradients could be decreased if the base inlet would be moved to the middle of the bioreactor. The future objective of this study is the implementation of the soft sensor for risk-based decision making and control in large-scale cultivations under consideration of techno-economic risks.

5 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.

6 Conflict of interest

None declared.

7 Nomenclature

a	Ordinate intercept of the SRC model
b_i	Linear regression coefficient for the i-th model parameter
C_{Gal}	galactose concentration (g L^{-1})
C_{Glc}	glucose concentration (g 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_{tCO}	total carbonic acid ($H_2CO_3^*$ and HCO_3^-) concentration (mol L^{-1})
C_{tNH}	total concentration of NH_4^+ and NH_3 (g L^{-1})
C_{tPh}	total concentration of H_3PO_4, $H_2PO_4^-$, and HPO_4^{2-} (g L^{-1})
C_{tZ}	total concentration of the unknown compound (dissociated and undissociated form) (mol L^{-1})
C_X	biomass concentration (g L^{-1})
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})
$K'_{r,Z}$	apparent reverse rate constant for the dissociation of the unknown component (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 measurements
pH_{opt}	optimal pH parameter in the pH function (-)
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 (C-mol L^{-1} h^{-1})
q_{NH}	volumetric ammonia consumption rate (mol L^{-1} 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})
RMSSE	root mean sum of squared errors (g L^{-1})
SRC_i	standardized regression coefficient of the i-th parameter
T	temperature in the bioreactor (K)
t	time variable (h)
t_{lag}	lag-time coefficient (h)
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})
$\hat{\theta}_{i,j}$	i-th parameter value used on the j-th Monte Carlo simulation

μ_{max}	maximum specific growth rate (h^{-1})
σ	standard deviation
σ_{pH}	spread parameter in the gaussian pH function
$\sigma_{\hat{\theta}_i}$	standard deviation of the estimated parameter
$\sigma_{\hat{y}_X}$	standard deviation of the biomass concentration distribution

8 References

[1] 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.

[2] ICH Q8(R2), Pharmaceutical Development, 8 (2009) 1–28. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf.

[3] J. Rantanen, J. Khinast, The Future of Pharmaceutical Manufacturing Sciences, J. Pharm. Sci. 104 (2015) 3612–3638. doi:10.1002/jps.24594.

[4] A.S. Rathore, H. Winkle, Quality by design for biopharmaceuticals, Nat. Biotechnol. 27 (2009) 26–34. doi:10.1038/nbt0109-26.

[5] S. Gnoth, M. Jenzsch, R. Simutis, A. Lübbert, Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control, J. Biotechnol. 132 (2007) 180–186. doi:10.1016/j.jbiotec.2007.03.020.

[6] J. Glassey, K. V. Gernaey, C. Clemens, T.W. Schulz, R. Oliveira, G. Striedner, C.F. Mandenius, Process analytical technology (PAT) for biopharmaceuticals, Biotechnol. J. 6 (2011) 369–377. doi:10.1002/biot.201000356.

[7] S.M. Mercier, B. Diepenbroek, R.H. Wijffels, M. Streefland, Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations, Trends Biotechnol. 32 (2014) 329–336. doi:10.1016/j.tibtech.2014.03.008.

[8] A.S. Rathore, S. Mittal, M. Pathak, V. Mahalingam, Chemometrics application in biotech processes: Assessing comparability across processes and scales, J. Chem. Technol. Biotechnol. 89 (2014) 1311–1316. doi:10.1002/jctb.4428.

[9] W. Sommeregger, B. Sissolak, K. Kandra, M. von Stosch, M. Mayer, G. Striedner, Quality by control: Towards model predictive control of mammalian cell culture bioprocesses, Biotechnol. J. 12 (2017) 1600546. doi:10.1002/biot.201600546.

[10] D. Solle, B. Hitzmann, C. Herwig, M. Pereira Remelhe, S. Ulonska, L. Wuerth, A. Prata, T. Steckenreiter, Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation, Chemie-Ingenieur-Technik. 89 (2017) 542–561. doi:10.1002/cite.201600175.

[11] J. Kager, C. Herwig, I.V. Stelzer, State estimation for a penicillin fed-batch process

combining particle filtering methods with online and time delayed offline measurements, Chem. Eng. Sci. 177 (2018) 234–244. doi:10.1016/j.ces.2017.11.049.

[12] 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. doi:10.1016/j.tibtech.2017.07.002.

[13] C.F. Mandenius, R. Gustavsson, Mini-review: Soft sensors as means for PAT in the manufacture of bio-therapeutics, J. Chem. Technol. Biotechnol. 90 (2015) 215–227. doi:10.1002/jctb.4477.

[14] E. Stocker, G. Toschkoff, S. Sacher, J.G. Khinast, Use of mechanistic simulations as a quantitative risk-ranking tool within the quality by design framework, Int. J. Pharm. 475 (2014) e245–e255. doi:10.1016/j.ijpharm.2014.08.055.

[15] ICH Q9, Quality Risk Management, (2005). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf.

[16] S. Adam, D. Suzzi, C. Radeke, J.G. Khinast, An integrated Quality by Design (QbD) approach towards design space definition of a blending unit operation by Discrete Element Method (DEM) simulation, Eur. J. Pharm. Sci. 42 (2011) 106–115. doi:10.1016/j.ejps.2010.10.013.

