Development of software sensors for on-line monitoring of baker's yeast fermentation process

Dissertation for obtaining the Doctoral Degree of Natural Science

(Dr. rer. nat.)

Faculty of Natural Science

University of Hohenheim

Institute of Food science and Biotechnology

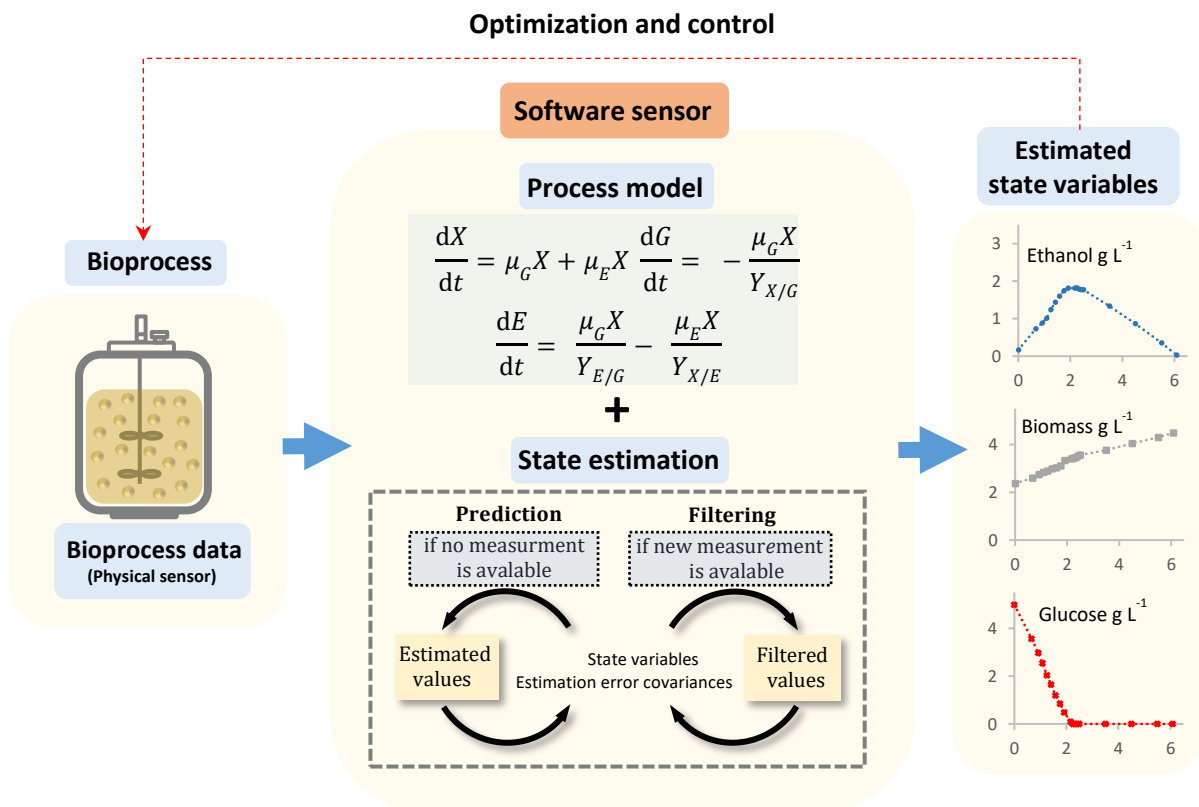Department of Process Analytics and Cereal Science

Submited by

Abdolrahim Yousefi-Darani

Born on September 15th. 1987 in Daran/Iran

Stuttgart 2021

Abdolrahm Yousefi-Darani

Development of software sensors for on-line monitoring of baker's yeast fermentation process

Contents

1. Preface

1.1. Acknowledgements

This thesis was created form September 2017 to December 2020 at the department of Process Analytics and Cereal Science of the University of Hohenheim. I want to show my gratitude to all the people who allowed me to write it.

First I want to thank my supervisor Prof. Dr. Bernd Hitzmann for his friendly and helpful assistance and support. Numerous suggestions and discussions helped me to significantly improve this work. Without his guidance and constant feedback this PhD would not have been achievable.

Apart from my supervisor, my thanks also go out to the support I received from Dr. Olivier Paquet-Durand, I am hugely appreciative for his assistance in computer programming in MATLAB and other helpful suggestions. Last but not least, my deep appreciation goes out to my actual as well as former colleagues in the department of Process Analytics and Cereal Science.

My special thanks goes to my family. I would like to thank my parents whose love and guidance are with me in whatever I pursue. Without them and their long lasting support, this thesis would not have been possible.

1.2. Co-authors

This work has been partially published previously with the knowledge and approval of the supervisor Professor Dr. Bernd Hitzmann. The scientific work presented in this thesis was partially conducted in cooperation with co-authors from the University of Hohenheim.

Application of fuzzy logic control for the dough proofing process.

By Abdolrahimahim Yousefi-Darani, Olivier Paquet-Durand, Bernd Hitzmann. Published in Food and Bioproducts Processing. Volume 115, pages 36-46, February 2019 (DOI: 10.1016/j.fbp.2019.02.006).

Olivier Paquet-Durand assisted in implementing the controller algorithm.

Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in _Saccharomyces cerevisiae_ batch cultivation.

By Abdolrahim Yousefi-Darani, Majharulislam Babor, Olivier Paquet-Durand, Bernd Hitzmann Published in Biosystems Engineering, volume 198, pages 198-209. August 2020 (DOI: 10.1016/j.biosystemseng.2020.08.004).

Majharulislam Babor carried out the experiments of yeast cultivations and determined the off-line values.

The Kalman filter for the supervision of cultivation processes.

By Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, Bernd Hitzmann. Published in Advances in Biochemical Engineering/Biotechnology (in press, DOI: 10.1007/10_2020_145).

Olivier Paquet-Durand assisted in implementing the extended Kalman filter algorithm.

Parameter and state estimation of backers yeast cultivation with a gas sensor array and unscented Kalman filter.

By Abdolrahimahim Yousefi-Darani, Olivier Paquet-Durand, Jörg Hinrichs, Bernd Hitzmann Published in Engineering in life science (in press, DOI: 10.1002/elsc.202000058).

Olivier Paquet-Durand assisted in implementing the unscented Kalman filter algorithm.

1.3. Publication lists

Peer-reviewed publications

- Abdolrahimahim Yousefi-Darani, Olivier Paquet-Durand, Bernd Hitzmann

  Application of fuzzy logic control for the dough proofing process. Food and Bioproducts Processing 115 (2019): 36-46.

- Abdolrahim Yousefi-Darani, Majharulislam Babor, Olivier Paquet-Durand, Bernd Hitzmann

  Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in *Saccharomyces cerevisiae* batch cultivation. Biosystems Engineering 198 (2020): 198-209.

- Abdolrahimahim Yousefi-Darani, Olivier Paquet-Durand, Jörg Hinrichs, Bernd Hitzmann

  Parameter and state estimation of backers yeast cultivation with a gas sensor array and unscented Kalman filter. Engineering in life science (2020).

- Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, Bernd Hitzmann

  The Kalman filter for the supervision of cultivation processes.
  Advances in Biochemical Engineering/Biotechnology (2020)

- Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, Viktoria Zettel and Bernd Hitzmann

  Closed loop control system for dough fermentation based on image processing
  Journal of Food Process Engineering, 41(5), e12801.

- Yousefi-Darani, A.R., Paquet-Durand, O., Hitzmann, B. and Zettel, V.

  Real-time automated control system for dough fermentation. Chemie Ingenieur Technik, 90(9), pp.1236-1237.

- Yousefi-Darani, A., Paquet-Durand, O. and Hitzmann, B.,

  Real-time monitoring of ethanol concentration during *Saccharomyces cerevisiae* cultivation. Chemie Ingenieur Technik, 92(9), pp.1216-1216

- Paquet-Durand, O., Zettel, V., Yousefi-Darani, A., Hitzmann, B.,

  The Supervision of Dough Fermentation Using Image Analysis Complemented by a Continuous Discrete Extended Kalman Filter. *Processes*, *8*(12), 1669.

Oral Presentations

- **ESBES - 12.09.2018**

   Closed loop control system for dough fermentation bases on image processing

- **WIG - May.2018**

   Automatisierung der Gare von Teiglingen basierend auf Bildverarbeitung und unterschiedlichen Regelalgorithmen.

Poster Presentations

- **Smart sensors - 01.10.2018**

   Application of smart sensors in controlling the proofing process in bread making

- **ISEKI Food - 04.07.2018**

   Real time automated control system for dough fermentation

- **DECHEMA - 14.09.2018**

   Closed loop control system for dough fermentation bases on image processing.

- **DECHEMA - 23.09.2019**

   Online monitoring of yeast cultivation with electronic nose

2. Summary

2. 1. Summary (English)

Software sensors and bioprocess are well-established research areas which have much to offer each other. Under the perspective of the software sensors area, bioprocess can be considered as a broad application area with a growing number of complex and challenging tasks to be dealt with, whose solutions can contribute to achieving high productivity and high-quality products.

Although throughout the past years in the field of software sensors and bioprocess, progress has been quick and with a high degree of success, there is still a lack of inexpensive and reliable sensors for on-line state and parameter estimation. Therefore, the primary objective of this research was to design an inexpensive measurement system for on-line monitoring of ethanol production during *the* backer's yeast cultivation process. The measurement system is based on commercially available metal oxide semiconductor gas sensors. From the bioreactor headspace, samples are pumped past the gas sensors array for 10 s every five minutes and the voltage changes of the sensors are measured. The signals from the gas sensor array showed a high correlation with ethanol concentration during cultivation process.

In order to predict ethanol concentrations from the data of the gas sensor array, a principal component regression (PCR) model was developed. For the calibration procedure no off-line sampling was used. Instead, a theoretical model of the process is applied to simulate the ethanol production at any given time. The simulated ethanol concentrations were used as reference data for calibrating the response of the gas sensor array. The obtained results indicate that the model-based calibrated gas sensor array is able to predict ethanol concentrations during the cultivation process with a high accuracy (root mean square error of calibration as well as the percentage error for the validation sets were below 0.2 gL^{-1} and 7 %, respectively). However the predicted values are only available every five minutes. Therefore, the following plan of the research goal was to implement an estimation method for continues prediction of ethanol as well as glucose, biomass and the growth rates. For this reason, two nonlinear extensions of the Kalman filter namely the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) were implemented separately for state and parameter estimation. Both prediction methods were validated on three different cultivation with variability of the substrate concentrations. The obtained results showed that both estimation algorithms show satisfactory results with respect to estimation of concentrations of substrates

and biomass as well as the growth rate parameters during the cultivation. However, despite the easier implementation producer of the UKF, this method shows more accurate prediction results compared to the EKF prediction method.

Another focus of this study was to design and implement an on-line monitoring and control system for the volume evaluation of dough pieces during the proofing process of bread making. For this reason, a software sensor based on image processing was designed and implemented for measuring the dough volume. The control system consists of a fuzzy logic controller which takes into account the estimated volume. The controller is designed to maintain the volume of the dough pieces similar to the volume expansion of a dough piece in standard conditions during the proofing process by manipulating the temperature of the proofing chamber. Dough pieces with different amounts of backer's yeast added in the ingredients and in different temperature starting states were prepared and proofed with the supervision of the software sensor and the fuzzy controller. The controller was evaluated by means of performance criteria and the final volume of the dough samples.

The obtained results indicate that the performance of the system is very satisfactory with respect to volume control and set point deviation of the dough pieces.

2.2. Zusammenfassung

Softwaresensoren und Bioprozese sind gut etablierte Forschungsgebiete, die sich gegenseitig viel befruchten können. Unter dem Blickwinkel der Softwaresensorik kann der Bioprozess als ein breites Anwendungsgebiet mit einer wachsenden Zahl komplexer und anspruchsvoller Aufgabenstellungen betrachtet werden, deren Lösung zur Erzielung hoher Produktivität und qualitativ hochwertiger Produkte beitragen kann. Obwohl in den letzten Jahren auf dem Gebiet der Softwaresensoren und des Bioprozesses rasch und mit großem Erfolg Untersuchung erzielt wurden, fehlt es immer noch an kostengünstigen und zuverlässigen Sensoren für die Online-Zustands- und Parameterschätzung. Daher war das primäre Ziel dieser Forschung die Entwicklung eines kostengünstigen Messsystems für die Online-Überwachung der Ethanolproduktion während des Kultivierungsprozesses von Backhefe. Das Messsystem basiert auf kommerziell erhältlichen Metalloxid-Halbleiter-Gassensoren. Die Headspace-Proben des Bioreaktors werden alle fünf Minuten für 10 s an der Gassensor-Anordnung vorbeigepumpt und die Spannungsänderungen der Sensoren werden gemessen. Die Signale des Gassensorarrays zeigten eine hohe Korrelation mit der Ethanolkonzentration während des Kultivierungsprozesses.

Um die Ethanolkonzentrationen aus den Daten des Gassensorarrays vorherzusagen, wurde ein Hauptkomponenten-Regressionsmodell (PCR) verwendet. Für das Kalibrierungsverfahren ist keine Offline-Probenahme notwendig. Stattdessen wird ein theoretisches Modell des Prozesses genutzt, um die Ethanolproduktion zu jedem beliebigen Zeitpunkt zu simulieren. Die kinetischen Parameter des Modells werden im Rahmen der Kalibration bestimmt. Die simulierten Ethanolkonzentrationen wurden als Referenzdaten für die Kalibrierung des Ansprechverhaltens des Gassensorarrays verwendet. Die erhaltenen Ergebnisse zeigen, dass das modellbasierte kalibrierte Gassensorarray in der Lage ist, die Ethanolkonzentrationen während des Kultivierungsprozesses mit hoher Genauigkeit vorherzusagen (der mittlere quadratische Fehler der Kalibrierung sowie der prozentuale Fehler für die Validierungssätze lagen unter 0,2 gL^{-1} bzw. 7 %). Die vorhergesagten Werte sind jedoch nur alle fünf Minuten verfügbar. Daher war der folgende Plan der Untersuchung die Implementierung einer Schätzmethode zur kontinuierlichen Vorhersage von Ethanol sowie von Glukose, Biomasse und der Wachstumsrate. Aus diesem Grund wurden zwei nichtlineare Erweiterungen des Kalman Filters, nämlich der erweiterte Kalman Filter (EKF) und der unscented Kalman Filter (UKF), getrennt für die Zustands und Parameterschätzung implementiert. Beide

Vorhersagemethoden wurden an drei verschiedenen Kultivierungen mit Variabilität der Start substratkonzentrationen validiert. Die erhaltenen Ergebnisse zeigen, dass beide Schätzungsalgorithmen zufriedenstellende Ergebnisse hinsichtlich der Schätzung der Konzentrationen von Substraten und Biomasse sowie der Parameter der Wachstumsrate während der Kultivierung ermitteln. Trotz der einfacheren Implementierung des UKF zeigt diese Methode jedoch genauere Vorhersageergebnisse im Vergleich zur EKF-Vorhersagemethode.

Ein weiterer Schwerpunkt dieser Untersuchung war der Entwurf und die Implementierung eines Online-Überwachungs- und Regelungssystems für die Volumenauswertung von Teigstücken während des Gärprozesses bei der Brotherstellung. Aus diesem Grund wurde ein auf Bildverarbeitung basierendes Überwachungssystem zur Messung der Teigvolumenauswertung entworfen und implementiert. Das Regelsystem besteht aus einem Fuzzy-Logic-Controller, der das gemessene Volumen für die Regelung nutzt. Die Regelung ist so ausgelegt, dass das Volumen der Teiglinge mit Werten des Volumens eines Teiglings unter Standardbedingungen während des Gärprozesses vergleicht und die Temperatur der Gärkammer entsprechend anpasst. Teiglinge mit unterschiedlichen Hefemengen in den Zutaten und verschiedenen Temperaturstartwerten wurden vorbereitet und unter Anwendung des Fuzzy-Reglers gegärt. Der Regler wurde anhand von Leistungskriterien und dem Endvolumen der Teigproben bewertet.

Die erhaltenen Ergebnisse zeigen, dass die Leistung des Systems in Bezug auf die Volumenregelung und die Sollwertabweichung der Teigstücke sehr zufriedenstellend ist.

3. Introduction and outline

3. 1. Introduction

3. 1. 1. Software sensors

Successful operation, control and optimization of biotechnological process depend on reliable real-time available measurements of the process variables. Although some hardware sensors are readily available, they often have several drawbacks: cost, sample destruction, discrete-time measurements, processing delay, sterilization, disturbances in the hydrodynamic conditions inside the bioreactor, etc. It is therefore of interest to use software sensors (Bogaerts and Wouwer 2003; Venkateswarlu 2005). The central idea behind a software sensor is to use easily accessible on-line data for the estimation of other process variables that are either difficult to measure or only measured at low frequency (Kadlec et al., 2009).

A schematic diagram demonstrating the main components of a software sensor as defined in Luttmann et al. (2012) is shown in Figure 1.
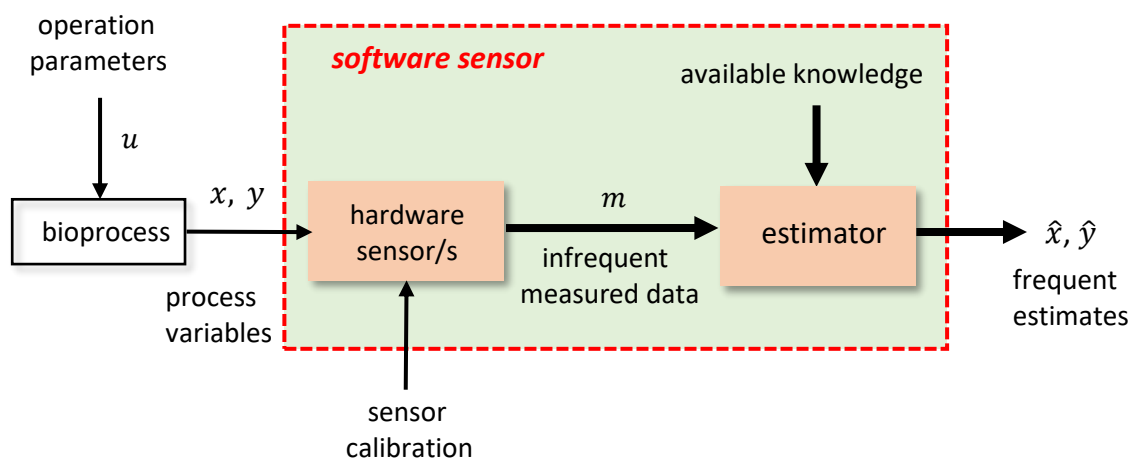
Fig. 1. *Software sensor structure*

As illustrated in Figure 1, a software sensor consists of one or more hardware (physical) sensor, knowledge of the process and an estimator. In the following parts of this sections, a brief overview of the individual elements of a software sensor are presented.

3. 1. 1. 1. Process knowledge

A critical element of a software sensor is the available knowledge. In any kind of process, knowledge is defined as the ability to describe relationships between critical process variables and critical quality or performance attributes. In the context of bioprocesses, knowledge could be expressed through mathematical models (Studer et al., 1998).

According to the paradigms that are employed, bioprocess models are classified as mechanistic, empirical (data-driven) and hybrid models. A short description of these models are as follows:

- **Mechanistic models**

In simple terms, mechanistic models are description of the biological phenomenon of a bioprocess using microscopic mass balance formulated by means of a set of ordinary differential equations, in which the rates of biochemical transformations (in the mass balance equations) are described by kinetic models. Most of these kinetic models are built upon a few classic models such as the Monod model, the Droop model, and the logistic model (Vatcheva et al., 2006). Due to their good data-fitting abilities across a wide range of bioprocesses, these models are still widely applied for monitoring, control and optimization of bioprocesses. However mechanistic models possess poor predictive capabilities for highly nonlinear and complex biological systems (del Rio-Chanona et al., 2019). This is caused by the construction principles of the kinetic models, which lump hundreds of metabolic pathways into just a few parameters which are assembled in a simple model structure. Furthermore, as metabolic pathways are strongly dependent on the cultivation conditions, any changes in these conditions also have an impact on the values of the model parameters. Therefore, the estimated parameters from one specific set of data may no longer apply to the same system operated under a different operating condition (Fouchard et al., 2009; Adesanya et al., 2014).

- **Empirical models**

Empirical models are based on direct measurement and extensive data records. Compared with kinetic models, empirical models contain more parameters and structures for data regression, enabling the inclusion of distinct process behaviours collected at different operating conditions in a single model. This allows the model to accurately interpolate the performance of untested new processes operated over a range of operating conditions (Zhang et al., 2019). The drawbacks of the empirical models is firstly they are heavily reliant on the quality and quantity of data sets. Second, most empirical models calculate the values of state variables at fixed time intervals. However, data sampled at a plant are often obtained at different time intervals, which would result in missing information for empirical model construction (Baughman and Liu, 2014).

- **Hybrid models**

Most bioprocesses deal with a high number of process variables and uncertainties. Therefore, it is difficult to describe the process only with kinetic equations. In hybrid models which are based on macroscopic mass balance equations, the biochemical transformations rates are described by means of a machine learning tool such artificial neural networks and/ or fuzzy logic algorithms. Machine learning tools would consider the influence of more process variables on the conversion rates, resulting in more reliable values.

In terms of application of the models, models are classified as dynamic and static models. Dynamic models include differential equations, typically over time or location coordinates which allow prediction. Static models are correlations which cannot provide time-dependent simulation results. Hence, they are not applicable for prediction over time or location, they are not commonly applied in bioprocess development (Kroll et al., 2017).

3. 1. 1. 2. Measurement system

- **Measurement type**

There are different ways of performing a measurement and taking samples from a bioprocess. These methods can be categorized as in-line, at-line or off-line as illustrated in Figure 2.
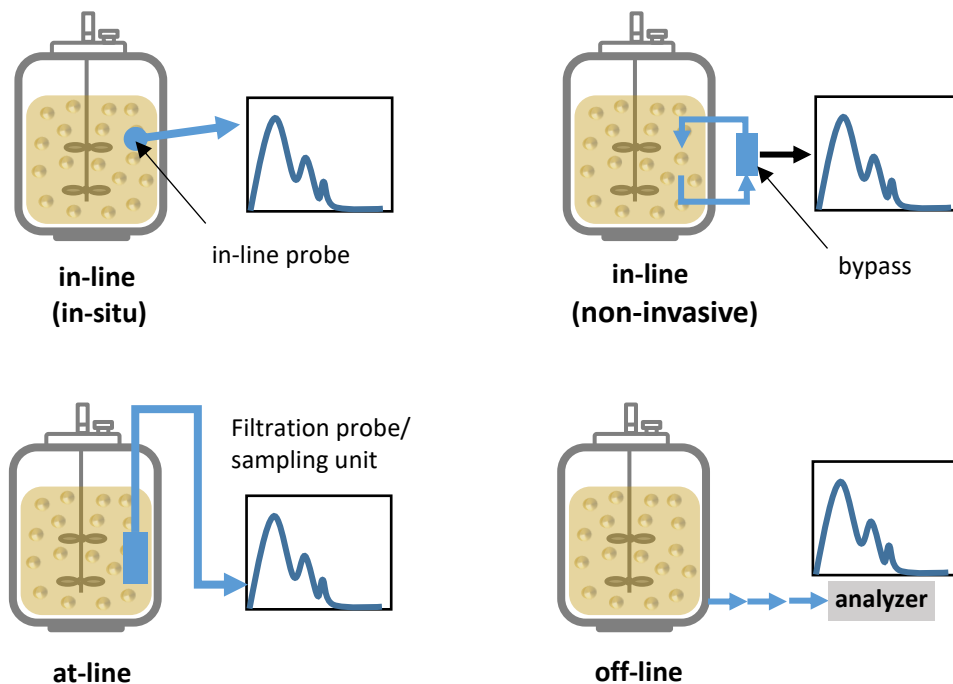
Figure 2. *Different types of sampling methods in bioprocess. Adapted from Claßen et al., 2017.*

Off-line measurements are collected manually or automatically and then transferred to the laboratory to be analysed. This causes long time delays, so that such measurements cannot contribute to the control of the dynamic process behaviour (Claßen et al., 2017; Lourenço et al., 2012).

An in-line sensor produces data continuously, and it is in direct contact with the process medium or separated from the bioreactor by a bypass (non-invasive sensor). By providing continuous information, these sensors are enablers of continuous process control.

An at-line sensors analyse samples near to the bioreactor. Samples are collected regularly (manually or automatically) and are transported to the analysis system for further analysis. Data are generated according to the analysis frequency at certain time intervals.

- **Sensor type**

Bioprocess variables can be measured in the gas, liquid, and solid phases of a bioprocess with a wide variety of measurement techniques (Fig. 3). In the gas-phase, monitoring of several gases, especially oxygen, ethanol and carbon dioxide, is important. The changes in the concentrations of these gases provide information about cell growth, metabolism, and productivity. In the liquid phase, the concentrations of various nutrients, metabolites and products, as well as dissolved gases like oxygen and carbon dioxide, must be monitored. This monitoring can be performed with in-line sensors or via the large variety of analytical systems available in chemical and biochemical laboratories. The biomass which is the most important variable in a bioprocess is a representative of the solid phase. The biomass can be characterized by its concentration or, more importantly, by its metabolic activity. Different analytic methods such as turbidity sensors, impedance sensors and NMR spectroscopy are used to determine the biomass concentration (Biechele et al., 2015).
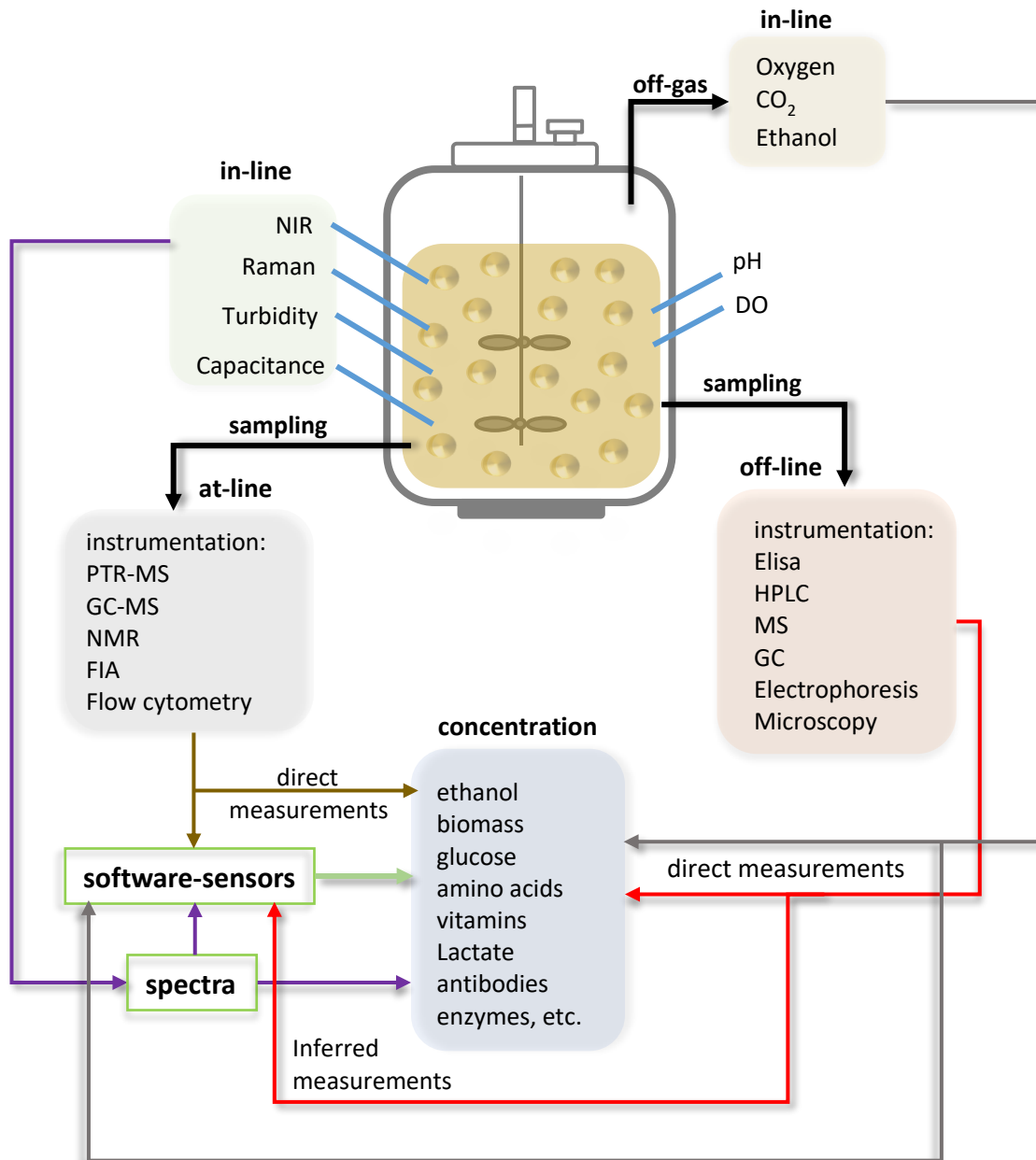
Figure 3. *Schematic of bioprocess monitoring techniques. NIR near-infrared spectroscopy, FTIR Fourier-transform infrared spectroscopy, HPLC high-performance liquid chromatography, ELISA enzyme-linked immunosorbent assay, GC gas chromatography, MS mass spectrometry, PTR-MS proton transfer reaction mass spectrometry, MALDI-TOF-MS matrix-assisted laser desorption ionization time-of-fight mass spectrometry, NMR nuclear magnetic resonance, FIA flow-injection analysis. Adapted from Veloso & Ferreira, (2017).*

3. 1. 1. 3. State estimator

The state estimator is the 'brain' of the software sensor. Many kinds of state estimation methods have been proposed and most of them have been applied in bioprocesses. Bioprocess state estimation methods can be categorized in five main classes. Each technique has its own advantages and drawbacks. This section aims to presents a brief overview of each method, which can be advantageously applied to bioprocesses.

13

- **Component balance based methods**

By these methods, estimations are carried out using theoretical and experimental derived relationships between the measured variables and the estimated values. The basic feature of this method is to represent the conversion of substrate to cell mass by reasonable assumptions regarding the regulatory structure of the organism (Zhao, 1996; Venkateswarlu, 2005).

- **Observer based methods**

Observers are used as state estimators in deterministic system (no randomness is involved in the development of estimated states of the process). As one of the basic state estimation methods, the Luenberger observer has been provided by David. G. Luenberger (Luenberger, 1964) to solve the problem of state estimation of deterministic linear systems (Wang et al., 2016). The principle of the Luenberger observer is that by combining a known measurement with knowledge of the process (process mathematical model), the process state can be estimated.

Consider the following linear discrete-time system.

$$x_{[k]} = \mathrm{A}x_{[k-1]} + \mathrm{B}u_{[k-1]} \tag{1}$$

where $x_{[k]}$ is the process state vector, $u_{[k-1]}$ is an input vector and A and B are matrices with proper dimensions. Outputs of the system are functions of the state vector $x_{[k]}$ and are represented with the vector $z_{[k]}$:

$$z_{[k]} = \mathrm{H}x_{[k]} \tag{2}$$

Where matrix H is the measurement matrix which relates the process states to the measurements.

The observer model of the system can be derived from the above equations.

Using the initial conditions of the state variables, the state $x_{[k]}$ and the outputs $z_{[k]}$ can be simulated in response to a time sequence of inputs. However, the predicted values would differ from the actual outputs of the system if the model parameters are not exactly known.

The Luenberger observer is a real-time simulator with a feedback mechanism for recursively correcting its estimated state by determining the prediction error, which reflects the discrepancy between the output of the system $z_{[k]}$ and the predicted measurements $\mathrm{H}x_{[k]}$.

The prediction error is then multiplied by the matrix **L** (observer gain) and fed back to the state equation. The state of the observer (estimated state) and the output of the state are defined with the following equations.

$$x_{O,[k]} = \mathrm{A}x_{O,[k-1]} + \mathrm{B}u_{[k]} + \mathrm{L}\left(z_{[k]} - \mathrm{H}x_{[k]}\right) \qquad (3)$$

$$z_{O,[k]} = \mathrm{H}x_{O,[k]} \qquad (4)$$

Where $x_{O,[k]}$ is the current observed state, $x_{O,[k-1]}$ is the previous observed state and $z_{O,[k]}$ is the observed outputs.

Note that the observer gain matrix (L) remains constant throughout the whole process and is selected using methods such as pole placement. Once setup correctly, the Luenberger observer is capable of generating estimates accurately.

As already mentioned, the Luenberger observer is designed and implemented in linear systems. Most bioprocesses often include nonlinearities. Therefore, the Luenberger observer cannot provide satisfactory results for such systems. Nonlinear observers are better suited for state estimation in nonlinear processes.

Nonlinear observers can be classified in two major classes namely the Luenberger-based observers and the finite-dimensional system observers.

The Luenberger-based observers involves the extended versions of the classical Luenberger observer (Dochain, 2003; Alonso et al., 2004; Tronci et al., 2005, Fissore et al., 2007, Vries et al., 2010). The extended Luenberger observer (ELO), sliding mode observer (SMO), adaptive state observer (ASO), generic and back stepping observers are examples of observers falling into this class. This type of observer is suitable for less complex linear systems with relatively simpler computational methods (Bejarano et al., 2007).

The second category is the finite-dimensional system observers, which include the reduced-order, low-order, high-gain, asymptotic and exponential observers. These finite-dimensional system observers are designed for processes whose dynamics are described by ordinary differential equations (ODEs) (Bitzer and Zeitz, 2002).

The attributes, advantages and limitations according to each method are given in Ali et al. (2015). The main difference between each method is according to their ability to take into account the measurement errors, to the necessity of using an accurate model for the reaction kinetics (Bogaerts, 1999), to the fact that they are only based on local approximations

(linearization of the nonlinear models) or not (Gauthier et al., 1992)) and to their speed of the estimation convergence which can be arbitrarily fixed or is only depending on the culture conditions (Bastin and Dochain, 1991; Bogaerts and Hanus, 2001).

- **Kalman filter based methods**

The Luenberger's state observer is strictly used for state estimation in deterministic systems. In most bioprocess systems, the process model as well as the sensor (physical) output are to some extent disturbed by noise. For such systems, Kalman filter based methods are used as state estimators (randomness is involved in the development of estimated states of the process).

The Kalman filter is used to provide optimal estimates of unmeasured states for time varying linear systems in the presence of noise (stochastic systems) by combining information from a process mathematical model with on-line process measurements.

In the Kalman filter, the process model defines the evaluation of the state from time $k-1$ to time k as:

$$\mathbf{x}_{[k]} = A\mathbf{x}_{[k-1]} + B\mathbf{u}_{[k-1]} + \mathbf{w}_{[k-1]} \tag{5}$$

Where x is the state vector, u is the process input and w is the Gaussian process noise vector that is assumed to be zero-mean with the covariance Q. Matrix A relates the state at the previous time step k-1 to the state at the current step k, matrix B relates the control input to the state variables x.

The process model is paired with the measurement model that describes the relationship between the state and the measurement at the current time step k as:

$$\mathbf{z}_{[k]} = C\mathbf{x}_{[k]} + \mathbf{v}_{[k]} \tag{6}$$

Where $\mathbf{z}_{[k]}$ is the output of the system, matrix C is the measurement matrix which relates the process state $\mathbf{x}_{[k]}$ to the measurements and v is the Gaussian measurement noise vector which is assumed to be zero-mean with the covariance R.

The Kalman filter algorithm consist of two recursive steps. The flow chart of the Kalman filter algorithm is presented in Figure 4.
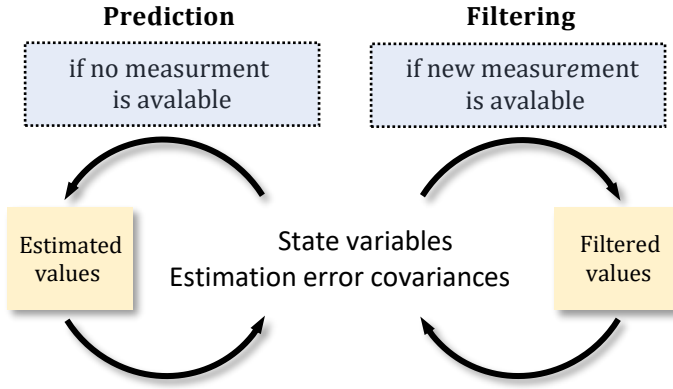
Figure 4. *Kalman filter flowchart.*

In the first step (the prediction step), the process model is used to predict the state variables and the estimation error covariance's by the following equations.

$$\mathbf{x}_{[k]} = A\mathbf{x}_{[k-1]} + B\mathbf{u}_{[k-1]} \tag{7}$$

$$P_{[k]} = AP_{[k-1]}A^{T} + Q \tag{8}$$

In the above equations $x_{[k]}$ is the state variables estimate at time *k* which is deduced from a previous estimation of the state $x_{[k-1]}$ at time *k* - 1. The new term **P** is called the state error covariance matrix which encrypts the error covariance of the predicted state values. $\mathbf{P}_{[k]}$ is the new prediction error covariance matrix at time *k* and $\mathbf{P}_{[k-1]}$ is the previous estimated error covariance matrix at time *k* - *1*. Whenever a measurement is available a correction step is performed.

In the correction step the predicted model estimates are combined with the measured values to provide corrected estimates by the following equations.

$$K_{[k]} = P_{[k]}C\left(R + CP_{[k]}C^{T}\right)^{-1} \tag{9}$$

$$P_{f,[k]} = P_{[k]}\left(1 - K_{[k]}H\right)^{2} + K^{2}R \tag{10}$$

$$x_{f,[k]} = x_{[k]} + K_{[k]}\left(z_{[k]} - Cx_{[k]}\right) \tag{11}$$

First the measurement prediction error which reflects the discrepancy between the true outputs $z_{[k]}$ and the predicted outputs $Cx_{[k]}$ is calculated and multiplied by the so called Kalman gain $K_{[k]}$ to update the estimated state variables. Therefore the filtered state variables

$x_{f,[k]}$ are obtained. In the similar manner the filtered estimation error covariance $P_{f,[k]}$ is obtained. The Kalman gain $K_{[k]}$ is chosen to minimise the estimated error covarianc and unlike the Observer gain, $K_{[k]}$ is continuously changed with each iteration to converge the states rapidly in a robust manner.

With the filtered values as initial condition, the simulation of the process as well as the estimation error covariance's can be carried out until the next measurement is obtained and everything repeats again.

- **Artificial neural network based methods**

Artificial neural networks (ANN) are powerful tools for state estimation in highly nonlinear dynamic systems. The main advantage of ANN-based methods for state estimation is that they do not need any prior knowledge about the kinetic growth rate, but the disadvantage is the large amount of data sets required for training and testing the neural network.

Neural networks are comprised of a great number of interconnected neurons (nodes). The choice of the architecture of the network depends on the task to be performed. For modelling of physical systems, a feedforward layered network is normally used which has an input layer containing multiple or single inputs, at least one hidden layer and an output layer containing multiple or single outputs. The input layer receives information from an external source and passes the information to the network for processing, the hidden layer receives information from the input layer and does all the information processing and the output layer receives the processed information from the network, and sends the result out to an external receptor (Baughman and Liu, 2014).

Each neuron is a processor which performs a weighted sum of all inputs from other neurons or from outside of the network. Predictions are obtained by passing the weighted sum through a nonlinear transfer function (sigmoid, hyperbolic tangent, sinusoid, threshold, etc.) (Thibault et al., 1990). The predicted values are then subtracted from the desired values to calculate the prediction error. The prediction error is minimized by performing iterations so that the network is optimized.

- **Fuzzy logic based methods**

Fuzzy logic systems are intelligent models based on fuzzy sets and fuzzy logic (Zadeh 1965), which have a wide range of applications, e.g., pattern recognition, intelligent control, data mining (Zhang et al., 2017). Fuzzy logic systems can also be used for state estimation by

mapping the input and output of a system, using a set of fuzzy rules and a corresponding inference mechanism.

Similar to ANN based methods, fuzzy logic based methods do not need any prior knowledge about the kinetic growth rate. However, fuzzy logic gains superiority over neural networks as system complexity increases because it allows for an easier design, system identification, and reassignment after a change. Moreover, fuzzy rules can process numerical as well as symbolic data and offers some transparency (Patnaik, 1997; Georgieva and de Azevedo, 2009).

3. 1. 2. Gas sensor array

The analysis of fermenter exhaust gas can be used to obtain important knowledge about the microbial activity. For instance, in the aerobic fermentation of baker's yeast, carbon dioxide and water is produced from glucose, and in anaerobic fermentation, ethanol is produced. Ethanol may be also produced in the aerobic fermentation if excess glucose is presented. Production of ethanol is unwanted since it may inhabit growth of microorganisms. An absence of dissolved oxygen would indicate a situation where fermentation conditions are anaerobic. Therefore, measurement of oxygen, CO_2 and ethanol concentrations are crucial in order to have an optimized processing.