[17] R. Spann, J. Glibstrup, K. Pellicer Alborch, S. Junne, P. Neubauer, C. Roca, D. Kold, A. Eliasson Lantz, G. Sin, K. V. Gernaey, U. Krühne, CFD predicted pH gradients in lactic acid bacteria cultivations, Biotechnol. Bioeng. (2018).

[18] R. Spann, C. Roca, D. Kold, A. Eliasson Lantz, K. V. Gernaey, G. Sin, A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations, Biochem. Eng. J. 135 (2018) 49–60. doi:10.1016/j.bej.2018.03.016.

[19] J. Villadsen, J. Nielsen, G. Lidén, Bioreaction Engineering Principles, Springer US, Boston, MA, 2011. doi:10.1007/978-1-4419-9688-6.

[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. doi:10.1007/s002530051232.

[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. doi:10.1016/j.fbp.2014.09.007.

[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. doi:10.1016/S0141-0229(01)00466-5.

[23] W. Fu, A.P. Mathews, Lactic acid production from lactose by< i> Lactobacillus plantarum</i>: kinetic model and effects of pH, substrate, and oxygen, Biochem. Eng. J. 3 (1999) 163–170.

[24] 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.

[25] 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. doi:10.1016/S0043-1354(99)00334-6.

[26] R.M.C. Dawson, Data for Biochemical Research, Clarendon Press, Oxford, 1969.

[27] R.E. Loewenthal, G.A. Ekama, G.R. Marais, Mixed weak acid/base systems Part I - Mixture characterisation., Water SA. 15 (1989) 3–24.

[28] C.W. Davies, Ion Association, Butterworth, Londond, 1962.

[29] 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. doi:10.1080/00401706.1979.10489755.

[30] 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. doi:10.1016/j.watres.2009.03.048.

[31] I. Cameron, R. Raman, Process Systems Risk Management, Elsevier Academic Press, 2005.

[32] H. Siegumfeldt, K. Björn Rechinger, M. Jakobsen, Dynamic changes of intracellular pH in individual lactic acid bacterium cells in response to a rapid drop in extracellular pH., Appl. Environ. Microbiol. 66 (2000) 2330–2335. doi:10.1128/AEM.66.6.2330-2335.2000.

[33] A. Amanullah, C.M. McFarlane, A.N. Emery, A.W. Nienow, Scale-down model to simulate spatial pH variations in large-scale bioreactors, Biotechnol. Bioeng. 73 (2001) 390–399. doi:10.1002/bit.1072.

[34] J.E. Jiménez-Hornero, I.M. Santos-Dueñas, I. García-García, Optimization of biotechnological processes. The acetic acid fermentation. Part III: Dynamic optimization, Biochem. Eng. J. 45 (2009) 22–29. doi:10.1016/j.bej.2009.01.011.

[35] L. Mears, S.M. Stocks, G. Sin, K. V. Gernaey, A review of control strategies for manipulating the feed rate in fed-batch fermentation processes, J. Biotechnol. 245 (2017) 34–46. doi:10.1016/j.jbiotec.2017.01.008.

[36] D.P. Petrides, C.A. Siletti, The Role of Process Simulation and Scheduling Tools in the Development and Manufacturing of Biopharmaceuticals, in: Proc. 2004 Winter Simul. Conf. 2004., IEEE, 2004: pp. 960–965. doi:10.1109/WSC.2004.1371568.

[37] C.L. Gargalo, P. Cheali, J.A. Posada, K. V. Gernaey, G. Sin, Economic Risk Assessment of Early Stage Designs for Glycerol Valorization in Biorefinery Concepts, Ind. Eng. Chem. Res. 55 (2016) 6801–6814. doi:10.1021/acs.iecr.5b04593.

[38] A. Hasanly, M. Khajeh Talkhoncheh, M. Karimi Alavijeh, Techno-economic assessment of bioethanol production from wheat straw: a case study of Iran, Clean Technol. Environ. Policy. 20 (2017) 357–377. doi:10.1007/s10098-017-1476-0.

[39] M.D. McKay, J.D. Morrison, S.C. Upton, Evaluating prediction uncertainty in simulation models, Comput. Phys. Commun. 117 (1999) 44–51. doi:10.1016/S0010-4655(98)00155-6.

[40] E.L. Paul, V.A. Atiemo-Obeng, S.M. Kresta, eds., Handbook of Industrial Mixing, John Wiley \& Sons, Inc., Hoboken, NJ, USA, 2003. doi:10.1002/0471451452.

[41] U. Rehman, W. Audenaert, Y. Amerlinck, T. Maere, M. Arnaldos, I. Nopens, How well-mixed is well mixed? Hydrodynamic-biokinetic model integration in an aerated tank of a full-scale water resource recovery facility, Water Sci. Technol. 76 (2017) 1950–1965. doi:10.2166/wst.2017.330.

[42] C. Langheinrich, A.W. Nienow, Control of pH in large-scale, free suspension animal cell bioreactors: Alkali addition and pH excursions, Biotechnol. Bioeng. 66 (1999) 171–179. doi:10.1002/(SICI)1097-0290(1999)66:3<171::AID-BIT5>3.0.CO;2-T.