Many researches have been dedicated to design and implementation of measurement devices for oxygen and carbon dioxide analyses in bioprocesses and many commercial sensors are already available. However, there is still a lack of a cheap, reliable and accurate measurement device for on-line measurement of ethanol concentration for bioprocesses.

Infrared gas analysers could be used for measuring ethanol concentrations in the gaseous phase. The infrared gas analyser measures trace gases by determining the absorption of an emitted infrared light source through the air sample. The main drawback of this method is that the outlet gas of a bioreactor contains a considerable amount of moisture, which would negatively affect the absorption of the emitted infrared light source. An alternative method is using gas sensor arrays (GSA) also known as electronic nose.

A GSA is an instrument consisting of an array of reversible but only semi-selective gas sensors coupled to a pattern recognition algorithm (Persaud and Dodd 1982).

A variety of sensor types such as conducting polymer sensors (CP), piezoelectric surface acoustic wave (SAW), thickness shear mode (TSM), metal oxide semiconductor (MOS), metal oxide semiconductor field effect transistor (MOSFET) and electrochemical (EC) sensors could

be used in a GSA. Due to the low cost and simple working principle of metal oxide gas sensors, they are the most employed type of sensors in GSA.

In MOS, the metal oxide material simply measures the change of resistance. Usually, oxygen molecules will attract free electrons in metal oxide material, forming oxygen ions. In this process, free electrons inside the materials are consumed, resulting in band bending and an electron depletion region at the materials surfaces. The target gas molecules are either acting as reducing gas or oxidizing gas, corresponding to donor or acceptor of charge carriers. When the materials are placed in the environment of these gases, the resistance of the semiconductor metal oxide will increase or decrease depending upon the type of majority carriers in the sensing materials and types of target gases (Shankar and Rayappan, 2015; Chen et al., 2019).

The performance improvement of GSA system mainly contains two kinds of optimization: sensitive material selection (sensor array optimization) and signal processing and data evaluation (Fig. 5).
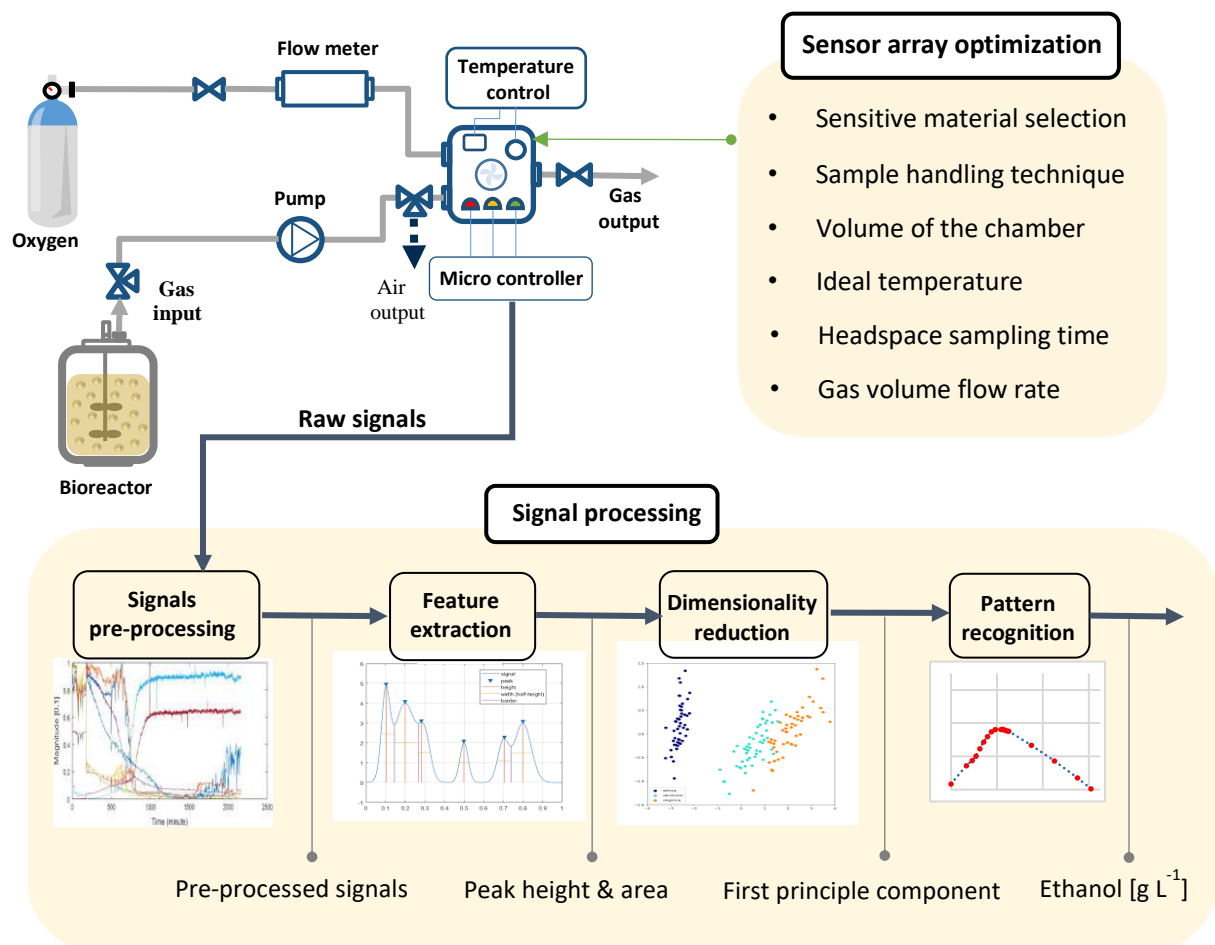
Figure 5. *Performance improvement steps of a GAS system.*

For selecting the sensors in a GSA, a typical approach is to first analyse the chemical composition of the samples and then select sensors which are responsive to those chemical groups with cross sensitivity to form the sensor array. Then from this, the sensors are tune and adjust in the array according to the responses, which is called sensitive material selection and sensor array optimization (Yan et al., 2015).

Signal processing and data evaluation is a key point of performance improvement of a GSA system. A basic signal processing and data evaluation structure of a GSA system contains the following main steps

- **Data pre-processing**

Signal pre-processing is used to prepare the data for further processing. This can involve noise reduction and smoothing the raw signals and to line arise the signals output to compensate for concentration fluctuations in the response vector by using a normalisation procedure. The functions carried out in this step should be considered carefully, as the quality of some aspects of the data will be improved while others are diminished. For example, normalization will make each set of data directly comparable; however weak signals dominated by noise will have their relative noise magnified greatly (Hines et al., 1999; Gardner and Taylor, 2009).

- **Feature extraction**

The aim of feature extraction is to extract robust information from the characteristic sensor response with less redundancy, which can represent the different "fingerprint" patterns well, to ensure the effectiveness of the subsequent pattern recognition algorithm. There are many feature extraction methods which have been used in E-nose applications. Most of these feature extraction methods can be roughly divided into three types: the first type is to extract signal features such as maximum values, integrals, differences and derivatives from the original response curves of sensors. The second type is based on curve fitting, which fits the response curves based on a specific model and extracts a set of fitting parameters as the features. The third one is based on some transforms, and very often the fast Fourier transform (FFT) and the discrete wavelet transform (DWT) (Llobet et al., 1997; Llobet et al., 2002; Carmel et al., 2003; Yan et al., 2015).

- **Dimensionality reduction**

The aim of dimensionality reduction is to find a low-dimensional mapping that preserves most of the information in the original feature vector. In the field of artificial olfaction, principle components analysis and discriminant function analysis are commonly used. These evaluation procedures present high-dimensional data sets in a manner so that there is maximum variance in a minimum number of dimensions (Musatov et al., 2010; Hartyáni et al., 2013).

- **Pattern recognition**

The response vector generated by the sensor array are then analysed using a pattern recognition technique (PRT). The PRT methods may be divided into supervised and non-supervised methods although a combination of both can be used.

In most cases there are two stages used in the pattern recognition process. First, the output of the sensor array is tagged with a descriptor and the classes are learned and grouped according to their description. Then the response from an unknown vector may be classified using the relationship found from the known vector. This process is known as supervised learning. Artificial neural network (ANN) is the best-known supervised technique.

In unsupervised learning there are no descriptors, the classes are learned based on some form of similarity measures. The most applied unsupervised technique is principal component analyses (PCA) (Scott et al., 2006; Peris and Escuder-Gilabert, 2009).

3. 1. 3. Baker's yeast fermentation

Bakers' yeast is a generic name for yeast products used primarily for leavening purposes which are strains of *Saccharomyces cerevisiae*. These organisms are single celled fungi which reproduce by budding and bear their sexual spores in an ascus. As fungi, yeasts are eukaryotic organisms distinct from more primitive prokaryotic organisms (Trivedi et al., 1986).

Baker's yeast production is one of the oldest food biotechnologies and may be considered as a "ripened technology" (Gélinas, 2012).

The objective of baker's yeast manufacturing is to harvest as fast as possible the highest amount of living cell mass at the lowest cost. First, fermentation tanks must be seeded with the strongest and purest microbial starters, otherwise unwanted microorganisms will be harvested instead. Second, such microorganisms must be fed under thoroughly controlled conditions to optimize yeast biomass and gassing power enough to raise dough, otherwise bakers will experience variations in bread volume (Gélinas, 2014). The yeast is recovered from

the final fermenter by using centrifugal action to concentrate the yeast solids. The yeast solids are subsequently filtered by a filter press or a rotary vacuum filter to concentrate the yeast further. Next, the yeast filter cake is blended in mixers with small amounts of water, emulsifiers, and cutting oils. After this, the mixed press cake is extruded and cut. The yeast cakes are then either wrapped for shipment or dried to form dry yeast (Barker and Williamson, 1992).

The fermentation of bakers' yeast can be performed in three different modes: continuous, batch and fed-batch. The continuous modes is based on the continuous addition of substrate and simultaneous removal of the products. In a batch mode no substrate is added to the initial charge, nor is any product removed, until the end of the process. In the fed-batch operation substrate is fed either through injections or continuously during fermentation and the product is removed at the end of the process (Reyman, 1992).

When *S. cerevisiae* is grown aerobically using glucose as substrate, biomass and ethanol are produced and a diauxic pattern can be observed. After depletion of glucose, ethanol is consumed by the cells (Zhang et al., 1997). Accordingly, a mathematical model that reflects the ethanol and biomass production and glucose consumption during batch cultivation of *S. cerevisiae* in an ideal stirred tank reactor can be described by the following equations (Solle et al., 2003).

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu_G X + \mu_E X \tag{10}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} = -\frac{\mu_G X}{Y_{X/G}} \tag{11}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mu_G X}{Y_{E/G}} - \frac{\mu_E X}{Y_{X/E}} \tag{12}$$

Here G, E and X are the glucose, ethanol and the biomass concentrations, respectively. μ_G and μ_E are the specific growth rates on glucose and ethanol, respectively. $Y_{X/G}$, $Y_{E/G}$ and $Y_{X/E}$ are the yield coefficients with respect to the conversion from glucose to biomass, glucose to ethanol and ethanol to biomass, respectively.

The diauxic growth can be considered in the model by the fact, that μ_G is only greater than zero if glucose is present, and then there is no growth on ethanol ($\mu_E = 0$ h^{-1}) (glucose

repression). If glucose has been consumed, then the cells have just ethanol as the only carbon source, so μ_G = 0 h^{-1} but μ_E is greater than zero until ethanol is consumed. The above assumptions can be formulated into the following equations:

$$\mu_G = \begin{cases} 0 & G = 0 \\ \mu_{G0} & G > 0 \end{cases} \tag{13}$$

$$\mu_E = \begin{cases} 0 & G > 0 \text{ or } E = 0 \\ \mu_{E0} & G = 0 \text{ and } E > 0 \end{cases} \tag{14}$$

In order to perform a simulation of the process, the parameters of the model should be determined. Typical values of the yield coefficients can be found in literature.

In a fed-batch process, substrate feed rate plays a significant role to attain a high productivity of cultivation processes. The below equations are a theoretical model of a fed-batch process (Kristensen, 2002).

$$\frac{dX}{dt} = \mu_S X - \frac{FX}{V} \tag{15}$$

$$\frac{dS}{dt} = -\frac{\mu_S X}{X/S} + \frac{F(S_F - S)}{V} \tag{16}$$

$$\frac{dV}{dt} = F \tag{17}$$

$$F = \frac{\mu_S X V}{Y_{X/S}(S_F - S_{const})} - \frac{\mu_E X}{Y_{X/E}}, \text{ if } \frac{dS}{dt} = 0 \tag{18}$$

where X is cell mass, S is substrate concentration, V is working volume and F is a substrate feed rate. μ_S is the specific growth rate on the substrate. $Y_{X/S}$ and $Y_{X/E}$ are the yield coefficient for the substrate with respect to biomass and ethanol respectively. S_F is the substrate concentration of the feed solution. S_{const} is the substrate concentration in the bioreactor at the start of a fed-batch process.

On-line monitoring and control of fed-batch cultivations have been studied and reported extensively in literature (Veloso et al., 2009; Jianlin et al., 2010; Dewasma et al., 2013; Lisci et al., 2020). On-line monitoring of batch cultivation is also crucial in order to monitor the state and if necessary change the operation conditions to achieve high productivity over the process. However, on-line monitoring and state estimation in batch cultivations remains

briefly addressed in the literature. Therefore, the focus of this study was to design and implement software sensors for the batch cultivation process of *S. cerevisiae.*

3. 1. 4. Dough fermentation

Dough is a multiphase and multicomponent system mainly composed of carbohydrates, proteins, lipids, water and air. The dough ingredients, as well as the processing conditions, determine the macroscopic structure of baked products which, in turn, is responsible for their appearance, texture, taste and stability. To build up this structure, the ingredients are mixed and kneaded, the dough leavened and baked (Autio and Laurikainen, 1997; Romano et al., 2007). Since the fluffiness and the flavour of the final product is influenced very much during leavening process, this step is one of the quality-determining steps in the production of baking goods. Leavening is done by CO_2 production which is associated with yeast activity. The produced CO_2 is penetrated and trapped in the viscoelastic texture of the dough which results in the dough volume expansion. A desirable loaf volume is achieved only if the dough provides a favourable environment for yeast growth and gas generation and, at the same time, possesses a gluten matrix capable of maximum gas retention (Sahlström et al., 2004).

The production of fermentation metabolites in dough can be affected by different factors, including availability of fermentation sugars in dough, ingredients/ recipe of the dough, dough fermentation conditions, yeast growth conditions, and the genetic makeup of the yeast strain (Struyf et al., 2017).

The control and optimization of the leavening process in the bread-making ensures final texture and quality of the backed products. In order to implement a control strategy, two major elements are required; a measurement device and a mathematical model.

For measuring the volume evaluation of dough during the fermentation process, several methods have been reported in literature. Some author's follow the evolution of the volume or height of the dough sample to characterize this phase (Kilborn et al., 1981). Some authors study the microstructure of the bread dough sample using radio-graphic methods like X-rays (Babin et al., 2006) and magnetic resonance imaging MRI (Bonny et al., 2004) or microscopic methods like scanning electron microscope SEM (Gan et al., 1995). In these methods, the fraction of the volume occupied by the gas on 2D (or 3D) images is measured to characterize the fermentation phase (Skaf et al., 2009).

For the mechanistic mathematical description of the volume evolution during the proofing process, certain assumptions are necessary: the dough has a Newtonian behavior, so the

Bernoulli equation can be used complemented by the continuity equation, during kneading just nitrogen is introduced into the dough (no CO_2), the bubbles are spherical and homogeneous distributed over the dough, no new bubbles are built during fermentation, all bubbles have the same radius, the yeast and therefore the CO_2 production is homogeneously distributed over the dough, growth of yeast is not considered, only CO_2 diffusion is considered, no coalescence and disproportionation are considered, all the gas retained in the dough and the ideal gas law can be applied. For modelling the volume evaluation of dough which is based on the work of (Stanke et al., 2014), by considering the assumptions, the increasing bubble radius over time can be described by the following equation.

$$\frac{dR}{dt} = \frac{3nR_GT}{16\pi R^2\mu} - \frac{P_DR}{4\mu} - \frac{\gamma}{2\,\mu} \tag{19}$$

Where R is the bubble radius, n is the amount of CO_2 in moles in the bubble, R_G the universal gas constant, T the temperature, P_D the pressure in the liquid dough, γ the surface tension and μ is viscosity. Because of the concentration difference of CO_2 in dough and bubble, a transport caused by diffusion of CO_2 into the bubble takes place, which is proportional to the concentration difference. The change of the number of moles of CO_2 in the bubble with time is described by Eq. (20)

$$\frac{dn}{dt} = 4DR\pi(C_D - C^*) \tag{20}$$

D is the diffusion coefficient, C^* the carbon dioxide concentration which is in equilibrium with its partial pressure at the bubble surface and C_D the carbon dioxide concentration in the dough. CO_2 concentration in dough increases with the production by yeast cells and decreases with the diffusion into all bubbles. Thus, change of CO_2 concentration in dough is described by Eq. (21)

$$\frac{dC_D}{dt} = Q_{CO_2} - 4N_bDR\pi(C_D - C^*) \tag{21}$$

Q_{CO_2} is carbon dioxide production rate and N_b is the number of bubbles per volume unit.

The concentration of C^* and P_B are followed from Henrys law which is described in Eq. (22) and (23).

$$C^* = \frac{P_B}{H} \tag{22}$$

$$P_B = \frac{n - n_0}{n} - P_{total} \tag{23}$$

The temperature dependency of carbon dioxide production rate is described by Arrhenius equation which is presented in Eq. (24).

$$Q_{CO_2} = A e^{-E_a/(R_G T)} \tag{24}$$

Where A is the pre-exponential factor (constant for each chemical reaction and E_a is the activation energy for the reaction.

3. 2. Outline

Commercialization of bioprocesses will require a continued emphasis on bioprocess monitoring and control. In the past, the application of process monitoring and control to bioprocesses has been limited by the availability of suitable sensors. New developments have combined enhanced computational power and a broad set of data analytical techniques with molecular biology to address some of these limitations (Reilly and Charles, 1990). Accordingly, the focus of this study was in two-fold; the first objective was to design and implement a software sensor for on-line monitoring of baker's yeast batch cultivation process and secondly to design a monitoring and control system for the supervision of the fermentation (proofing) process in bread making.

3. 2. 1. Baker's yeast cultivation

Up to now, there is still a lack of inexpensive and robust commercially available sensor that allow real-time monitoring of important variables in the baker's yeast cultivation process. Therefore, in the firs publication '' **Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in *Saccharomyces cerevisiae batch cultivation*''** the main goal was to design an inexpensive measurement system which is capable of measuring at least one process variable during the cultivation process. For this reason, a gas sensor array as well as a bioreactor head space sampling system was designed to analyse the exhaust gas of the bioreactor. The signals from the gas sensor array showed a high correlation toward ethanol concentration during cultivations, therefore the signals were used for predicting

ethanol concentration. For the calibration procedure no off-line samples were used. Instead, a theoretical model of the process was applied to simulate the ethanol production at any given time.

With the proposed method, it was possible to predict ethanol concentration with a high accuracy. However, ethanol was predicted only every five minutes (due to the headspace sampling producer) and no information from other process variables such as biomass and glucose concentration were obtained. Therefore, in the second publication **"The Kalman filter for the supervision of cultivation processes"** an extended Kalman filter (EKF) based software sensor was designed and implemented for continues monitoring of ethanol concentration during cultivations of baker's yeast. The software sensor employs the discrete ethanol measurements from the gas sensor array and predicts continues ethanol, glucose and biomass concentrations. Furthermore, the specific growth rates and their maximum values were also predicated with the proposed method.

Usually the EKF shows good prediction results. Nevertheless, in spite of the satisfactory results, it has some disadvantages. It is reliable for systems which are almost linear on the time scale of the update intervals; it requires the calculation of Jacobians at each time step, which may be difficult to obtain for higher order systems; it does linear approximations of the system at a given time instant, which may introduce errors in the estimation, leading to a state divergence over time (Julier and Uhlmann, 1997). Therefore, in the third publication **"Parameter and state estimation of backers yeast cultivation with a gas sensor array and unscented Kalman filter"** an unscented Kalman filter (UKF) was used for predicting the process variables.

The UKF is another nonlinear extension of the Kalman filter which is very similar to the EKF, but instead of approximating the non-linear process model by calculating the Jacobian of the dynamics for the determination of the estimation error variance, the transformed probability distributions are approximated directly. This is done by representing the distribution by a set of chosen sample points (sigma points), transforming these points by the non-linear model function, and then approximating the mean and variance of the transformed distribution, by the mean and variance of the transformed points (Julier et al., 2000).

The obtained results show that with the proposed UKF algorithms, it is possible to estimate the maximal specific growth rates as well as continuous ethanol, glucose and biomass concentrations with a better accuracy compared to the EKF based method.

3. 2. 2. Dough fermentation process

The fermentation (proofing) process of dough is one of the quality-determining steps in the production of baking goods. In order to ensure final product quality, robust measurement system and control strategies are required. Therefore, for the dough fermentation process, in the publication ''**Application of fuzzy logic control for the dough proofing process''** a measurement system (software sensor) based on image analyses was developed for measuring the volume evaluation of dough pieces and a fuzzy logic controller was designed to maintain the volume of the dough pieces similar to volume expansion of a dough piece under standard conditions. The fuzzy controller uses the measured volume from the imaging system and comperes it to a reference value (volume of a dough piece in standard conditions). The fuzzy controller manipulates the temperature and humidity of the proofing chamber according to the difference in the measured volume. The obtained results indicate that the performance of the system is very satisfactory with respect to volume control.

4. Publications

3. 1. Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in *Saccharomyces cerevisiae* batch cultivation.

By Abdolrahim Yousefi-Darani, Majharulislam Babor, Olivier Paquet-Durand, Bernd Hitzmann

Published in Biosystems Engineering, volume 198, pages 198-209. August 2020.

Research Paper

Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in *Saccharomyces cerevisiae* batch cultivation

Abdolrahim Yousefi-Darani[*], Majharulislam Babor, Olivier Paquet-Durand, Bernd Hitzmann

Department of Process Analytics and Cereal Science, University of Hohenheim, Stuttgart, Germany

The ethanol concentration in batch cultivation with the yeast *Saccharomyces cerevisiae* was predicted on-line using a gas sensor array. Headspace samples were pumped past the gas sensors array for 10 s every five minutes and the voltage changes of the sensors were measured. For the calibration procedure no off-line sampling was used. Instead, a theoretical model of the process has been applied to simulate the ethanol production at any given time. However, the kinetic parameters of the simulation model are unknown at the beginning of the calibration. It is demonstrated that these kinetic parameters of the theoretical process model can be acquired from the response of the gas sensor array alone. The calculated parameters result in a simulation model that is at least as accurate as a model whose parameters have been acquired by least squares fitting to off-line measurements. The root mean square error of calibration as well as the percentage error for validation sets were below 0.2 g L^{-1} and 7%, respectively.

The obtained results indicate that the model-based calibrated gas sensor array can be a cheap alternative to other tools, such as spectroscopy-based methods, that are used for monitoring yeast cultivations.

1. Introduction

Microorganisms are industrially used to produce either generic biomass or specific substances such as enzymes, amino acids or antibiotics. When biomass production is performed, usually a high yield is targeted. For *Saccharomyces cerevisiae* (*S. cerevisiae*) this is achieved by purely oxidative metabolism (Madigan et al., 2017). *S. cerevisiae* can use glucose as the energy and carbon source for aerobic respiration. When a critical glucose concentration is exceeded, this microorganism produces ethanol. This is known as the Crabtree effect and results in a lower biomass yield which might, depending on the goal of the cultivation, be undesirable (De Deken, 1966). Therefore, in order to achieve high productivity of biomass, continuous real-time monitoring of ethanol concentration is required.

Bioprocess variables are of a chemical, physical, or biological nature and can be measured in the gas, liquid, and on

solid phases of a bioprocess. On-line measurements of these variables make great demands on the sensing device. It is easier to meet these demands in the gas phase environment than in the liquid phase for various reasons: in the gas phase, the number of interfering substances is smaller and the mechanical stress on the sensor membrane is lower than in the stirred bioreactor liquid. In addition, prevention of so-called sensor fouling by cell adhesion is not necessary and a sterile barrier in the form of a mass filter can be easily introduced into the gas stream (Wild et al., 1996).

Numerous methods have attempted to measure the concentration of volatile organic compounds (VOCs) in the vapour phase. In recent years, particularly due to recent technological developments in sensor technology and computing power, gas sensor arrays (electronic nose techniques) have become valuable tools for VOC measurements. Generally, the sensor array technique is attractive for a number of significant features, such as the relatively fast assessment of headspace, a quantitative representation or qualitative identification of a gas and using cheap chemical sensors which can be easily integrated in current production processes. It has thus become particularly suitable for the continuous monitoring of microbial fermentation processes (Jiang et al., 2015). Recent applications of gas sensor arrays for monitoring fermentation process have been reported in the literature (Buratti & Benedetti, 2016; Ghosh et al., 2017; Hidayat et al., 2018; Li et al., 2019; Tan et al., 2018, 2019). However, only a few works have demonstrated the application of gas sensor arrays for monitoring ethanol concentration during S. cerevisiae cultivation (Bachinger & Mandenius, 2001; Lidén et al., 1998; Mandenius et al., 1997). In order to predict a specific volatile compound with a gas sensor array, chemometric modelling techniques are required. In the previous studies, the calibration methods for the chemometric models were limited to data-driven calibration methods. The main disadvantage of data-driven calibration methods is the huge amount of off-line data necessary to calculate a reliable model.

An alternative to the data-driven calibration method, which is a time consuming task, is model-based calibration (MBC). A statistical model-based approach for developing calibration models does not require the time-expensive collection of samples for off-line measurements. Furthermore, this approach addresses some of the shortcomings of traditional calibration methods to study the entire system response which results in robust calibration. Lin et al. (2007) gave a systematic approach for development of data-driven soft sensors. MBC approaches have been implemented on spectroscopy-based monitoring systems. Solle et al. (2003), as well as Paquet-Durand et al. (2017a), used this evaluation technique for the prediction of biomass, glucose, and ethanol during a S. cerevisiae cultivation. Furthermore, Paquet-Durand et al. (2017b) applied this method for evaluation of fluorescence measurements during several parallel cultivations of H. polymorpha in a microtitre plate.

Based on fluorescence measurements, Ödman et al. (2009) and Solle et al. (2003) evaluated yeast cultivation using glucose as substrate and developed chemometric models, one for the glucose consumption phase with concomitant ethanol production and a separate one for the ethanol consumption phase (after glucose depletion). They stated that it was difficult to use one and the same model for both phases. Paquet-Durand et al. (2017a) examined artificial neural networks for the correlation of the fluorescence spectra with glucose, biomass and ethanol concentrations. They implemented a model-based training approach for training one single neural network for the whole process. They reported an accurate prediction of glucose and biomass (error of prediction below 5%) though the prediction error for ethanol was 10%. This is due to ethanol not being fluorescent so it could only be determined indirectly from the spectra. Therefore, fluorescence-based monitoring methods are not the most accurate methods for predicting ethanol concentrations during S. cerevisiae cultivation process. In this contribution, ethanol concentration during yeast cultivation was predicted using a gas sensor array and chemometric modelling. The main contribution of this paper can be summarised as follow:

- Design and implementation of a gas sensor array and headspace sampling system in order to achieve accurate prediction of ethanol concentration in the liquid phase during S. cerevisiae batch cultivation.
- Implementation of a MBC algorithm for the calibration of the gas sensor array. Instead of using off-line measurements, simulated process variables were used to determine parameters of the chemometric model. The kinetic parameters of the process model are unknown at the beginning and are also determined during this procedure.

The results of the proposed calibration method are compared with a classical calibration method (CCM) in which the parameters of the model are acquired by least squares fitting to off-line measurements.

The remaining paper is organised as follows. Section 2 provides the materials and methods which were applied in this study. Section 3 provides the results and Section 4 concludes this paper.

2. Material and methods

2.1. Strain and cultivation conditions

Three batch fermentations of S. cerevisiae, named BC1, BC2, and BC3, were performed. Cultivations were carried out in a 3 L stainless steel tank bioreactor (Minifors, Infors HT, Bottmingen, Switzerland) with a working volume of 1.35 L and equipped with temperature and pH control. 5 g of baker's yeast S. cerevisiae (fresh backer's yeast, Omas's Ur-Hefe) was used in all pre-cultivations. The baker's yeast was suspended into 100 mL Schatzmann medium containing 0.34 g L^{-1} $MgSO_4 \cdot 7H_2O$, 0.42 g L^{-1} $CaCl_2 \cdot 2H_2O$, 4.5 g L^{-1} $(NH_4)_2SO_4$, 1.9 g L^{-1} $(NH_4)_2HPO_4$, 0.9 g L^{-1} KCl. After shaking for 10 min they were added into the bioreactor. The medium used for batch cultivations was the same as for the suspended cells, but with 5 g L^{-1}, 7 g L^{-1} and 9 g L^{-1} glucose for BC1, BC2, and BC3 respectively and 1 mL L^{-1} trace elements solution (0.015 g L^{-1} $FeCl_3 \cdot 6H_2O$, 9 mg L^{-1} $ZnSO_4 \cdot 7H_2O$, 10.5 mg L^{-1} $MnSO_4 \cdot 2H_2O$, and 2.4 mg L^{-1} $CuSO_4$ $5H_2O$) and 1 mL L^{-1} vitamin solution (0.06 g L^{-1} myoinositol, 0.03 g L^{-1} Ca-

pantothenate, 6 mg L^{-1} thiamine HCl, 1.5 mg L^{-1} pyridoxine HCl, and 0.03 mg L^{-1} biotin). All three batch runs were operated under the same conditions, that is, a constant temperature of 30 °C and pH maintained at 5. The aeration and agitation rates were kept constant at 3.5 L min^{-1} and 450 rpm, respectively.

2.2.   Off-line analysis

Samples were regularly taken from the bioreactor and put into pre-weighed and pre-dried micro centrifuge tubes. The samples without the supernatant were left in a drying oven at 103 °C for 24 h. They were then cooled down for 30 min before weighing. The supernatant of the samples after centrifugation was examined by HPLC (ProStar, Variant, Walnut Creek, CA, USA) to determine the concentration of ethanol. The supernatant was firstly filtered through a 0.45 μm pore size polypropylene membrane filter (VWR, Darmstadt, Germany). Subsequently, 20 μL was injected into a Rezex ROA-organic acid H+ (8%) column (Phenomenex, Aschaffenburg, Germany) operated at 70 °C with 5 mM H$_2$SO$_4$ as an eluent at 0.6 mL min^{-1} flow rate. The concentration of ethanol was calculated by Galaxie™ Chromatography software (Varian, Walnut Creek, CA, USA).

2.3.   Gas sensor array and on-line sampling

In this study, a gas sensor array was developed and implemented for real-time monitoring of ethanol concentration during yeast cultivation. The developed system is built in three parts, namely measurement chamber, electronics and mechanics.

The measurement chamber includes commercially available metal oxide semiconductor (MOS) gas sensors. Selecting the proper sensors is always challenging in volatile compound measurement with gas sensor arrays.

During S. cerevisiae cultivation, ethanol is the main volatile compound produced. Whenever ethanol is present, chemical sensors cannot distinguish other volatile compounds from the much higher ethanol background. Therefore, in this contribution only MOS gas sensors with high sensitivity to ethanol (according to the manufacturer's instructions) were used (TGS 822, TGS 813 and MQ3). The MOS gas sensors were placed inside a chamber with a volume of 250 mL to measure the ethanol concentration of the incoming gas. A circulation fan was placed inside the chamber for homogeneous distribution of the gas. In order to eliminate the drift effect of sensor signal which may arise from temperature variation under long time measurements, the temperature of the measurement chamber was kept constant at 42 °C. For this reason, a temperature sensor (DHT22, Aosong Electronics Co., Ltd) as well as two heating elements were placed inside the measurement chamber. The temperature of the chamber was controlled with a closed loop temperature control system.

The electronics part covers the sensor circuit, control circuit, micro controllers and a power supply circuit for generating required different voltages for sensors (5 V) and valves (12 V). A micro controller (Arduino Nano) containing a 10-bit ADC (analogue to digital convertor) was used to convert the electrical signals from the MOS sensors and the temperature

sensor to digital signals. The digital signals were sent to another micro controller (Arduino mega 2560) via I2C communication protocol. Data from the Arduino mega was sent to a computer (Intel Core i3, 2933 MHz, 4 GB RAM) via serial port communication for further signal processing and data extraction. The mechanical part consists of Teflon tubing, pump (Schwarzer Precision, Essen, Germany) and solenoid vales. Figure 1 gives a schematic diagram of the measurement system.

The bioreactor headspace sampling procedure consisted of an automated sequence of internal operations which was performed every 5 min during the cultivation process. Headspace sampling contained the following main stages: exposure stage, purging stage and sensor regeneration stage. In the exposure stage, the headspace of the bioreactor was passed into the sensor chamber for 10 s at a flow rate of 400 mL min^{-1} with a diaphragm pump (Schwarzer Precision, Essen, Germany). The exposure stage was followed by the purging stage. In this phase, clean air was drawn into the sampling pipe (the pipe connected from the bioreactor to the measurement chamber) to clear the remaining gas from the previous measurement. Simultaneously the measurement chamber was flushed out with a stream of oxygen at a flow rate of 450 mL min^{-1} for 180 s. These values as well as the flow rate of the pump were all determined by technical conditions of our measurement set-up and practical considerations. After the flushing step was over, the input and output valves of the sensor chamber was kept closed for 110 s so that the readouts from sensors reached the same level as before measurement (sensor regeneration stage). The sampling procedure was controlled via a set of miniature solenoid valves interfaced by a micro controller (Arduino Nano).

2.4.   Signal pre-processing and feature extraction

Analogue data obtained from the output voltage responses of the sensors were sampled at 2 s^{-1} frequency during the whole cultivation process. Figure 2 shows a typical sensor circuit and its interface diagram.

In Fig. 2, V_h is a fixed voltage for the heater of the sensor (5 V), V_C is the upper reference voltage (5 V), R_L is the load resistor and R_s is the sensor resistance. With pure air R_s is high. With the presence of detectable gases, R_s changes with the variation of gas concentration. V_C is a fixed voltage (5 V). By measuring the voltage on the resistor R_L, the sensor response (V_s) can be calculated by the following equation:

$$V_s = \frac{R_L}{R_L + R_s} \times V_C \tag{1}$$

Each sensor reacts differently to volatiles in the headspace gas of the bioreactor. The output voltage is related to the ethanol concentration but it doesn't directly include the concentration levels. However, it is known that if the concentration level changes, the output voltage responses of the sensors change also (Kiani et al., 2016; Omatu & Yano, 2016).

For ethanol prediction, the sensor response should first be further pre-processed to obtain comprehensible signals. Second, some features from the pre-processed signals should be extracted. In the next step, dimensionality reduction should be performed on the extracted features. Dimensionality
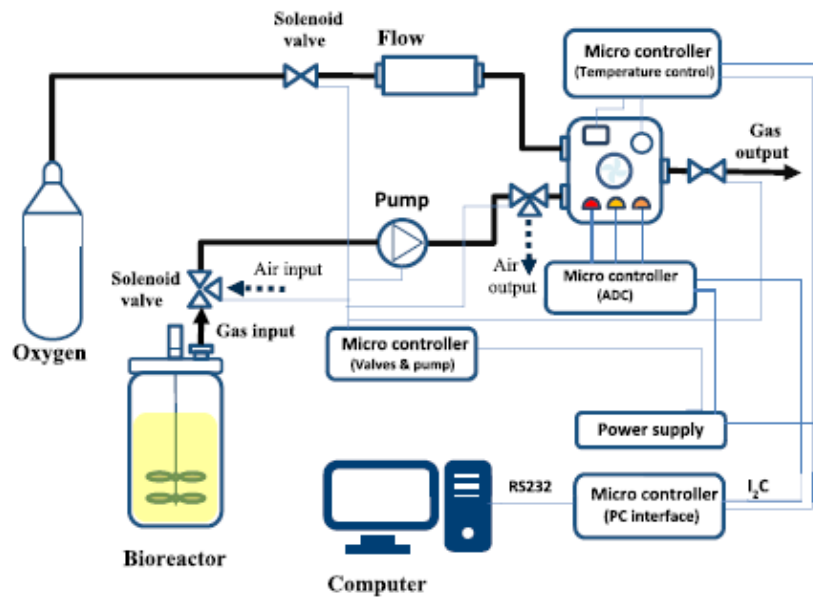
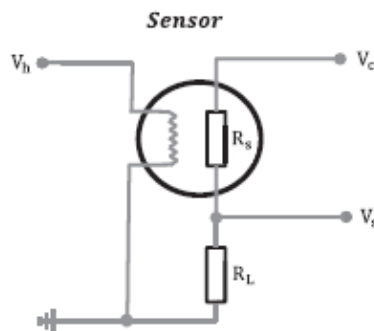Fig. 1 — Schematic diagram of the measurement system.

Fig. 2 — Basic gas sensor interface circuit diagram.

reduction projects the feature vector onto a lower dimensional space in order to avoid problems associated with high-dimensional, sparse datasets and redundancy (Aguilera et al., 2012). Finally, the reduced features should be imported into the prediction model. Figure 3 shows the block diagram of the proposed ethanol prediction algorithm.

One of the simplest methods of signal pre-processing, which is also widely used for drift compensation, is the transformation of individual sensor signals based on the initial value of the sensor response. This process compensates for noise, drift and also for inherently large or small signals (Di Carlo & Falasconi, 2012). For this reason the following equation was applied.

$$S_i = \frac{V_{si} - V_{s0,i}}{V_{s0,i}} \qquad (2)$$

where V_{si} is the response (the voltage of the sensor) of the ith sensor, $V_{s0,i}$ is its baseline and S_i corresponds to the modified signal.

The next step in the ethanol prediction algorithm is extracting useful features from the output signals. In each measurement cycle, the sensors are exposed to the headspace gas of the bioreactor for 10 s, which causes changes in the output signals. In the next step the odorant is flushed out of the sensor using the oxygen gas and the sensor returns back to its baseline. The time during which the sensor is exposed to the odorant is referred to as the transient phase, while the time it takes the sensor to return to its baseline resistance is called the recovery phase. In order to exploit the obtained information in the transient phase, two representative features were extracted from each sensor:

- Peak height: calculated as the difference between the maximal value and the baseline value of sensor response in the transient phase.
- Peak area: calculated as the area under the signal response in the transient phase.

In total, each measurement cycle was characterised by 6 variables (i.e., 3 sensors × 2 features per sensor). In order to quantify the amount of useful information for predicting ethanol concentration from all the variables, principal component analyses (PCA) was performed. The process of PCA is to find a new coordinate system of the mean centred data set, whose axes are perpendicular and have maximal variance
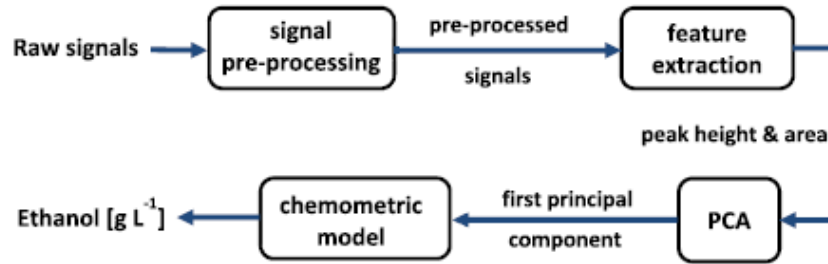
Fig. 3 – Block diagram of the proposed ethanol prediction algorithm.

in decreasing order. The direction containing most of the variance of the data is called the first PC. The second PC carries the maximum variance of the remaining data and so on. These PCs are statistically unrelated to each other (AS, 2006; Otto, 1999).

2.5. Theoretical model

When *S. cerevisiae* is grown aerobically using glucose as substrate, biomass and ethanol are produced and a diauxic pattern can be observed. After depletion of glucose, ethanol is consumed by the cells (Zhang et al., 1997). Accordingly, we have used a mathematical model that reflects the ethanol and biomass production and glucose consumption during batch cultivation of *S. cerevisiae*. The model was formulated by Solle et al. (2003) and has been reported in literature previously (Assawajaruwan, Eckard, et al., 2017; Assawajaruwan, Reinalter, et al., 2017; Paquet-Durand et al., 2017a). This mathematical model is based on the following assumptions: the cultivation of the yeast cells is performed in an ideal stirred tank reactor, there is no concentration gradient in the reactor, the change of physicochemical properties of the cells does not change the growth rates and yield coefficients except for the main substrates (glucose and ethanol) and the growth rates are the same for all cells and no limitation in growth can be observed (change of the main substrate from glucose to ethanol is very fast).

The mathematical model can be described by the following differential equations:

$$\frac{dX}{dt} = \mu_G X + \mu_E X \tag{3}$$

$$\frac{dG}{dt} = -\frac{\mu_G X}{Y_{X/G}} \tag{4}$$

$$\frac{dE}{dt} = \frac{\mu_G X}{Y_{E/G}} - \frac{\mu_E X}{Y_{X/E}} \tag{5}$$

Here G, E and X are the glucose, ethanol and the biomass concentrations, respectively. μ_G and μ_E are the specific growth rates on glucose and ethanol, respectively. $Y_{X/G}$, $Y_{E/G}$ and $Y_{X/E}$ are the yield coefficients with respect to the conversion from glucose to biomass, glucose to ethanol and ethanol to biomass, respectively.