[43] A. Nørregaard, C. Bach, U. Krühne, U. Borbjerg, K. V. Gernaey, Hypothesis-driven compartment model for stirred bioreactors utilizing computational fluid dynamics and multiple pH sensors, Chem. Eng. J. (2019). doi:10.1016/j.cej.2018.08.191.

[44] C. Haringa, W. Tang, A.T. Deshmukh, J. Xia, M. Reuss, J.J. Heijnen, R.F. Mudde, H.J. Noorman, Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines, Eng. Life Sci. 16 (2016) 652–663. doi:10.1002/elsc.201600061.

[45] C. Haringa, W. Tang, G. Wang, A.T. Deshmukh, W.A. van Winden, J. Chu, W.M. van Gulik, J.J. Heijnen, R.F. Mudde, H.J. Noorman, Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization, Chem. Eng. Sci. 175 (2018) 12–24. doi:10.1016/j.ces.2017.09.020.

9 Figure legends

Figure 1. Simplified bioreactor setup with dimensions in cm (A), the steady-state velocity profile predicted by the CFD model (B), and the compartment model (C). A bioreactor with three six-blade Rushton turbines, four baffles, and a liquid volume of 700 L was used. The ammonia (alkali inlet) was added at the bottom of the bioreactor. 7 compartments were designed based on the axial velocities of the steady-state CFD solution resembling a stirrer speed of 130 rpm.

Figure 2. Probabilistic soft sensor to predict the pH gradient (left column), biomass growth, lactic acid production, and substrate consumption (right column). The soft sensor using the compartment model was applied to data of a 700 L *S. thermophilus* batch cultivation. Predictions of the pH at the controlling position (blue line), the pH at the bottom compartment (blue dots), the pH in the top compartment (blue dashed line), biomass (cyan), lactic acid (magenta) and lactose (green) are shown. The soft sensor outputs that were created with the on-line information after 2, 4, and 6 h of cultivation time are shown in the boxes A, B, and C, respectively. The off-line measurements for biomass (gray dots with standard deviation), lactic acid (grey squares), and lactose (gray circles) are shown for comparison only but were not used for the on-line update of the parameters.

Figure 3. Compartment model predictions and measurements of the pH gradient in the 700 L *S. thermophilus* cultivation with a stirrer speed of 130 rpm. Monte Carlo simulation of the compartment model with the mixed weak acid/base model (dashed line) and measurements (dots) that were recorded every 1 s. The 95 % confidence interval is shown.

Figure 4. Probability distribution of the target biomass yield and production quantity. The probability distribution of the biomass yield after 6 h of cultivation as predicted by the Monte Carlo simulation (A); Cumulative distribution function of the yield with respect to the target yield (B); Cumulative distribution function of the total biomass production per batch (C). The grey shaded area under the cumulative distribution function represents the risk.

Figure 5. On-line risk quantification during the cultivation. The risk as biomass production loss per batch was quantified on-line based on the output of the probabilistic soft sensor that was updated in 5 min intervals. Limited on-line measurements were available in the beginning of the cultivation that did not enable a proper risk quantification (Initialization phase). With more on-line data, the dynamic model parameters could be updated in the soft sensor allowing the risk prediction (Risk prediction phase).

Figure 6. pH gradients predicted by the compartment model when the base would be added at different positions. Base addition below the middle impeller (A), below the top impeller (B), and from the top of the bioreactor (C).

10 Tables

Table captions

Table 1. Properties of the compartment model at 130 rpm (half of the bioreactor was modelled).

Table 2 Kinetic parameters of the dynamic model for the *S. thermophilus* cultivation.

Table 3. Kinetics of the mixed weak acid/base model. f_m and f_d are the mono- and divalent activity coefficients, respectively; see Musvoto et al. [25] and Loewenthal et al. [27].

Table 4. Methodology of the probabilistic soft sensor

Table 1. Properties of the compartment model at 130 rpm (half of the bioreactor was modelled).

Compartment interconnection	Interface area [m^2]	Velocity [m s^{-1}]	Compartment no.	Volume [m^3]
1 ↔ 2	0.1754	0.0693	1	0.0388
2 ↔ 3	0.1839	0.0476	2	0.0364
3 ↔ 4	0.1754	0.0810	3	0.0671
4 ↔ 5	0.1839	0.0527	4	0.0768
5 ↔ 6	0.1754	0.0669	5	0.0396
6 ↔ 7	0.1847	0.0541	6	0.0806
			7	0.0191

Table 2 Kinetic parameters of the dynamic model for the *S. thermophilus* cultivation.