The diauxic growth can be considered in the model by the fact that μ_G is only greater than zero if glucose is present, and

then there is no growth on ethanol ($\mu_E = 0$ h^{-1}) (glucose repression). If glucose has been consumed, then the cells have just ethanol as the only carbon source, so $\mu_G = 0$ h^{-1} but μ_E is greater than zero until ethanol is consumed. The above assumptions can be formulated into the following equations:

$$\mu_G = \begin{cases} 0 & G = 0 \\ \mu_{G0} & G > 0 \end{cases} \tag{6}$$

$$\mu_E = \begin{cases} 0 & G > 0 \text{ or } E = 0 \\ \mu_{E0} & G = 0 \text{ and } E > 0 \end{cases} \tag{7}$$

In order to perform a simulation of the process, the parameters of the model should be determined. Typical values of the yield coefficients can be found in the literature (Ödman et al., 2009). In this contribution, the yield coefficients have been fixed to following values, which are similar to those used by Paquet-Durand et al. (2017a):

$$Y_{X/G} = 0.175 \text{ g}_X \text{ g}_G^{-1}, \ Y_{E/G} = 0.473 \text{ g}_E \text{ g}_G^{-1} \text{ and } Y_{X/E} = 0.598 \text{ g}_X \text{ g}_E^{-1}$$

The values for the specific growth rates (μ_{G0} and μ_{E0}) are determined during the MBC procedure by numerical integration of the differential equations using a Runge–Kutta-method.

2.6. Model-based calibration procedure

In order to predict ethanol concentrations from the data of the gas sensor array, the following principal component regression (PCR) model was applied (the chemometric model).

$$c_E = p_0 + (p_1 \times PC_1) + (p_2 \times PC_1^2) \tag{8}$$

where c_E is the predicted ethanol concentration, PC_1 is the first principal component of the gas sensor array data and p_0, p_1 and p_2 are the parameters of the chemometric model.

The simulated ethanol concentrations calculated from the process model were used as reference data for calibrating the response of the gas sensor array. In order to calculate the simulated ethanol concentrations, the values of the specific growth rates were required. For obtaining these values the following procedure was applied:

During the first step, roughly estimated starting values of the specific growth rates (μ_{G0} and μ_{E0}) are used and the simulated ethanol concentration is calculated. During the calibration procedure, the evaluation of the simulated ethanol concentration is compared with the predicted ethanol

concentration and the sum of squared differences is calculated. In the next step, the error of prediction is minimised by implementing an optimisation algorithm. The algorithm changes the process model parameters (μ_{G0} and μ_{E0}) as well as the parameters of the chemometric model (p_0, p_1 and p_2). All the steps are processed in a cycle until the minimum of the sum of squared differences is obtained. The flowchart of the MBC procedure is presented in Fig. 4.

The optimisation method which was used to minimise the error of prediction is a particle swarm optimisation algorithm. This algorithm works by improving a population of candidate solutions called particles, which are the parameters of the mathematical models (here the specific growth rates as well as the parameters of the chemometric model). The particles are flying through the search space and the velocity of each particle is determined by the position of its best-known performance as well as the position of the overall swarm's best known performance. The swarm iteratively moves to the best solution. A more detailed description can be found in the literature (e.g. Wang et al., 2014).
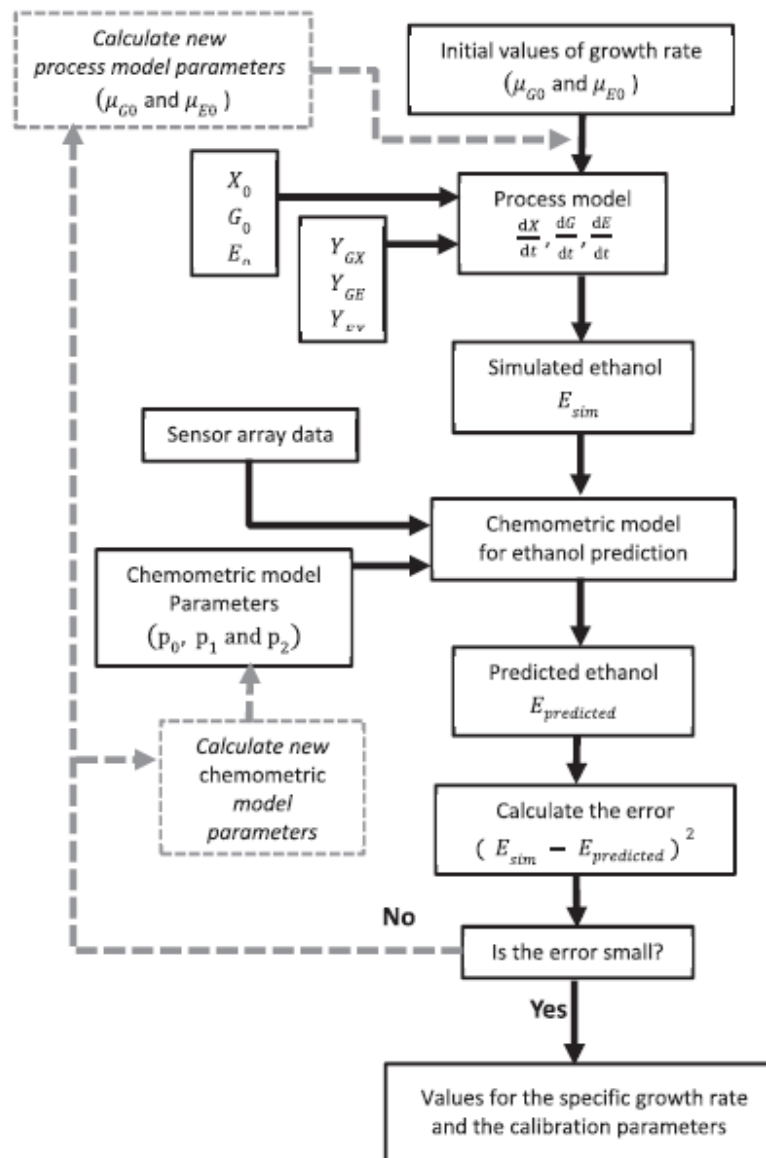

Fig. 4 – Flowchart of the model-based calibration procedure to get optimal parameters for the process model as well as the calibration model.
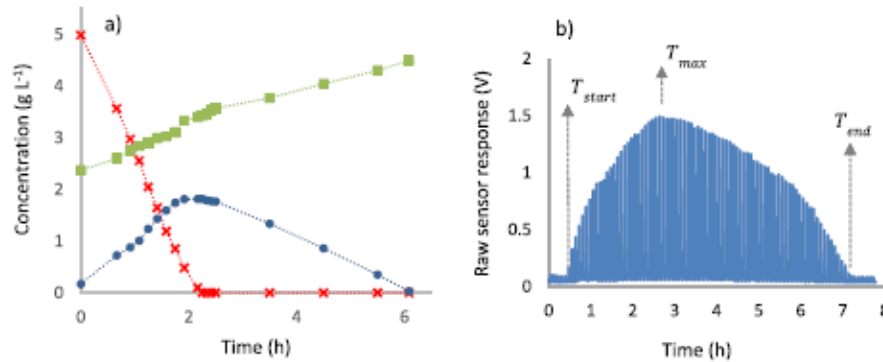
Fig. 5 — (a) Evaluation of fermentation variables (ethanol production (●), glucose consumption (×) and biomass production (■)) during yeast batch cultivation BC2 and (b) the raw signal response of one sensor (TGS 822) to the same cultivation.

By applying this MBC method, appropriate values for the parameters of the theoretical process model (μ_{G0} and μ_{E0}) can be estimated. Furthermore, the optimal parameters of the calibration model are calculated and are used for predicting ethanol concentration.

2.7. Calibration using off-line values

Ethanol concentration was also predicted by a classical calibration approach. The same chemometric model was used for this approach. The chemometric model was calibrated using the off-line ethanol concentrations.

For both calibration methods, data from all three cultivations were carried out separately for the calibration models (with each calibration method, three different calibration models were made). Each calibration model was evaluated individually with the data from the other two cultivations.

The calibration models were characterised by the root-mean-squared error of calibration (RMSEC), standard error of calibration (SEC) with respect to the maximum ethanol concentration, root-mean-squared error of prediction (RMSEP) and standard error of prediction (SEP) with respect to the maximum ethanol concentration.

The RMSEC describes how a model fits the calibration and RMSEP evaluates a calibration model versus a sample set from a new cultivation. They are calculated with the following equation:

$$\text{RMSEC or RMSEP} = \sqrt{\sum_{i=1}^{N} \frac{(\hat{Y}_i - Y_i)^2}{N}} \tag{9}$$

where $\hat{Y}_i$ represents the calculated concentration during calibration (RMSEC) or prediction (RMSEP) and Y_i is the concentration determined by reference analysis (either simulated or off-line concentrations). N stands for the measurement count.

RMSEC and RMSEP provide an estimate of the prediction error in the same unit as the initial data (g L^{-1}), while SEC and SEP provide the prediction error with respect to the maximum ethanol concentration in the reference data in terms of percentage. SEC and SEP are calculated as follow:

$$\text{SEC or SEP (\%)} = \frac{\sqrt{\sum_{i=1}^{N} \frac{(\hat{Y}_i - Y_i)^2}{N}}}{Y_{max}} * 100\ \% \tag{10}$$

where $\hat{Y}_i$ is the calculated concentration determined by the chemometric model and Y_i is the concentration determined by reference analysis (either simulated or off-line concentrations). N stands for the measurement count and Y_{max} is the maximum ethanol concentration in the reference data.

3. Results and discussion

3.1. Response of the gas sensor array

The gas sensor array response pattern from a cultivation of microorganisms may have various origins. The sensors can, if sensitive enough, respond to specific volatile compounds emitted by the microorganisms in the culture or to the

Table 1 — Predicted values of the specific growth rates and the parameters of the chemometric models using MBC and CCM calibration approaches.								
Cultivation	μ_G (h^{-1})	μ_E (h^{-1})	p_0 (g L^{-1})		p_1 ((g L^{-1})/(unit))		p_2 (g L^{-1}/Unit^2)	
	MBD	MBD	MBD	CCM	MBD	CCM	MBD	CCM
BC1	0.15	0.08	0.04	−0.23	0.38	0.79	1.37	0.09
BC2	0.15	0.07	0.04	−0.3	0.32	1.0	0.51	0.06
BC3	0.16	0.07	0.02	−0.1	0.85	0.27	1.24	0.09

Table 2 – Yield coefficient of all three cultivations as well as the selected yield coefficient for the simulation model from literature.

Cultivation	$Y_{X/G}(g_X/g_G)$	$Y_{E/G}(g_E/g_G)$	$Y_{X/E}(g_X/g_E)$
BC1	0.158	0.458	0.552
BC2	0.184	0.483	0.586
BC3	0.159	0.464	0.593
Simulation	0.175	0.473	0.598
Mean ± SD	0.167 ± 0.01	0.468 ± 0.01	0.577 ± 0.01

emission from the nutrient components of the medium (Bachinger & Mandenius, 2000).

In order to determine the sensitivity of the gas sensor array to the nutrient components of the medium as well as to the volatile compounds emitted by the cells, the response pattern of the gas sensor array from BC2 (before and after adding the suspended cell culture) was compared to the off-line values of ethanol, glucose and biomass which were measured during the same cultivation. For this reason the bioreactor was filled with 1 L of the medium and a glucose solution (to a final concentration in bioreactor of 7 g L^{-1}) and five sampling cycles were performed from the headspace of the bioreactor. Immediately before the 6th sampling cycle, the suspended cell was added and the sampling cycles were

performed every 5 min until the end of the cultivation process (Fig. 5).

Figure 5(a) indicates the evolution of glucose, ethanol and biomass during the cultivation (BC2) and Fig. 5(b) illustrates the raw signal response of one of the gas sensors during the same cultivation.

In Fig. 5(b), the five peaks before T_{start} indicate the response of the sensor to the components of the medium. At T_{start} the suspended cells are added to the bioreactor (inoculation time), T_{max} is the peak with the highest value and T_{end} is the last peak where the ethanol is depleted and whose height is the same as before T_{start}.

High sensitivity of the sensor to ethanol produced during the cultivation process can be described by comparing the general pattern of the peak heights (Fig. 5(b)) with the ethanol concentration (Fig. 5(a)).

In Fig. 5(a), while the glucose is decreasing during the glucose phase, the products (ethanol and biomass) are increasing. At around 2 h, the glucose is depleted and the cultivation shifts to the ethanol consumption phase. After around 7 h, the ethanol is also depleted and no more ethanol remained in the culture broth. Similar behaviour of the sensor response can be observed from Fig. 5(b). In Fig. 5(b), the peak height before T_{start} and after T_{end} is the same. In addition, there are no changes in the peak heights after T_{end} (the region where all the ethanol is consumed). From these illustrations it can be seen that the sensor is highly sensitive to ethanol but
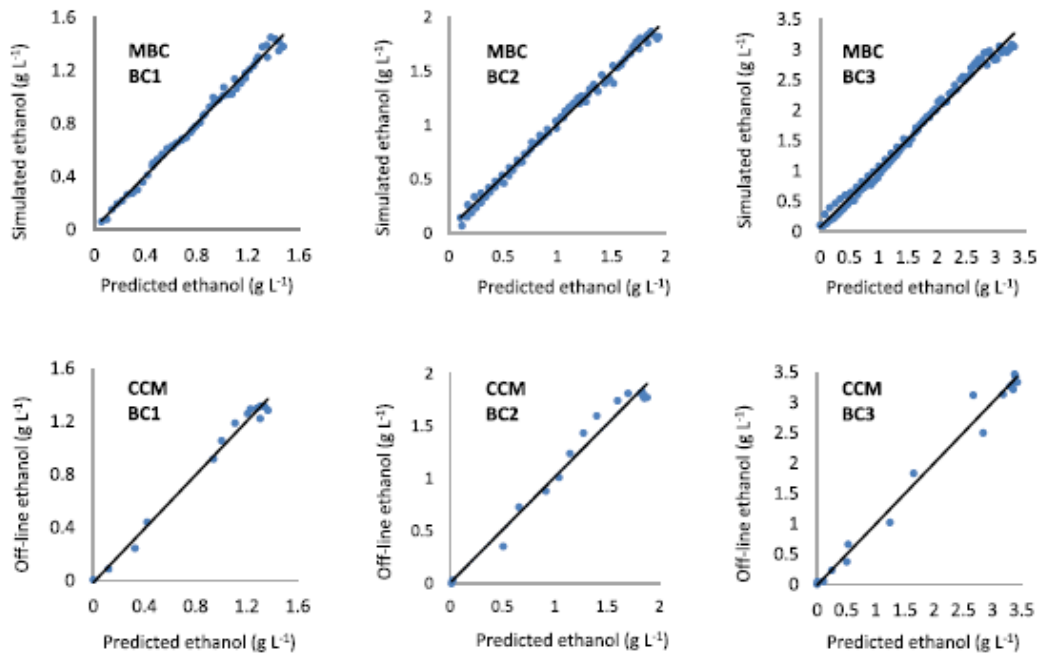
Fig. 6 – Predicted versus simulated ethanol concentrations using the MBC approach for all three cultivations (BC1–BC3) as well as the predicted verses off-line ethanol concentrations using the CCM approach for all three cultivations (BC1–BC3).

38

Table 3 — RMSEC and SEC for both calibration methods.

Cultivation	MBC		CCM	
BC1	0.01 g L^{-1}	3.3%	0.17 g L^{-1}	4.9%
BC2	0.04 g L^{-1}	2.5%	0.09 g L^{-1}	5.2%
BC3	0.03 g L^{-1}	2.4%	0.05 g L^{-1}	4.1%

not so sensitive to the other components of the cultivation medium.

3.2. Calibration models

For the MBC approach, the proposed method in section 2.6 was applied to the sensor data gathered from the headspace gas analysis of the bioreactor during each cultivation process separately. The evaluation of the predicted ethanol concentrations from the gas sensor array was compared with the simulated ethanol concentrations. The sum of squared differences was calculated and minimised by the particle swarm optimisation method. With this approach, the parameters of the chemometric models (p_0, p_1 and p_2) as well as the growth rates of the simulation model (μ_{G0} and μ_{E0}) were obtained.

For the CCM, the off-line ethanol concentrations (measured from the off-line samples taken during the cultivation) were fitted to the response of the gas sensor array and the sum of squared differences was minimised. The predicted values for the specific growth rates on glucose and ethanol (obtained from the MBC approach) as well as the parameters of the PCR model using both calibration approaches are presented in Table 1.
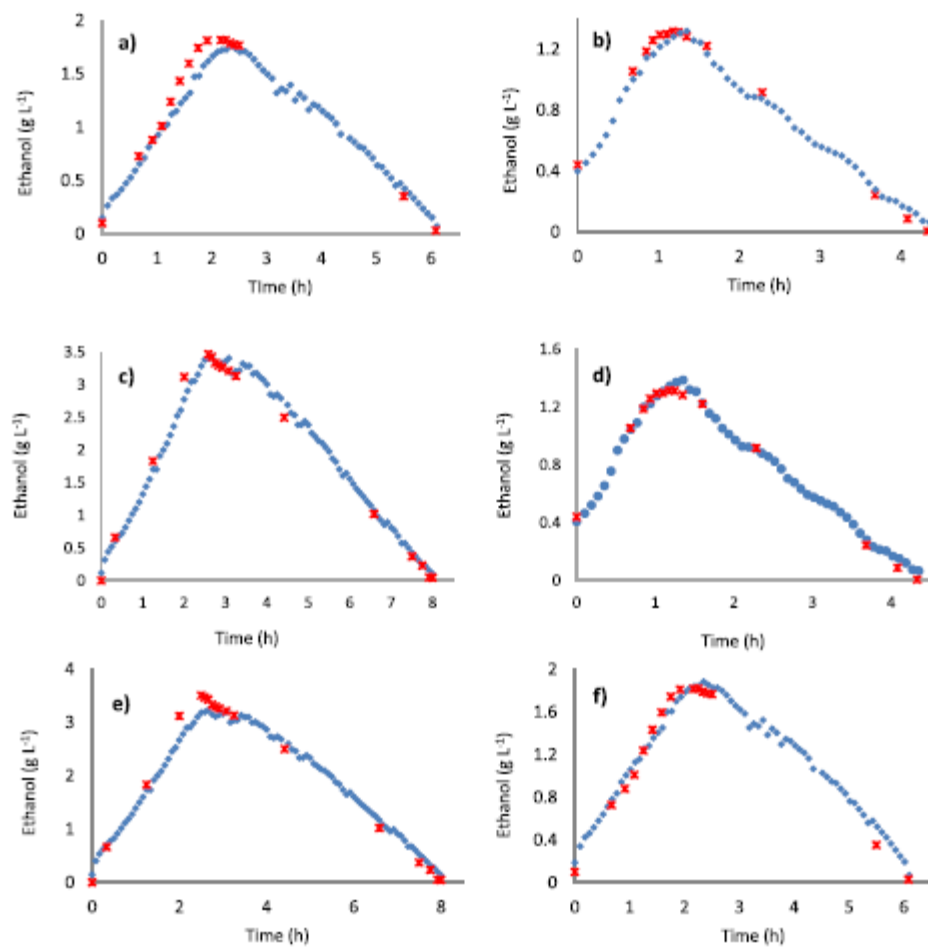
Fig. 7 – Predicted (dots) verses off-line (crosses) ethanol concentrations using the MBC approach. (a) and (b) present the predicted and off-line ethanol concentrations (using BC3 for calibration) during BC2 and BC1 respectively. (c) and (d) present the predicted and off-line ethanol concentrations (using BC2 for calibration) during BC3 and BC1 respectively. (e) and (f) present the predicted and off-line ethanol concentrations (using BC1 for calibration) during BC3 and BC2 respectively.

Data in Table 1 reveal that there is no significant difference between the growth parameters (μ_{G0} and μ_{E0}) from cultivations with different initial conditions. This shows that the yeast cells have regulatory mechanisms to be able to balance the cellular activity in different conditions. Furthermore, the values for μ_{G0} and μ_{E0} that were obtained by fitting the theoretical process model directly to the off-line data of the same cultivations were 0.15 h^{-1} and 0.074 h^{-1}, respectively. So the growth rate values are equal. These values also fit well with the range of S. cerevisiae's growth rates reported in literature (Boender et al., 2009). Therefore, the parameter estimation method can be considered reliable.

As previously mentioned in section 2.1, the yield coefficients of the simulation model were obtained from literature. In order to ensure that no changes in these coefficients occurred, the yield coefficient of each cultivation was also obtained by fitting the process model to the off-line data. The obtained values were compared to the chosen values from literature (Table 2).

Between three different cultivations as well as the simulated process model, the values of yield coefficients have no significant difference due to the low standard deviation. This indicates that the yield coefficients chosen from literature are in the correct range and the simulation model can describe the process accurately.

As a method of assessing the fit of the calibration models to the data, the correlation plots were prepared (Fig. 6). The predicted versus simulated ethanol concentrations using the MBC approach for all three cultivations (BC1–BC3) as well as the predicted versus off-line measured ethanol concentrations using the CCM approach for all three cultivations (BC1–BC3) are presented.

The RMSEC and SEC were chosen as the numerical tools for accuracy assessment of the calibration models. The values are given in Table 3.

The results in Table 3 indicate that the most suitable calibration method for the determination of the ethanol concentration is the MBC approach (RMSEC is below 3.5% in all 3 cultivations). This was to be expected due to the difference in the number of data used during calibration, because for the CCM approach just 13 samples were collected and analysed off-line. However, with a relatively small number of training data, the CCM approach is also a reliable method for the determination of ethanol concentration (RMSEC is below 5.5% in all 3 cultivations).

3.3. Validation of the calibration models

In order to see if the calibration models are able to predict the ethanol concentration during new process runs, each calibration model obtained from a cultivation run was validated with the data from the other two cultivations which have different initial concentrations. Figure 7(a) and (b) present the predicted (using data from BC3 for the calibration model) as well as the off-line measurements of ethanol concentration as a function of time during BC2 and BC1 respectively. Figure 7(c) and (d) present the predicted (using data from BC2 for the calibration model) as well as the off-line measurements of ethanol concentration as a function of time during BC3 and BC1 respectively. Figure 7(e) and (f)

present the predicted (using data from BC1 for the calibration model) as well as the off-line measurements of ethanol concentration as a function of time during BC2 and BC2 respectively.

Figure 7 indicates that the predicted ethanol concentration using the MBC approach corresponds well with the off-line measurements during all cultivations. Furthermore, the ethanol production phase and ethanol consumption phase are clearly indicated by the predicted ethanol values without a significant time delay (in comparison with the off-line values). In S. cerevisiae batch cultivation, the metabolic shift of the yeast cells (shifting from ethanol production to ethanol consumption) is a critical point which indicates a significant change in its metabolism and can be observed using the gas sensor array. Ethanol prediction with the MBC approach was compared with the ethanol prediction using the CCM approach. For this reason RMSEP and SEP were calculated and the results are shown in Table 4.

Table 4 reveals that, by using the MBC approach, the SEP of prediction is below 7% in all cases except when BC2 is used for calibration and the data set from BC1 is used for validation (SEP = 9.3%). However, even though in the MBC approach no off-line measurement were used during the calibration procedure of the chemometric model, the prediction corresponds very well with the simulated values. Furthermore, in comparison with the predictions from using the CCM, lower prediction errors are obtained.

The larger errors of prediction from the CCM are because not so many off-line samples are used for calibration. Therefore, by increasing the number of off-line samples, more accurate predictions might be obtained. However this would be a time-consuming and expensive approach.

When using the MBC approach, the percentage error greatly depends on the kinetic parameters of the simulation model that are obtained from the optimisation algorithm. If they are close to the real values, the process model will describe the process sufficiently accurately. Therefore, improving the optimisation process can lead to even lower percentage errors. Of course, the percentage errors also depend on how good the off-line samplings were performed and how accurate they are. However, the prediction error for ethanol concentrations is below 10% in all three cultivations

Table 4 – The RMSEP and SEP obtained with the MBC and CCM approach.

Using BC1 for the calibration model				
Validation with	MBC		CCM	
BC2	0.16 g L^{-1}	9.3%	0.1 g L^{-1}	6.0%
BC3	0.08 g L^{-1}	6.3%	0.1 g L^{-1}	7.6%
Using BC2 for the calibration model				
Validation with	MBC		CCM	
BC1	0.63 g L^{-1}	5.6%	0.34 g L^{-1}	9.7%
BC3	0.06 g L^{-1}	5.0%	0.11 g L^{-1}	9.0%
Using BC3 for the calibration model				
Validation with	MBC		CCM	
BC1	0.24 g L^{-1}	6.9%	0.29 g L^{-1}	8.5%
BC2	0.11 g L^{-1}	6.3%	0.13 g L^{-1}	7.2%

which is considered decent for a bioprocess especially when considering that no off-line values were used to achieve this result.

4. Conclusion

The signals from the gas sensor array described here showed a high correlation with ethanol concentration during cultivations of *S. cerevisiae growing on glucose*. Similar to any other indirect measurement method, a chemometric model is required for predicting the ethanol concentration from the signals of the gas sensor array. This approach normally requires off-line sampling for calibration purposes which is expensive and time-consuming. Alternatively, a simulation model can be used, if the correct parameters for the model are known. In the proposed method, the only requirement for calculating the parameters of the simulation model is the response of the gas sensor array from a single cultivation run. Then the parameters for the simulation model can be calculated by minimising the prediction error by optimising the kinetic parameter values of the simulation model as well as the parameter values of the chemometric model.

The proposed MBC method provided comparable results to the reference ethanol concentration values obtained by HPLC. Furthermore, compared to spectroscopy-supported models for ethanol prediction which have applied separate prediction models for each diauxic growth phases of the cultivation, in this investigation only a single model was applied for the two diauxic growth phases.

The proposed measurement system is inexpensive to implement, has just a few maintenance requirements and can be implemented on bioreactors with different volumes. However it could be assumed when dealing with much higher ethanol concentrations, the sensor measurement limit would be exceeded. In this case, the outlet gas of the bioreactor can be diluted with a reference gas. Overall, the gas sensor array seems to be a useful non-invasive tool for continuous monitoring of yeast cultivation and might also be used for closed loop control of processes involving ethanol measurements.

Declaration of competing interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

REFERENCES

Aguilera, T., Lozano, J., Paredes, J. A., Alvarez, F. J., & Suárez, J. I. (2012). Electronic nose based on independent component analysis combined with partial least squares and artificial neural networks for wine prediction. *Sensors*, 12(6), 8055—8072.

AS, C. P. (2006). *The Unscrambler Tutorials*. On line at http://www.camo.com/downloadsU9.

Assawajaruwan, S., Eckard, P., & Hitzmann, B. (2017a). On-line monitoring of relevant fluorophores of yeast cultivations due to glucose addition during the diauxic growth. *Process Biochemistry*, 58, 51—59.

Assawajaruwan, S., Reinalter, J., & Hitzmann, B. (2017b). Comparison of methods for wavelength combination selection from multi-wavelength fluorescence spectra for on-line monitoring of yeast cultivations. *Analytical and Bioanalytical Chemistry*, 409(3), 707—717.

Bachinger, T., & Mandenius, C.-F. (2000). Searching for process information in the aroma of cell cultures. *Trends in Biotechnology*, 18(12), 494—500.

Bachinger, T., & Mandenius, C. F. (2001). Physiologically motivated monitoring of fermentation processes by means of an electronic nose. *Engineering in Life Sciences*, 1(1), 33—42.

Boender, L. G., de Hulster, E. A., van Maris, A. J., Daran-Lapujade, P. A., & Pronk, J. T. (2009). Quantitative physiology of Saccharomyces cerevisiae at near-zero specific growth rates. *Applied and Environmental Microbiology*, 75(17), 5607—5614.

Buratti, S., & Benedetti, S. (2016). Alcoholic fermentation using electronic nose and electronic tongue. *Electronic Noses and Tongues in Food Science*, Elsevier, 291—299.

De Deken, R. (1966). The Crabtree effect: A regulatory system in yeast. *Microbiology*, 44(2), 149—156.

Di Carlo, S., & Falasconi, M. (2012). *Drift correction methods for gas chemical sensors in artificial olfaction systems: techniques and challenges*. INTECH Open Access Publisher.

Ghosh, S., Tudu, B., Bhattacharyya, N., & Bandyopadhyay, R. (2017). A recurrent Elman network in conjunction with an electronic nose for fast prediction of optimum fermentation time of black tea. *Neural Computing & Applications*, 31(2), 1165—1171.

Hidayat, S. N., Nuringtyas, T. R., & Triyana, K. (2018). Electronic nose coupled with chemometrics for monitoring of tempeh fermentation process. In *2018 4th International Conference on Science and Technology (ICST)*. IEEE.

Jiang, H., Zhang, H., Chen, Q., Mei, C., & Liu, G. (2015). Recent advances in electronic nose techniques for monitoring of fermentation process. *World Journal of Microbiology and Biotechnology*, 31(12), 1845—1852. https://doi.org/10.1007/s11274-015-1940-0

Kiani, S., Minaei, S., & Ghasemi-Varnamkhasti, M. (2016). A portable electronic nose as an expert system for aroma-based classification of saffron. *Chemometrics and Intelligent Laboratory Systems*, 156, 148—156.

Lidén, H., Mandenius, C.-F., Gorton, L., Meinander, N. Q., Lundström, I., & Winquist, F. (1998). On-line monitoring of a cultivation using an electronic nose. *Analytica Chimica Acta*, 361(3), 223—231.

Lin, B., Recke, B., Knudsen, J. K., & Jørgensen, S. B. (2007). A systematic approach for soft sensor development. *Computers & Chemical Engineering*, 31, 419—425.

Li, G., Yuan, L., Wang, X., Meng, Y., Li, J., Zhao, Y., & Peng, Y. (2019). Rapid quantification analysis of alcohol during the green jujube wine fermentation by electronic nose. In *IOP Conference Series: Earth and Environmental Science*. IOP Publishing.

Madigan, M., Bender, K., Buckley, D., Sattley, W., & Stahl, D. (2017). *Brock Biology of Microorganisms* (15th ed.). Essex, England: Pearson.

Mandenius, C. F., Eklöv, T., & Lundström, I. (1997). Sensor fusion with on-line gas emission multisensor arrays and standard process measuring devices in baker's yeast manufacturing process. *Biotechnology and Bioengineering*, 55(2), 427—438.

Ödman, P., Johansen, C. L., Olsson, L., Gernaey, K. V., & Lantz, A. E. (2009). On-line estimation of biomass, glucose and ethanol in Saccharomyces cerevisiae cultivations using in-situ multi-

wavelength fluorescence and software sensors. *Journal of Biotechnology, 144*(2), 102–112.

Omatu, S., & Yano, M. (2016). E-nose system by using neural networks. *Neurocomputing, 172*, 394–398. https://doi.org/10.1016/j.neucom.2015.03.101

Otto, M. (1999). *Chemometrics: Statistics and computer application in analytical chemistry*. Weinheim: Wiley-VCH.

Paquet-Durand, O., Assawarajuwan, S., & Hitzmann, B. (2017a). Artificial neural network for bioprocess monitoring based on fluorescence measurements: Training without offline measurements. *Engineering in Life Sciences, 17*(8), 874–880.

Paquet-Durand, O., Ladner, T., Büchs, J., & Hitzmann, B. (2017b). Calibration of a chemometric model by using a mathematical process model instead of offline measurements in case of a H. polymorpha cultivation. *Chemometrics and Intelligent Laboratory System, 171*, 74–79.

Solle, D., Geissler, D., Stärk, E., Scheper, T., & Hitzmann, B. (2003). Chemometric modelling based on 2D-fluorescence spectra without a calibration measurement. *Bioinformatics, 19*(2), 173–177.

Tan, J., Balasubramanian, B., Sukha, D., Ramkissoon, S., & Umaharan, P. (2019). Sensing fermentation degree of cocoa (Theobroma cacao L.) beans by machine learning classification models based electronic nose system. *Journal of Food Process Engineering, 42*(6). e13175.

Tan, C., Xie, D., Liu, Y., Peng, W., Li, X., Ai, L., Wu, C., Wen, C., Huang, X., & Guo, J. (2018). Identification of different bile species and fermentation times of bile arisaema based on an intelligent electronic nose and least squares support vector machine. *Analytical Chemistry, 90*(5), 3460–3466.

Wang, G.-G., Gandomi, A. H., Yang, X.-S., & Alavi, A. H. (2014). A novel improved accelerated particle swarm optimization algorithm for global numerical optimization. *Engineering Computations, 31*(7), 11-98-1220.

Wild, R., Citterio, D., Spichiger, J., & Spichiger, U. E. (1996). Continuous monitoring of ethanol for bioprocess control by a chemical sensor. *Journal of Biotechnology, 50*(1), 37–46.

Zhang, Z., Scharer, J., & Moo-Young, M. (1997). Mathematical model for aerobic culture of a recombinant yeast. *Bioprocess Engineering, 17*(4), 235–240.

4. 2. The Kalman filter for the supervision of cultivation processes.

By Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, Bernd Hitzmann. Published in Advances in Biochemical Engineering/Biotechnology. Page 1-30. November 2020.

The Kalman Filter for the Supervision of Cultivation Processes

Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, and Bernd Hitzmann

Contents

Abstract In the era of technology and digitalization, the process industries are undergoing a digital transformation. The available process models, advance sensor technologies, enhanced computational power and a broad set of data analytical

A. Yousefi-Darani (✉), O. Paquet-Durand, and B. Hitzmann
Department of Process Analytics and Cereal Science, Institute of Food Science and
Biotechnology, University of Hohenheim, Stuttgart, Germany
e-mail: rahim@uni-hohenheim.de

techniques enable solid bases for digital transformation in the biopharmaceutical industry.

Among various data analytical techniques, the Kalman filter and its non-linear extensions are powerful tools for prediction of reliable process information. The combination of the Kalman filter with a virtual representation of the bioprocess, called digital twin, can provide real-time available process information. Incorporation of such variables in process operation can provide improved control performance with enhanced productivity.

In this chapter the linear discrete Kalman filter, the extended Kalman filter and the unscented Kalman filters are described and a brief overview of applications of the Kalman filter and its non-linear extensions to bioreactors are presented. Furthermore, in a case study an example of the digital twin of the backer's yeast batch cultivation process is presented.

Graphical Abstract A digital twin of a bioreactor mirrors the processes of the real bioreactor. It contains the physical parts, the process model and prediction algorithm to predict the bioprocess variables. These values could be used for optimization and control of the process.
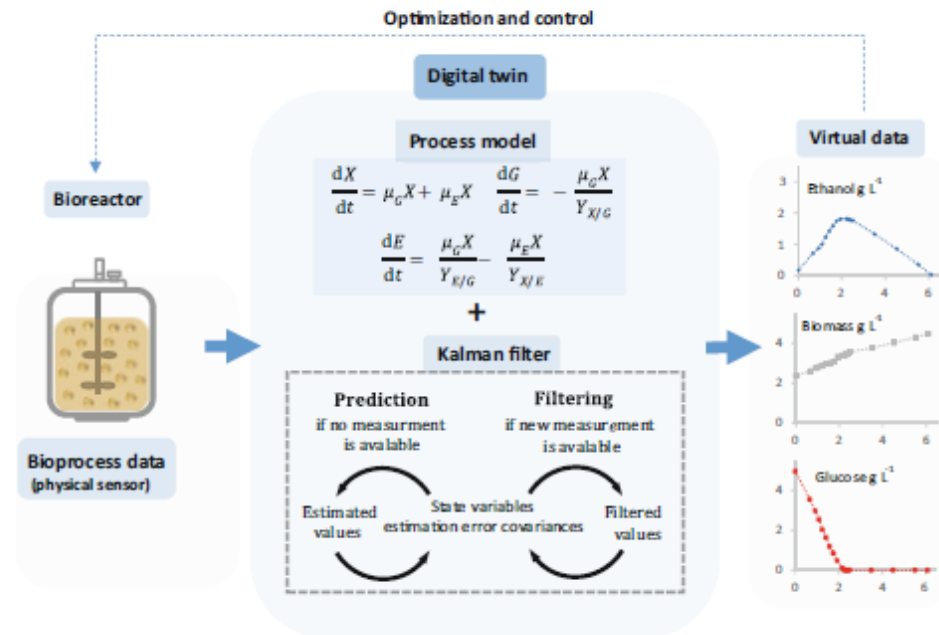
Abbreviations

A	State transition matrix
B	Process input transition matrix
C	Measurement model
CKF	Cubature Kalman filter
EKF	Extended Kalman filter
EnKF	Ensemble Kalman filter
F	Jacoby matrix of f()
f()	Non-linear function describing the process change
FIA	Flow injection analysis
H	Jacoby matrix of measurement model
h()	Measurement model
KF	Kalman filter
P	Estimation error covariance matrix
$\mathbf{p}$	Model parameter vector for estimation
Q	Process noise covariance matrix
R	Measurement noise covariance matrix
t	Time
UKF	Unscented Kalman filter
$\mathbf{v}$	Measurement noise vector
$\mathbf{w}$	Process noise vector
$\boldsymbol{x}$	State variables vector
$\mathbf{x}(t)$	State variable at continuous time k
$\mathbf{x}_{[k]}$	State variable at discrete time k
$\mathbf{x}_{f,[k]}$	Filtered state variable at discrete time k
$\mathbf{z}$	Measurement vector

1   Introduction

Bioprocesses are described as biological systems that are non-linear, complex and unsteady; thus development of precise control systems in order to achieve robust product quality and productivity can be challenging. The control of these processes can be significantly improved by online process monitoring followed by corrective actions. In this context, bioprocess digital twins are helpful tools.

Digital twins are virtual representations of the production process which enable pre-emptive process control by using online data to predict the process outcome in advance. They convert the physical process to a smart process and thus achieve the ultimate goal of the digital transformation. This enables unprecedented possibilities for timely and automated intervention to provide critical decision support during process development [1].

Digital twins mainly consist of a mathematical model which describes the dynamic behaviour observed in a biochemical reactor and a prediction or self-learning algorithm which estimates the cellular component concentrations and the process parameters that cannot be described mechanistically [2, 3].

Bioprocess mathematical models may generally be categorized into algebraic equations and dynamic models. Algebraic equations are developed from mass and component balances, from mass or heat transfer laws or even from elemental balances. Dynamic models usually consist of dynamic balances of conserved quantities in combination with kinetics to describe rate expressions as functions of the state variables. Detailed description of mathematical modelling of bioprocesses is covered by previous authors in greater details than space allows here [4–7]. The goal of this chapter is to highlight state estimation methods with a specific focus on the Kalman filter and its non-linear extensions.

For linear systems, the Luenberger observer and the Kalman filter, whose 60th anniversary occurred in 2020 [8], are the most applied methods for estimating parameters and process variables that cannot be measured directly. In the area of non-linear systems, particle filtering (PF), high gain observers, non-linear extensions of the Kalman filter such as the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) and many others have been proposed. However, due to the simple structure and low computational effort of non-linear extensions of the Kalman filter, these methods have gained more interest, and many research studies have been dedicated to the implementation of such filters for state and parameter estimation in bioprocess technologies. The main objective of this chapter is to discuss the applications of different Kalman filter algorithms in bioprocess technologies. Therefore, this chapter is organized as follows: in the next section, a brief overview of the Kalman filtering theory and its non-linear extensions will be discussed. Applications of the Kalman filter for the supervision of cultivation processes will be given in the third section, followed by a case study evaluating the implementation of an extended Kalman filter for developing a digital twin of the backer's yeast batch cultivation process. In the last section, a conclusion is presented.

2 Kalman Filtering Theory and Its Non-linear Extensions

The Kalman filter is a set of mathematical equations that provides an efficient computational solution of the least-squares method when the considered system is linear and the uncertainties are modelled by Gaussian random variables. When the system state dynamics is non-linear, then certain linearization methods are applied. The most prominent of these algorithms are the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), invented independently by several research groups. Different extensions of the Kalman filters differ in the way the estimation error is calculated. A brief overview of these methods are as follows.

2.1 The Kalman Filter

The Kalman filter is used to provide optimal estimates of unmeasured states for time varying linear systems in the presence of noise by combining information from a process mathematical model with online process measurements. The process model defines the evaluation of the state from time $k-1$ to time k as:

$$x_{[k]} = Ax_{[k-1]} + Bu_{[k-1]} + w_{[k-1]} \tag{1}$$

where x is the state vector, u is the process input and w is the Gaussian process noise vector that is assumed to be zero-mean with the covariance Q. Matrix A relates the state at the previous time step $k-1$ to the state at the current step k, matrix B relates the control input to the state variables x.