Symbol	Value	Std. deviation	Reference
Biological model			
K_I	164 g L^{-1}	n.d.	[20]
K_{La}	19.80 g L^{-1}	0.05 g L^{-1}	[18]
$K_{P,La}$	0.24 L g^{-1}	0.03 L g^{-1}	[18]
$K_{P,pH1}$	20	n.d.	[18]
$K_{P,pH2}$	7	n.d.	[18]
K_S	0.79 g L^{-1}	n.d.	[20]
pH_{opt}	6.39	0.06	[18]
t_{lag}	updated in the soft sensor		
Y_{gal}	0.69 g g^{-1}	0.04 g g^{-1}	[18]
α	5.19 g g^{-1}	0.01 g g^{-1}	[18]
μ_{max}	Initial value: 2.06 h^{-1}, updated in the soft sensor		[18]
σ_{pH}	1.42	0.04	[18]
Mixed weak acid/base model			
$K'_{r,C1}$	10^7 s^{-1}		[25]
$K'_{r,LA}$	10^7 s^{-1}		[25]
$K'_{r,NH}$	10^{12} s^{-1}		[25]
$K'_{r,P1}$	10^8 s^{-1}		[25]
$K'_{r,P2}$	10^{12} s^{-1}		[25]
$K'_{r,W}$	10^{10} s^{-1}		[25]
$K'_{r,Z}$	10^7 s^{-1}		[25]
pK_{C1}	$3404.7/(T - 14.8435 + 0.03279 \cdot T)$		[27]
pK_{LA}	3.86		[26]
pK_{NH}	$2835.8/(T - 0.6322 + 0.00123 \cdot T)$		[27]
pK_{P1}	$799.3/(T - 4.5535 + 0.01349 \cdot T)$		[27]
pK_{P2}	$1979.5/(T - 5.3541 + 0.01984 \cdot T)$		[27]
pK_W	14		[27]
pK_Z	9.4		[18]
T	313 K		measured process condition
Initial Conditions			
$C_{Gal,t=0}$	0.0 g L^{-1}		
$C_{Glc,t=0}$	0.0 g L^{-1}		
$C_{P,t=0}$	0.0 g L^{-1}		
$C_{S,t=0}$	70 g L^{-1}	2.3 g L^{-1}	
$C_{tCO,t=0}$	$1.002 \cdot 10^{-5}$ $mol\ L^{-1}$		
$C_{tNH,t=0}$	0.005 g L^{-1}		
$C_{tPh,t=0}$	2 g L^{-1}		
$C_{tZ,t=0}$	2 mol L^{-1}		
$C_{X,t=0}$	0.025 g L^{-1}	$8 \cdot 10^{-4}$ g L^{-1}	

Table 3. Kinetics of the mixed weak acid/base model. f_m and f_d are the mono- and divalent activity coefficients, respectively; see Musvoto et al. [24] and Loewenthal et al. [26].

Dissociation process	Reaction	reaction rate vector	apparent equilibrium constant
Ammonium	$NH_4^+ \leftrightarrow NH_3 + H^+$	$K'_{r,NH} \cdot K'_{NH} \cdot [NH_4^+] - K'_{r,NH} \cdot [NH_3] \cdot [H^+]$	$K'_{NH} = 10^{-pK_{NH}}$
Phosphate 1	$H_3PO_4 \leftrightarrow H_2PO_4^- + H^+$	$K'_{r,P1} \cdot K'_{P1} \cdot [H_3PO_4] - K'_{r,P1} \cdot [H_2PO_4^-] \cdot [H^+]$	$K'_{P1} = 10^{-pK_{P1}}/f_m^2$
Phosphate 2	$H_2PO_4^- \leftrightarrow HPO_4^{2-} + H^+$	$K'_{r,P2} \cdot K'_{P2} \cdot [H_2PO_4^-] - K'_{r,P2} \cdot [HPO_4^{2-}] \cdot [H^+]$	$K'_{P2} = 10^{-pK_{P2}}/f_d$
Carbonate 1	$H_2CO_3^* \leftrightarrow HCO_3^- + H^+$	$K'_{r,C1} \cdot K'_{C1} \cdot [H_2CO_3^*] - K'_{r,C1} \cdot [HCO_3^-] \cdot [H^+]$	$K'_{C1} = 10^{-pK_{C1}}/f_m^2$
Lactate	$C_3H_6O_3 \leftrightarrow C_3H_5O_3^- + H^+$	$K'_{r,LA} \cdot K'_{LA} \cdot [C_3H_6O_3] - K'_{r,LA} \cdot [C_3H_5O_3^-] \cdot [H^+]$	$K'_{LA} = 10^{-pK_{LA}}/f_m^2$
Water	$H_2O \leftrightarrow OH^- + H^+$	$K'_{r,W} \cdot K'_W - K'_{r,W} \cdot [OH^-] \cdot [H^+]$	$K'_W = 10^{-pK_W}/f_m^2$
Unknown compound	$ZH^+ \leftrightarrow Z + H^+$	$K'_{r,Z} \cdot K'_Z \cdot [ZH^+] - K'_{r,Z} \cdot [Z] \cdot [H^+]$	$K'_Z = 10^{-pK_Z}/f_m^2$

Table 4. Methodology of the probabilistic soft sensor

Step 1 **Define the initial process conditions (x_0) of the real process and model**

Step 2 **Start the cultivation**

Iterate step 3 to 7 in 5 minutes intervals until t_{end} (cultivation completion)

Step 3 **Read on-line measurements**

pH and ammonia addition rate (pH, $q_{NH,add}$)

Step 4 **Update the kinetic parameters μ_{max} and t_{lag}**

Step 4.1 Data reconciliation

$$NH_4^+ + C_3H_5O_3^- = q_{NH,add} + q_P = 0$$

$$q_X = q_X / a$$

Step 4.2 Parameter update

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

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

Step 5 **Monte Carlo simulation of the model**

Step 5.1 Define the input uncertainty space (once/ not every interval) (σ_θ,σ_{x_0})

Step 5.2 Sample the independent input matrix (once/ not every interval)

SAMPLE MATRIX ($\Theta_{l \times N}$)

Step 5.3 Monte Carlo simulation

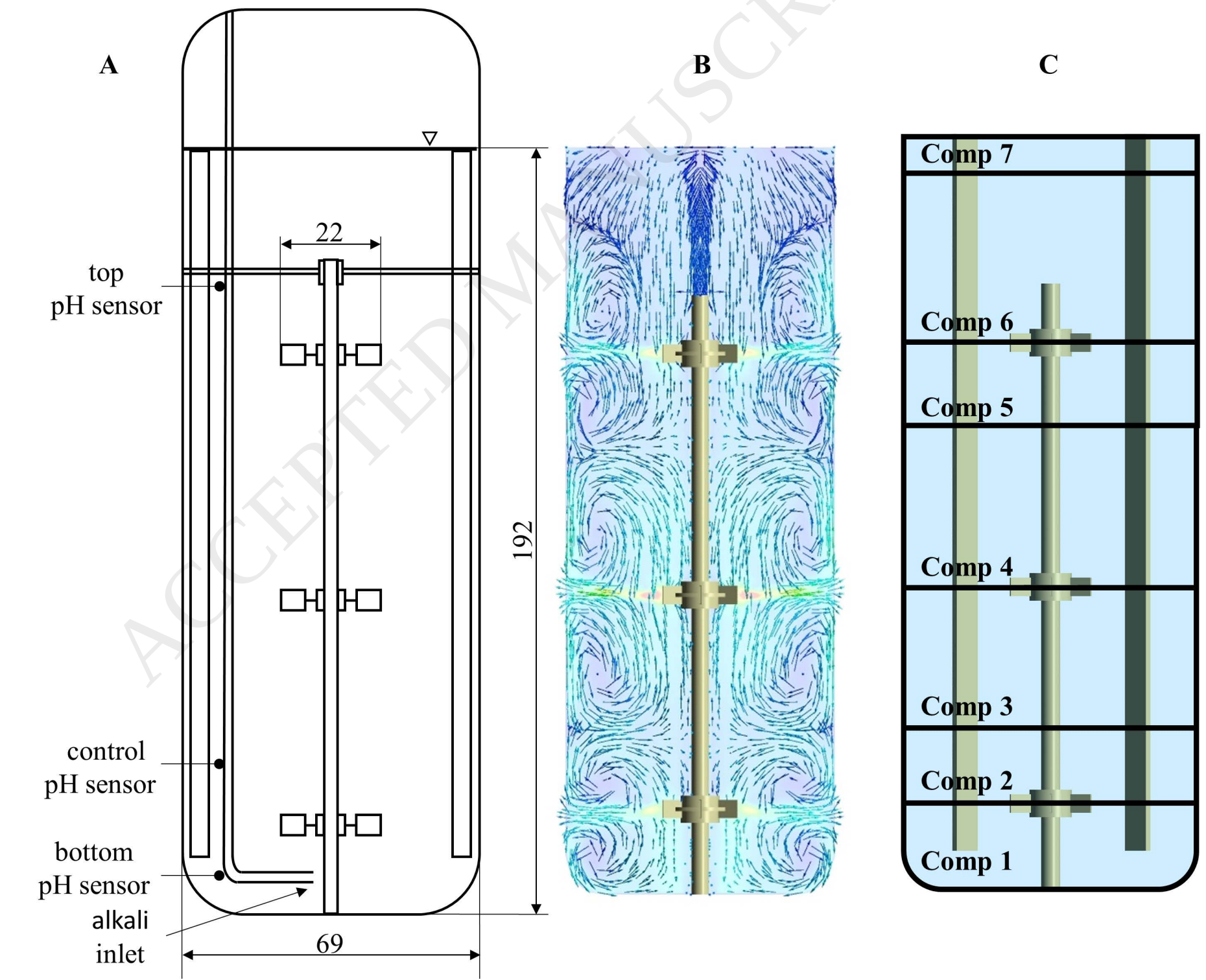