The process model is paired with the measurement model that describes the relationship between the state and the measurement at the current time step k as:

$$z_{[k]} = Cx_{[k]} + v_{[k]} \tag{2}$$

where z is the measurement vector and v is the Gaussian measurement noise vector which is assumed to be zero-mean with the covariance R. Matrix C relates the state to the measurement $z_{[k]}$. Since the measurements does not exhaustively inform on the current situation of the process, the KF aims to provide an estimate of the process state at time k, given the initial state of x_0, the measurements and the information of the system.

The Kalman filter algorithm consists of two steps which are summarized as follows:

- *Prediction step (time update)*: Using the initial condition, the process model is used to predict the state variables and the estimation error covariance's until the first measurement is available.

$$x_{[k]} = Ax_{[k-1]} + Bu_{[k-1]} \tag{3}$$

$$P_{[k]} = AP_{[k-1]}A^T + Q \tag{4}$$

In the above equations, $x_{[k]}$ is the state variables estimate at time k which is deduced from a previous estimation of the state $x_{[k-1]}$ at time $k-1$. The new term **P** is called the state error covariance matrix which encrypts the error covariance of the predicted state values. $\mathbf{P}_{[k]}$ is the new prediction error covariance matrix at time k and $\mathbf{P}_{[k-1]}$ is the previous estimated error covariance matrix at time $k-1$. Whenever a measurement is available, a correction step is performed:

- *Correction step (measurement update)*: In this step the predicted model estimates are combined with the measured values to provide corrected estimates.
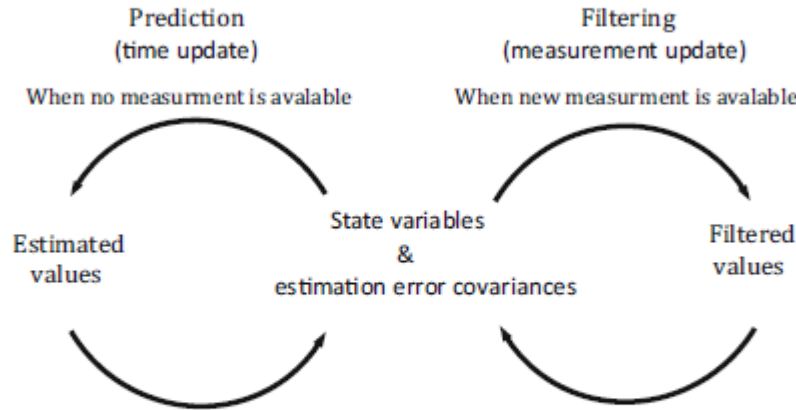
Fig. 1 The flow chart of the Kalman filter algorithm

$$x_{f,[k]} = x_{[k]} + K_{[k]}\left(z_{[k]} - Cx_{[k]}\right) \tag{5}$$

$$P_{f,[k]} = P_{[k]}\left(1 - K_{[k]}C\right)^2 + K^2 R \tag{6}$$

$$K_{[k]} = P_{[k]}C^T\left(R + CP_{[k]}C^T\right)^{-1} \tag{7}$$

The measurement prediction error, reflects the discrepancy between the true measurements $z_{[k]}$ and the predicted measurements $Cx_{[k]}$. The difference of both is multiplied by the so called Kalman gain and used to update the estimated state variables. Therefore the filtered state variables $x_{f,[k]}$ are obtained. In the similar manner, the filtered estimation error covariance $P_{f,[k]}$ is obtained. $K_{[k]}$ is chosen to minimizes the estimated error covariance

$$\frac{dP_f}{dK} = 0 \tag{8}$$

The measurement error variance must be compared with the estimation error variance to see how the filter is acting. For this purpose, a very rough treatment is necessary:

If $R \ll CP_{[k]}C^T$ then $K \approx C^{-1}$ and $x_{f,[k]} \approx C^{-1}z_{[k]}$; so the filtered is almost determined by the measured.

If $R \gg CP_{[k]}C^T$ then $x_{f,[k]} \approx x_{[k]}$; the filtered value is almost the estimated one and no influence of the measurement will be obtained.

With the filtered values as initial condition the simulation of the process as well as the estimation error covariance's can be carried out until the next measurement is obtained and everything repeats again. The flow chart of the Kalman filter algorithm is presented in Fig. 1.

2.2 Continuous-Discrete Extended Kalman Filter

As described in the previous section, the Kalman filter addresses the general problem of trying to estimate the state of a process that is governed by a linear differential equation system. In non-linear dynamic systems, the process model or the measurement model cannot be determined with multiplication of vectors and matrices. For such systems, a linearization should be performed. The linearization can be performed by different methods. The essential difference among different versions of the Kalman filters (extended Kalman filter, unscented Kalman filter and ensemble Kalman filter) consists in how they calculate the estimation error. A Kalman filter that linearizes about the current mean and covariance is referred to as an extended Kalman filter (EKF). A non-linear dynamic system can be described by the following differential equation:

$$\frac{dx(t)}{dt} = f(x(t), u(t)) + w(t) \tag{9}$$

With discrete measurements that are:

$$z_{[k]} = h\left[x(t_{[k]})\right] + v_{[k]} \tag{10}$$

The differential equation provide the continuous part, the measurements are the discrete part, where f is a non-linear function of the state variables x and the control input u. The non-linear function h in the measurement equation relates the current state to the measurement $z_{[k]}$. w and v are, respectively, the process noise vector and the measurement noise vector. These noises are assumed to be zero mean, white, and independent of each other, with respective covariance matrices Q and R.

To calculate the estimation error covariance matrix, the following differential equations have to be solved in parallel to the state differential equation.

$$\frac{dP(t)}{dt} = F(t)P(t) + P(t)F^T(t) + Q \tag{11}$$

Here the Jacobian matrix is used, which is given by the following equation:

$$F = \frac{\partial f}{\partial x}\bigg|_{x(t),\ u(t)} \tag{12}$$

The filtering is performed as follows:

$$K_{[k]} = P(t_k)H^T(t_k)\left[H(t_k)P(t_k)H^T(t_k) + R\right]^{-1} \tag{13}$$

$$\boldsymbol{x}_f\left(t_{[k]}\right) = \boldsymbol{x}\left(t_{[k]}\right) + K_{[k]}\left[\boldsymbol{z}_{[k]} - h\left[\boldsymbol{x}\left(t_{[k]}\right)\right]\right] \tag{14}$$

$$P_f\left(t_{[k]}\right) = \left[I - K_{[k]}H_{[k]}\right]P\left(t_{[k]}\right)\left[I - K_{[k]}H_{[k]}\right]^T + K_{[k]}RK_{[k]}^T \tag{15}$$

where $H_{[k]}$ is the Jacoby matrix of h[]:

$$H_{[k]} = \left.\frac{\partial h}{\partial x}\right|_{x_{[k]}} \tag{16}$$

Correspondingly to the KF algorithm, the EKF algorithm consists of two main parts including prediction step and the correction step.

As mentioned above, the basic framework for the EKF involves state estimation of a non-linear dynamic system. However, in some cases, prediction of $\boldsymbol{x}_k$ requires coupling both state estimation and parameter estimation [9]. Here a process model parameter $p(t)$ is considered to be time dependent and can be estimated by adding the parameter as an additional state variable whose differential equation is then given as

$$\frac{d\boldsymbol{p}(t)}{dt} = 0 \tag{17}$$

At every time step, the current estimate of the parameter $p(t)$ is used in the measurement filter. In the joint estimation method, model state variables and model parameters are included in a single joint state vector. Parameter estimation evolves in time along with state estimation, as observations are assimilated [10].

Other alternatives for parameter estimation with the KF include calibrating parameters outside the KF calculation with an outer optimisation routine [11–13], and parameter estimation in steady-state KF calculations where observations are climatological averages over the entire time period of interest [14], but in both of these two approaches the parameter estimation part of the calculation considers all observations at once rather than sequentially.

2.3 Other Non-linear Extensions of the Kalman Filter

As mentioned previously, when the system is non-linear and can be well approximated by linearization, then the EKF is a good option for state estimation; however EKF is not optimal if the system is highly non-linear, this is because only the mean is propagated through the non-linearity [15]. The unscented Kalman filter (UKF) is another non-linear extension of the Kalman filter which is a discrete time filtering algorithm. The UKF utilizes the unscented transformation for computing approximate solutions to the filtering problems.

A general framework for state estimation based on the UKF for this state space model is presented as follows:

In the first step, the initial values for the state and covariance estimation have to be set. Following this, the recursive estimation is performed by the prediction and correction steps. Within the prediction step, a priori state and covariance estimation utilizing the process model is performed. Using the unscented transformation, a set of sigma points are chosen. These sigma points characterize the current probability density function. Each point from the sigma matrix is propagated through the process model to calculate the estimations of state variables and the error covariance. Following this, a correction step is preformed when a measurement is received. This leads to the estimations of the filtered state variables and the filtered error covariance by calculating the Kalman gain.

The UKF has been used in various fields for non-linear sate estimations. However a couple of alternative approaches have emerged over the last few years, namely, the ensemble Kalman filter (EnKF) and the cubature Kalman filter (CKF) which are widely used when the process model is of extremely high order and non-linear, the initial states are highly uncertain and a large number of measurements are available [16, 17].

Similar to the UKF, the EnKF and CKF select a set of sample points (sigma points) in order to deal with the non-linearity of the system. In high-dimension systems, the weights of the sigma points in the UKF are prone to be negative, leading to low estimation accuracy.

In EnKF the error covariances are estimated approximately using an ensemble of model forecasts. The main concept behind the formulation of the EnKF is that if the dynamical model is expressed as a stochastic differential equation, the prediction error statistics, which are described by the Fokker–Plank equation, can be estimated using ensemble integrations, and the error covariance matrices can be calculated by integrating the ensemble of model states [16].

The cubature Kalman filter uses the spherical–radial cubature rule to generate some weighted sampling points to approximate integral in Bayesian estimation. A brief overview of the unscented Kalman filtering and sigma point filtering in general are given by van der Merwe [18].

3   Application of Kalman Filters in Bioprocess Monitoring

Here 41 recent published articles [19–60] in the period of 1991–2020 on application of the Kalman filter and its extensions for state and parameter estimation in bioprocesses are discussed. Due to space limitation, only some of the reported articles are presented in Table 1. The table is organized by classifying the articles into different categories, which include the type of the Kalman filter and the applied process model, the type of microorganism and the cultivation process mode, the measured process variable(s) and the objective of the filtering algorithm. This table would help understanding how the Kalman filter was explored chronologically to date. It should be mentioned that in some works more than one Kalman filter

Table 1 Extended Kalman filter application for cultivation processes

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
Extended Kalman filter/experimental application	Batch cultivation/E. coli	Dissolved oxygen mass balance	Noise filtering from dissolved oxygen measurements	Dissolved oxygen	Lee et al. [19]
Extended Kalman filter/experimental application	Fed-batch cultivation/ S. cerevisiae	Material balance equation with Monod growth rate kinetics	Parameter estimation and substrate prediction	Glucose concentration with FIA	Hitzman et al. [32]
Kalman filter/experimental application	Batch cultivation/ S. cerevisiae	Ideal stirred tank reactor model with Monod growth kinetics (glucose and ethanol as limiting substrates)	Noise filtering from predicted bioprocess variables	Biomass, glucose, and ethanol (with ultrasonic velocity)	Cha and Hitzmann [36]
Extended Kalman filter/experimental application	Fed-batch cultivation/ S. cerevisiae	A model for an ideal stirred tank reactor in combination with Monod growth kinetics	Noise filtering from predicted glucose	Glucose concentration with flow injection analyses (FIA)	Arndt and Hitzmann [37]
Extended Kalman filter/ simulation	Fed-batch cultivation/ S. cerevisiae	Cybernetic model of Jones and Kompala	Filtering out noise from the feed stream	Dilution rate or the gas–liquid mass transfer coefficient for oxygen	Patnaik [39]
Extended Kalman filter/ simulation	Fed-batch cultivation/E. coli	General dynamic model of bioreactors with Monod growth kinetics	Parameter estimation and biomass prediction	Dissolved and exhaust oxygen and carbon dioxide	Rocha et al. [40]
Extended Kalman filter/experimental application	Fed-batch cultivation/ Bordetella pertussis	A model with two parameters which are calculated using separate experiments	Estimation of specific growth rate, biomass, and oxygen mass transfer	Dissolved oxygen	Soons et al. [42]

(continued)

Table 1 (continued)

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
Unscented Kalman filter/experimental application	Fed-batch cultivation/ hybridoma cell culture	Overflow metabolism model	Noise reduction from predicted values	Predicted specific uptake and production rate	Henry et al. [41]
Extended Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Ideal stirred tank reactor model with Monod growth kinetics	Parameter, biomass, and glucose prediction	Glucose concentration with FIA	Klockow et al. [43]
Extended Kalman filter/experimental application	Fed-batch cultivation/*E. coli*	General dynamic model of bioreactors with Monod growth kinetics	Estimation of biomass, glucose, and acetate	Dissolved oxygen and carbon dioxide	Veloso et al. [44]
Unscented Kalman filter/ simulation	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass in the headspace with Monod growth kinetics	Estimation of biomass and substrate concentrations	Dissolved oxygen and carbon dioxide	Jianlin et al. [46]
Unscented Kalman filter/ simulation	Fed-batch/ hybridoma cell	Material balance equation with Monod growth kinetic	Prediction of acetate and glucose concentration	Biomass and dissolved oxygen	Dewasme et al. [48]
Extended Kalman filter/ simulation	Batch cultivation/ *S. cerevisiae*	Unstructured model for alcoholic fermentation with immobilized cells using Monod growth kinetics	Estimation of product, substrate, and biomass concentrations	Glucose and ethanol	Popova et al. [49]
Extended Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass in the headspace with Monod	Estimation of substrate and biomass concentrations	Substrate and biomass concentration with NIR spectrometer	Krämer and King [54]

(continued)

54

Table 1 (continued)

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
		growth kinetics			
Unscented Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass with Monod growth kinetics	Biomass and specific biomass growth rat estimation	Oxygen uptake and CO_2 formation rate	Simutis and Lübert [55]
Sigma point Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass in the headspace with Monod growth kinetics	Estimation of substrate and biomass concentrations	Substrate and biomass concentration with NIR spectrometer	Krämer and King [57]
Extended Kalman filter/ simulation	Fed-batch cultivation/ *S. cerevisiae*	Material balance equation with Monod growth rate kinetics	Ethanol prediction and state estimation	Temperature, do and substrate concentration	Lisci and Tronci et al. [60]

algorithm are examined. More detailed description of each category for all publications is presented in the following part of this section.

3.1   Type of Kalman Filter

According to the type of Kalman filter algorithm, the literature presented indicates there exist a considerable number of articles on implementation of EKF for state and parameter estimation. More than 60% of the applications (28 articles) have implemented EKF algorithms for their process. This is due to the fact that the cultivation process of microorganisms is a complex non-linear biochemical process and the EKF is a well-known state estimation method for non-linear systems. The linear Kalman filter which is almost exclusively used for state estimation in linear systems have also been used by some authors (3 articles). Although the EKF shows good prediction results and is widely used in literature, it presents some disadvantages. It is reliable for systems which are almost linear on the time scale of the update intervals; it requires the calculation of Jacobians at each time step, which may be difficult to obtain for higher order systems; it does linear approximations of the system at a given time instant, which may introduce errors in the estimation, leading then the state to diverge over time [9, 15]. For instance, in continuous or fed-batch

cultivations, despite continuous supply by a feed, the substrate concentration can drop to zero as the cell takes it up very fast. In such cultivations, linearization in the time and measurement update can lead to significant inaccuracies in the process, while the EKF assumes a certain probability for substrate concentrations below zero, even though this is physically impossible [54]. Therefore in recent years, application of other non-linear extensions of the Kalman filter is used. For example, Fernandes et al. [54] have implemented an UKF algorithm in order to estimate glucose and glutamine from biomass, lactate and ammonia measurement during fed-batch cultivation of hybridoma cells. The predictions were compared to the ones obtained with an EKF; they have reported the UKF achieves better level of accuracy. Krämer and King [57] have implemented a UKF in fed-batch cultivation of *S. cerevisiae* for noise filtering from predicted biomass values with NIR spectrometer. In another study, the same authors [54] have implemented an EKF for the same process. The authors have reported accurate predicted values in both studies; however there is no comparison between the two methods. Other types of the non-linear Kalman filtering method have also been reported in literature. Zhao et al. [53] have implemented a CKF for incorporating delayed measurements of biomass, substrate, and product concentration in fed-batch cultivation for penicillin production. Bavdekar et al. [47] have implemented an EnKF for overcoming delayed measurements of biomass, substrate and ethanol concentration in fed-batch cultivation of *S. cerevisiae*. Addressing the same delay problem Klockow et al. [43] complemented a ring buffer by an EKF and got satisfied results.

In order to indicate which Kalman filter extension describes the process better, numerical simulation runs are required. According to this perspective, a closer look to the presented articles indicates that most studies (31 articles) had relied on practical applications and simulation studies have been reported only 12 times.

3.2 Microorganism

Regarding the type of microorganism, the articles show that the majority of the research has focused on applying the Kalman filter or its extensions for state or parameter estimation during the cultivation of *S. cerevisiae* (19 articles) and *E. coli* (7 articles). The importance of these microorganisms for the biopharmaceutical industry is widely recognized, as *E. coli* and *S. cerevisiae* are the most important host microorganism used to produce recombinant proteins [58]. In addition, *S. cerevisiae* is also widely used for the production of the backers yeast as well as wine and beer. Only a few articles demonstrate state estimation in the cultivation process of other microorganisms. For instance, some authors have implemented state estimation methods for prediction of substrate and product concentration during cultivations of *Candida utilis* [30], *Penicillium chrysogenum* [46, 53] and *Kluyveromyces marxianus* [34].

3.3 Cultivation Mode

From an operational point of view, cultivation of microorganisms can be performed in batch, fed-batch and continuous modes. In fed-batch cultivation modes, set point control of the substrate concentration by manipulating the input flow rate is a matter of particular economic and scientific interest. In order to have an efficient control system, sufficient knowledge about the process state variables is required, which can be achieved by the state estimation methods such as the Kalman filter or its extensions. Therefore, previous studies have almost exclusively focused on the application of state estimation methods for fed-batch cultivations (34 publications). However, online monitoring and estimation of state variables in batch cultivations is also crucial in order to monitor the state and if necessary may improve it to achieve high productivity over the process. For instance, controlling the level of dissolved oxygen (DO) in the fermentation broth, effects the rate of microbial metabolism. Accordingly, Lee et al. [19] have implemented an EKF for noise filtering of dissolved oxygen measurements which were used for controlling the DO levels in batch cultivation of *E. coli*. This approach and, more generally, online monitoring and state estimation of variables in batch cultivations remain briefly addressed in the literature.

3.4 Bioprocess Phase

Mixing of medium and pre-cultures are performed during upstream processing phase and separation and purification of the product from biomass is performed during the downstream processing phase. In order to optimize cell growth and maximize the product yield, online monitoring and a tight control is required during both phases. The presented articles show there have been numerous studies to investigate the application of state estimation methods during the cultivation phase (39 papers). However, the articles indicate that only two authors had examined the application of Kalman filtering methods for state and variable estimation in downstream processing. For efficient and robust process development in the downstream processing phase, knowledge of the location and concentration of the product and key contaminants is also crucial. Holwill et al. [28] have used a low technology detection system involving the measurement of rate of change of absorbance at a single wavelength after addition of reagent to a representative sample stream. This provided online data detailing the performance of a continuous precipitation process. This information as well as a mathematical model which describes the fractional protein perception were fed into a control algorithm which was programmed to maintain predefined set points by feedback control through adjustments to the overall feed saturation. The Kalman filter was used for estimating the parameters of the model. Feidl et al. [59] developed a state estimation procedure for estimation

of antibody concentration by combining information coming from kinetic model and a Raman analyser, in the frame of an extended Kalman filter approach (EKF).

3.5 Measurement Device

An overview of measurement devices that are appropriate for the operation of bioprocesses is presented by Sonnleitner [61]. More specific details of different types of sensors and their measurement principles can be found in literature [62, 63]. The literature presented indicate that in *E. coli* cultivation, most authors have employed DO and CO_2 measurements from the exit gas or glucose measurements using flow injection analysis as the measurement in the Kalman filter algorithm. On the other hand, in *S. cerevisiae* cultivations, besides DO, CO_2 and glucose measurements, biomass measurements have also been widely applied. For example, Dewasme et al. [48] applied biomass measurements for their KF during an *E. coli* cultivation.