for 1:N

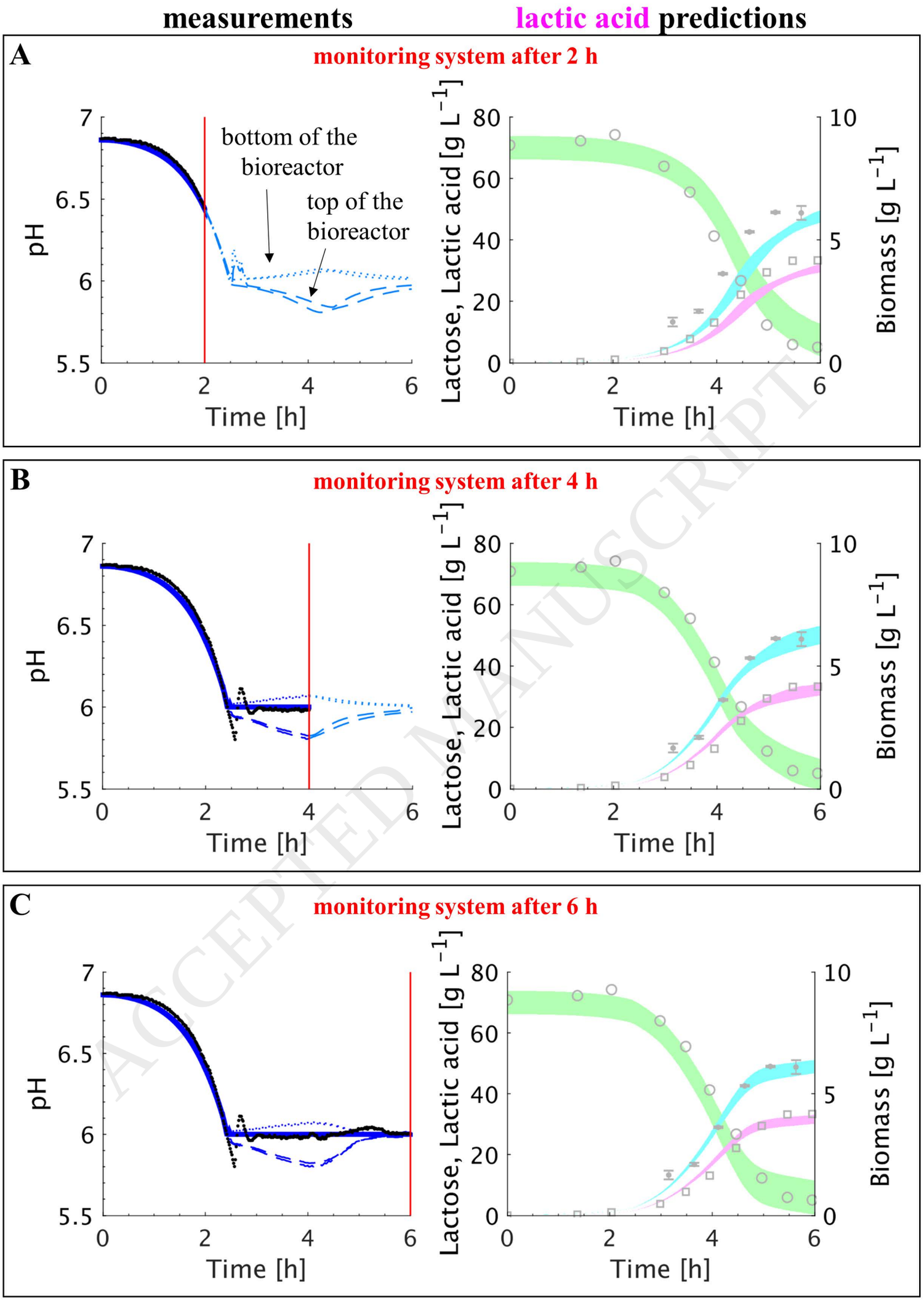
Solve y(t) = Model(θ_j,x_0)

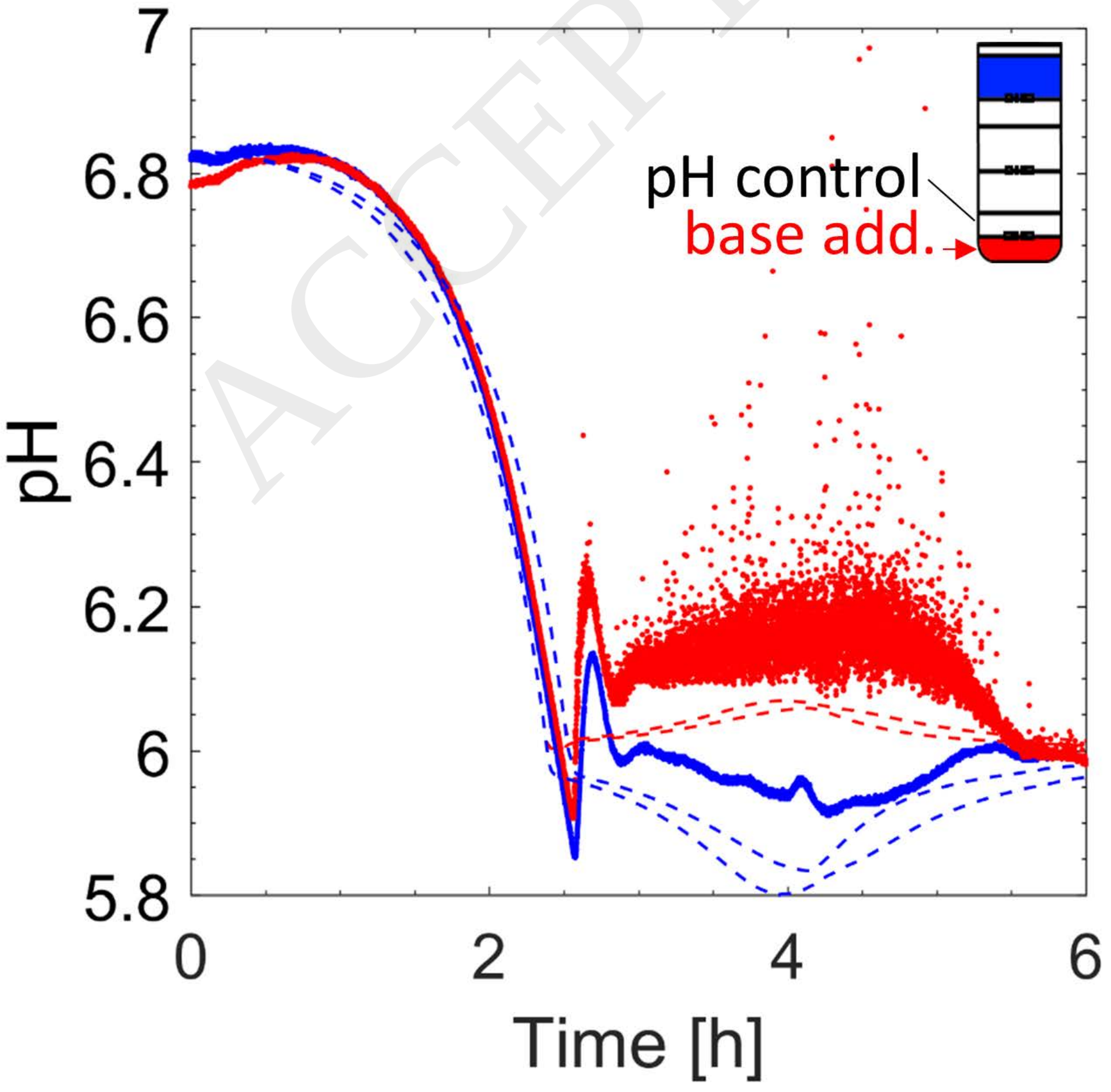
Step 6 **Process risk quantification**

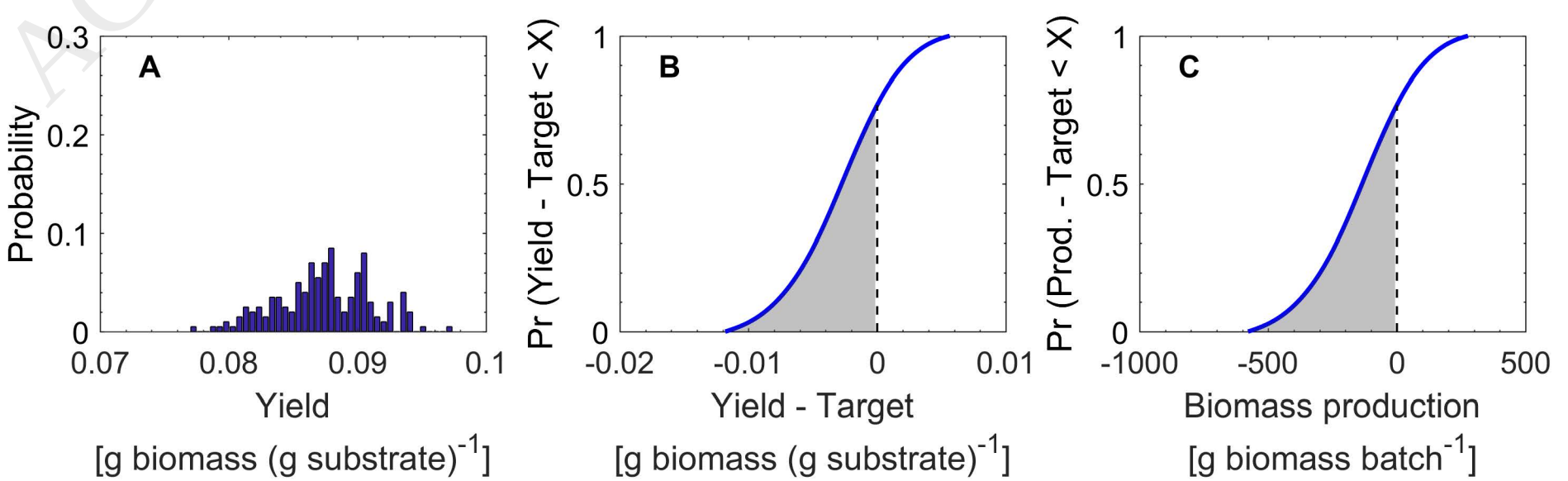
$$process\ risk = \sum_m consequence \cdot likelihood$$

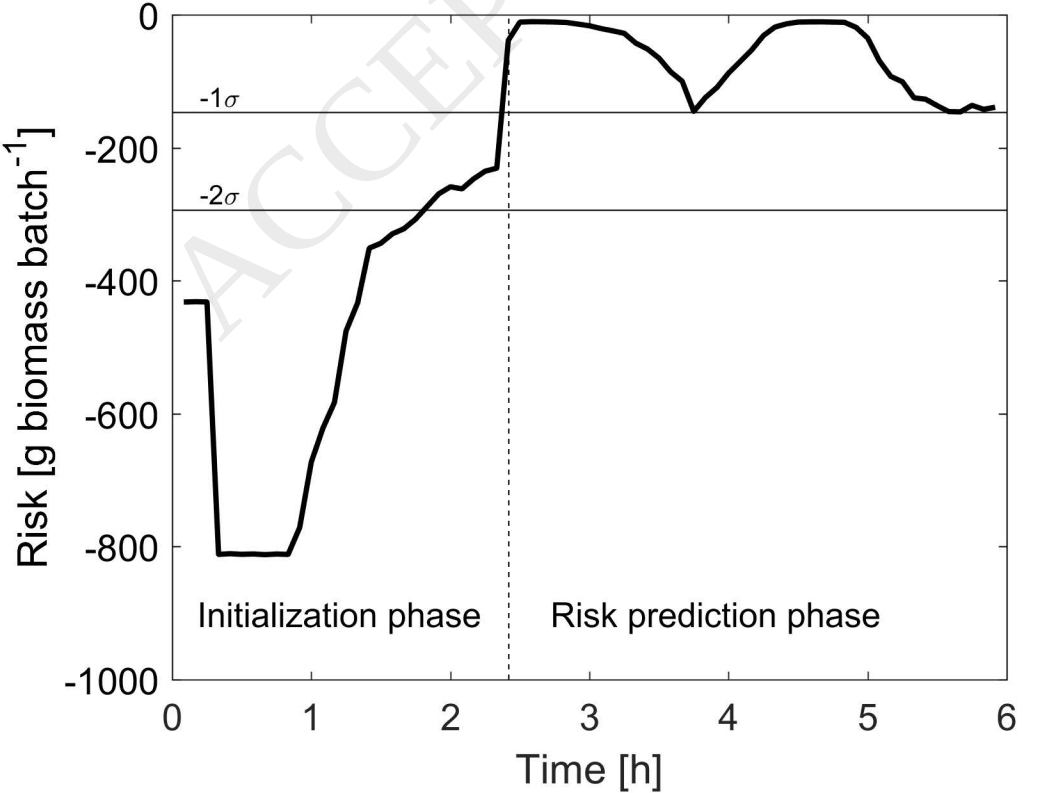
Step 7 **Save current state**

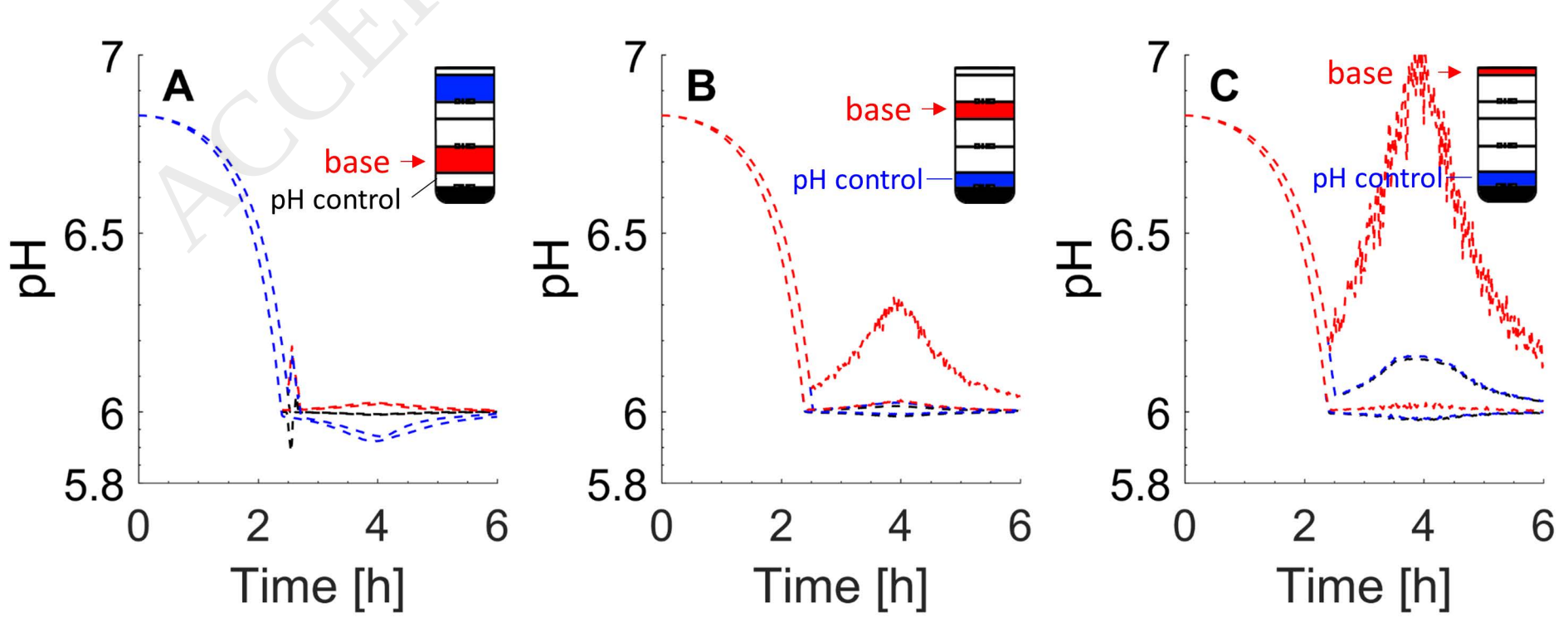





7
6.8
6.6
6.4
6.2
6
5.8
pH
0
2
4
6
Time [h]
pH control
base add.


A
Probability
0.3
0.2
0.1
0
0.07
0.08
0.09
0.1
Yield
[g biomass (g substrate)$^{-1}$]
B
Pr (Yield - Target < X)
1
0.5
0
-0.02
-0.01
0
0.01
Yield - Target
[g biomass (g substrate)$^{-1}$]
C
Pr (Prod. - Target < X)
1
0.5
0
-1000
-500
0
500
Biomass production
[g biomass batch^{-1}]


0
-200
-400
-600
-800
-1000
-1σ
-2σ
Risk [g biomass batch-1]
Initialization phase
Risk prediction phase
0
1
2
3
4
5
6
Time [h]


A
B
C
base
pH control
pH
Time [h]
7
6.5
6
5.8
0
2
4
6