3.6 Process Model

According to the articles presented, the general mass balance equations are the most common mathematical approach used for describing the process in state observing algorithms. An overview of typical models applied to bioprocesses is presented by Chhatre [64]. A wide variety of growth kinetics are developed for modelling of particular bioprocesses. The Monod growth model [65] is the most applied method for calculating the growth kinetics of microorganisms; it corresponds to a rational function in which the specific growth rate μ is only a function of a single limiting substrate concentration and is subjected to substrate saturation when $S \gg K_s$.

$$\mu = \mu_{max} \frac{S}{K_s + S} \tag{18}$$

where μ_{max} is the maximum specific growth rate, K_s is the Monod half-saturation constant, and S is the concentration of the limiting substrate. In the mentioned articles, all of the authors, which were growing *S. cerevisiae* and *E. coli,* have implemented the Monod growth kinetics. A modified Monod model was applied by Patnaik [35, 38] which is described in detail by Henson and Seborg [66] or Jones and Kompala [67]. Application of other methods for calculating the growth kinetics such as the Contois growth model [68] has also been reported. A feature of the Contois growth model is that growth rate depends upon the concentrations of both substrate and cell mass with the consequence that an inhibition is present at high cell concentrations. This growth kinetic has been implemented in a process model describing the growth behaviour of *Penicillium chrysogenum* in fed-batch

cultivations. A modified Contois model was applied by Jianlin et al. [48] and Zhao et al. [53] in an UKF and CKF algorithm for biomass and substrate prediction, respectively. The growth rate can also be represented by artificial neural networks. However this kind of models is not applied often in combination with a KF. Zorzetto and Wilson [27] have applied a hybrid model in an EKF algorithm which is based on the theory of limited respiratory with using artificial neural network for predicting the growth rates during fed-batch cultivation of *S. cerevisiae*.

Most of the process models which are reported in literature and are used in the Kalman filter algorithms are considered to be ideal stirred tank reactors, whereas production-scale operations are corrupted by noise. This problem is more sever in large-scale operations than in laboratory-scale fermentations [35]. This can describe why all applications of state estimation methods presented in Table 1 are performed in laboratory-scale bioreactors (most cultivations are performed in a 2–5 L bioreactor and one cultivation [57] have been performed in a 22 L bioreactor).

4 An Extended Kalman Filter for the Monitoring of a Yeast Cultivation

The integration of gas sensor array data in a non-linear state estimator has not been discussed previously in the literature. Yousefi-Darani et al. [69] have designed and implemented a model-based calibrated gas sensor array for online measurement of ethanol concentration in batch cultivation with the yeast *S. cerevisiae*. However the predicted values are only available every 5 min. Therefore in this work, in order to have continues values of ethanol concentration as well as the values of biomass, glucose and the maximal growth rates, we have implemented an EKF. In addition, the whole estimation producer could be considered as a digital twin of the baker's yeast batch cultivation process, which could be used for process optimization and control.

4.1 The Cultivation Process

The cultivation of *Saccharomyces cerevisiae* (fresh baker's yeast, Oma's Ur-Hefe) was carried out in a 2.5 L bioreactor (Minifors, Infors HT, Bottmingen, Switzerland) with a vessel of stainless steel working volume of 1.35 L equipped with a temperature (set point of 30°C) and pH (set point pH $= 5$) control unit. The aeration and agitation rates were kept constant at 3.5 L min^{-1} and 500 rpm, respectively. For the pre-culture, 5 g of the baker's yeast was suspended into 100 mL medium containing 0.34 g L^{-1} MgSO$_4$·7H$_2$O, 0.42 g L^{-1} CaCl$_2$·2H$_2$O, 4.5 g L^{-1} (NH$_4$)$_2$SO$_4$, 1.9 g L^{-1} (NH$_4$)$_2$HPO$_4$, 0.9 g L^{-1} KCl. The inoculation was performed after 10 min of shaking. The same medium supplemented with glucose to a final concentration of
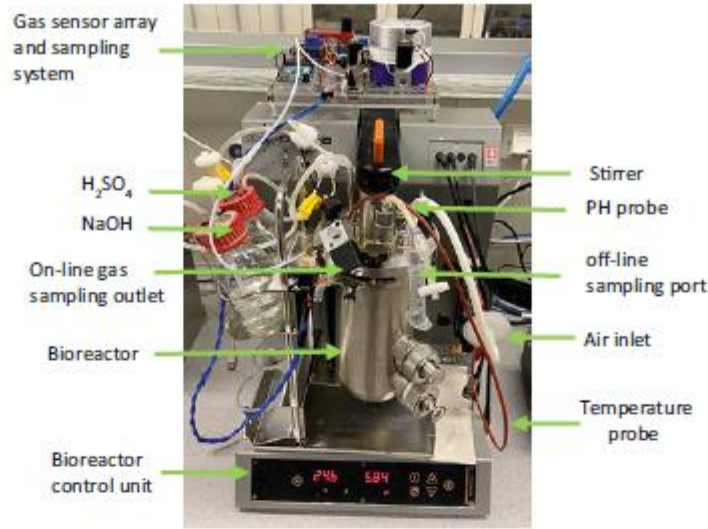
Fig. 2 Overview of the experimental setup

5 g L^{-1} as well as 1 mL L^{-1} trace elements solution (0.015 g L^{-1} FeCl$_3$·6H$_2$O, 9 mg L^{-1} ZnSO$_4$·7H$_2$O, 10.5 mg L^{-1} MnSO$_4$·2H$_2$O, and 2.4 mg L^{-1} CuSO$_4$ 5H$_2$O) and 1 mL L^{-1} vitamin solution (0.06 g L^{-1} myoinositol, 0.03 g L^{-1} Ca-pantothenate, 6 mg L^{-1} thiamine HCl, 1.5 mg L^{-1} pyridoxine HCl, and 0.03 mg L^{-1} biotin) was used for the cultivation. The experimental setup is presented in Fig. 2.

4.2 EKF Algorithm

The EKF uses discrete measurements of ethanol from the gas sensor array and estimates continuous online values of ethanol, biomass and glucose concentrations as well as the maximal growth rates in *S. cerevisiae* batch cultivation. A detailed description of the working principle of the EKF is presented in Sect. 2.2.

The EKF was implemented using the software Matlab® 2019a (version 9.6.0); the "Symbolic Math" toolbox (version 8.3) was used to calculate the estimation error covariance differential equation matrix (25 equations). For all calculations, a normal office PC (Intel Core® i5 8,500 with 8 GiB of RAM) with Window 10 was used. For the simulation, the system of in total 30 (5 + 25) differential equations was solved numerically using the explicit, Runge–Kutta-based ode45 method from Matlab. The Matlab code can be found in the appendix.

4.3 Online Ethanol Measurements

The online ethanol measurements were performed in a self-developed system equipped with commercially available metal oxide semiconductor (MOS) gas sensors (TGS 822, TGS 813 and MQ3). The sensors were located in a measuring chamber with a volume of 250 mL and operated in two cycles: a measurement cycle and a washing cycle. During the measurement cycle, the headspace gas was pumped into the measurement chamber for 10 s at a flow rate of 400 mL min^{-1} with a diaphragm pump (Schwarzer Precision, Essen, Germany). Then the chamber was flushed by pure oxygen for regeneration. A peak-shaped measurement signal is obtained, which was evaluated by using a chemometric model, which is described in detail in the literature [69]. Therefore, every 5 min a new ethanol measurement value is used by the Kalman filter. Figure 3 presents a schematic diagram of the online ethanol measurement system and the EKF for continuous state variables and parameter estimation.

Note that the EKF was carried out after the experiments were performed. The results, however, carry over to a true online application where the data is not analysed or modified in retrospect.
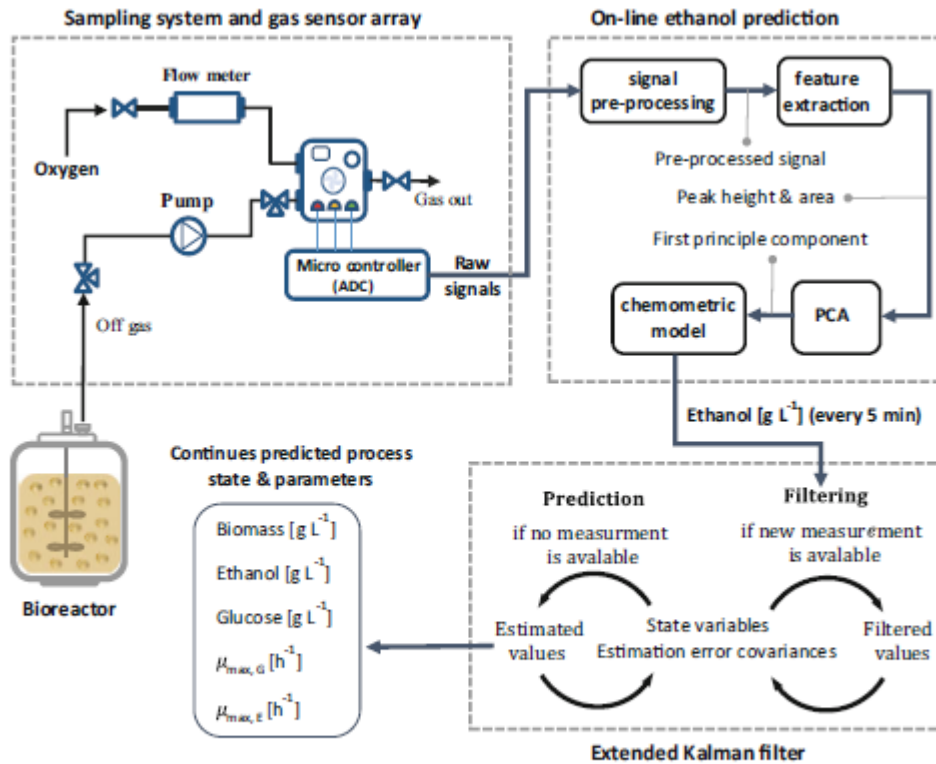
Fig. 3 Schematic diagram of the online ethanol measurement system and the EKF for continuous state variables and parameter estimation

4.4 Offline Measurements

For offline analysis, samples were regularly taken from the bioreactor and placed in pre-weighed and pre-dried micro centrifuge tubes. For biomass determination, the sample without supernatant were dried for 24 h at 103°C and after cooling for 30 min weighed. Using the filtrated supernatant (pore size filter, 0.45 µm, polypropylene membrane, VWR, Darmstadt, Germany), glucose and ethanol were determined by HPLC (ProStar, Variant, Walnut Creek, CA, USA); injection of 20 µL into a Rezex ROA-organic acid H+ (8%) column (Phenomenex, Aschaffenburg, Germany) and operated at 70°C with 5 mM H_2SO_4 as an eluent at 0.6 mL min^{-1} flow rate; software GalaxieTM Chromatography (Varian, Walnut Creek, CA, USA). The offline values were not used during the estimation of the state variables and are only taken to show that the estimates are accurate.

4.5 State Equations of the Cultivation Process

As bioreactor an ideal stirred tank reactor was assumed. As state variables, the biomass, glucose and ethanol concentrations as well as the maximal specific growth rate on glucose and ethanol were applied. Therefore, the following state equations are obtained:

$$\frac{d}{dt}\begin{bmatrix} X \\ G \\ E \\ \mu_{max,G} \\ \mu_{max,E} \end{bmatrix} = \begin{bmatrix} (\mu_G + \mu_E)X \\ -\dfrac{\mu_G}{Y_{GX}}X \\ \left(\dfrac{\mu_G}{Y_{GE}} - \dfrac{\mu_E}{Y_{EX}}\right)X \\ 0 \\ 0 \end{bmatrix} \tag{19}$$

were μ_G and μ_E are given as

$$\mu_G = \frac{\mu_{max,G} \cdot G}{K_G + G} \tag{20}$$

$$\mu_E = \frac{\mu_{max,E} \cdot E}{K_E + E} \cdot \left(1 - \frac{\mu_G}{\mu_{max,G}}\right)^2 \tag{21}$$

As one can see from the state equation, the Kalman filter is used to estimate the maximum specific growth rate on glucose $\mu_{max,\,G}$ and on ethanol $\mu_{max,\,E}$. The importance of the specific growth rate for the assessment of a cultivation is discussed by Galvanauskas et al. [70].

Table 2 Parameter values used for the simulation model

Parameter	Value	Description
K_G	$0.1\ \text{gL}^{-1}$	Monod constant glucose
K_E	$0.1\ \text{gL}^{-1}$	Monod constant ethanol
Y_{GX}	$0.17\ \text{gg}^{-1}$	Conversion factor glucose to biomass
Y_{GE}	$0.46\ \text{gg}^{-1}$	Conversion factor glucose to ethanol
Y_{EX}	$0.6\ \text{gg}^{-1}$	Conversion factor ethanol to biomass

Table 3 Initial conditions for the extended Kalman filter

Parameter	Value	Description
X_{t-0}	$2.4\ \text{gL}^{-1}$	Initial biomass concentration
G_{t-0}	$5.0\ \text{gL}^{-1}$	Initial glucose concentration
E_{t-0}	$0.1\ \text{gL}^{-1}$	Initial ethanol concentration
$\mu_{max,\ G}$	$0.14\ \text{h}^{-1}$	Initial maximal growth rate on glucose
$\mu_{max,\ E}$	$0.07\ \text{h}^{-1}$	Initial maximal growth rate on ethanol
P_{t-0}	$\begin{pmatrix} 0.02\ \text{g}^2\text{L}^2 & 0 & 0 & 0 & 0 \\ 0 & 0.02\ \text{g}^2\text{L}^2 & 0 & 0 & 0 \\ 0 & 0 & 0.02\ \text{g}^2\text{L}^{-2} & 0 & 0 \\ 0 & 0 & 0 & 0.02\ \text{h}^{-2} & 0 \\ 0 & 0 & 0 & 0 & 0.02\ \text{h}^{-2} \end{pmatrix}$	Initial estimation error covariance matrix

The extension to the ordinary Monod model for μ_E is applied, so that the transformation from glucose consumption to ethanol consumption is modelled. In Tables 2, 3, and 4 the parameters of the model as well as the initial values for the state equations and the initial values of the estimation error covariance are presented.

The Matlab code as well as the measured off- and online data of this example can be found in the appendix.

4.6 Results

In Fig. 4 the online and offline measured values of ethanol, the offline measured values of biomass and glucose as well as all the Kalman filter estimated values of all three bioprocess variables can be seen.

Figure 4 indicates the typical diauxic growth pattern of baker's yeast on glucose is obtained. First the glucose is consumed and biomass and ethanol are produced,

Table 4 Estimated measurement noise and process noise as well as measurement model for the EKF

Parameter	Value	Description
R	$0.0225\ \mathrm{g^2L^{-2}}$	Measurement noise variance
Q	$\begin{pmatrix} 0.001\ \mathrm{g^2L^2h^{-1}} & 0 & 0 & 0 & 0 \\ 0 & 0.001\ \mathrm{g^2L^2h^{-1}} & 0 & 0 & 0 \\ 0 & 0 & 0.001\ \mathrm{g^2L^{-2}h^{-1}} & 0 & 0 \\ 0 & 0 & 0 & 0.005\ \mathrm{h^{-3}} & 0 \\ 0 & 0 & 0 & 0 & 0.005\ \mathrm{h^{-3}} \end{pmatrix}$	Process noise covariance matrix
z	$(0\ 0\ 1\ 0\ 0)$	Measurement matrix (just ethanol is measured)
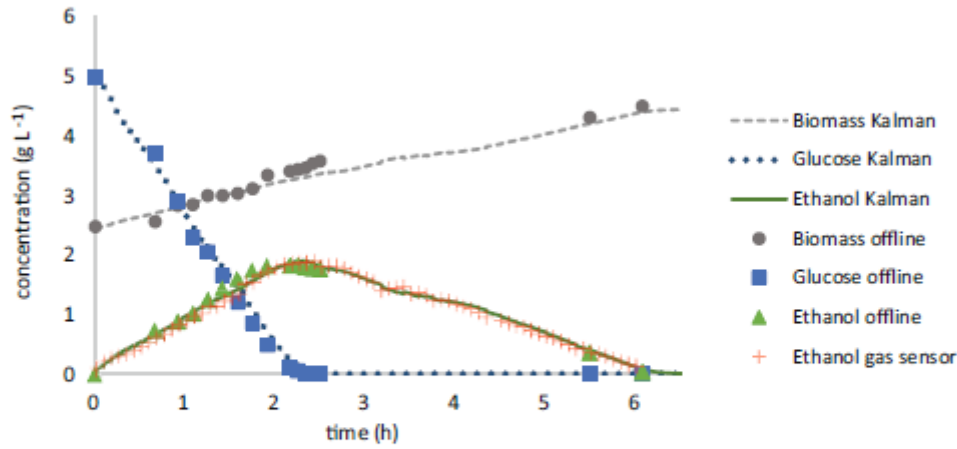
Fig. 4 Online and offline values for biomass, glucose and ethanol as well as EKF estimates for these values

**Table 5** Prediction error of EKF values compared to offline measurements

	Glucose	Ethanol	Biomass
RMSEP	0.12 g L^{-1}	0.14 g L^{-1}	0.12 g L^{-1}
Error	5.6%	2.8%	6.2%
R^2	0.96	0.99	0.97

then ethanol is converted to biomass. The offline measurements and its corresponding estimated values fit quite well together as can be seen in Table 5.

The root mean squared error of prediction (RMSEP) of glucose is 0.12 g L^{-1}. The ethanol offline values during glucose consumption are mostly higher than the online measured and the predicted ones; in overall their RMSEP is 0.14 g L^{-1}. All ethanol online measurements seems to be a little bit shifted in time compared to the offline values, which might indicate the time delay due to gas transport from the fermentation broth through the headspace of the reactor to the measurement system. The biomass has a RMSEP of 0.12 g L^{-1}, but the highest deviation can be seen shortly after ethanol is used as substrate. The values shortly before ethanol consumption might not be predicted accurately, because the model describing the switching from glucose to ethanol might be suboptimal.

In order to investigate the influence of the measurement frequency on the performance of the EKF, we decreased the measurement frequency of the online ethanol measurements to one per hour. The results of the estimated values with the EKF are presented in Fig. 5.

Still the overall behaviour of the estimated values is the same. However, the sampling frequency has an influence on the corrections of the estimated state during filtering. Larger step changes are observed in the estimated values whenever a new measurement is available. However, even if the sampling frequency is changed to one per hour, the overall behaviour is predicted well.
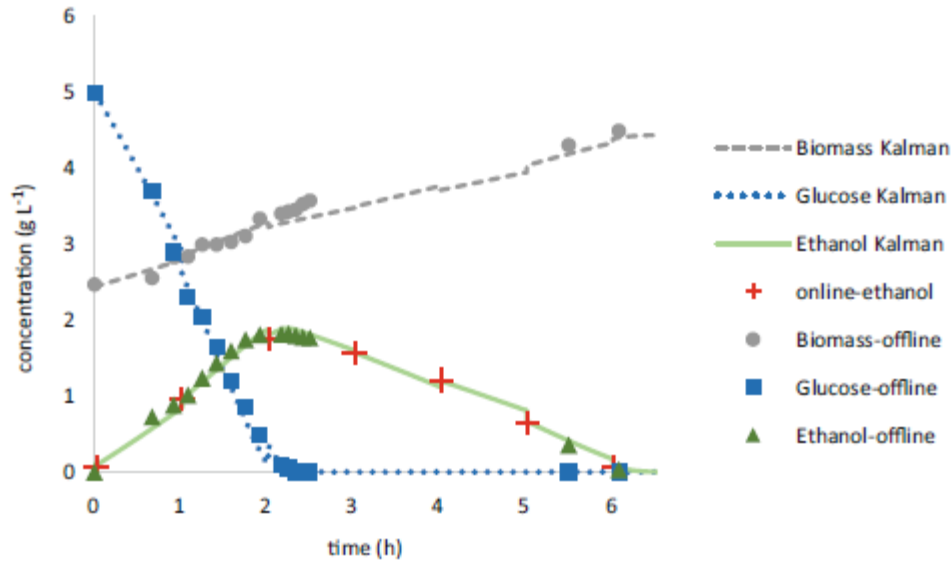
Fig. 5 Online (every 1 h) and offline values for biomass, glucose and ethanol as well as EKF estimates for these values

Obviously with a higher sampling frequency, these step changes are smaller. Nevertheless, with a 5 min sampling time, the EKF was able to follow the true states of the system with a reasonably small error. More detailed information about the influence of the sampling frequency on the accuracy of the Kalman filter estimates can be found in literature [71, 72].

The EKF was also used for predicting the specific growth rates and their maximum values.

In Fig. 6 the estimated maximum specific growth rates with respect to glucose $\mu_{max,\ G}$ and ethanol $\mu_{max,\ E}$ as well as specific growth rates itself (μ_G and μ_E for glucose and ethanol respectively) are presented.

After inoculation, the specific growth rate and its maximum value with respect to glucose are increasing from 0.14 h^{-1} to more than 0.18 h^{-1}. However shortly thereafter they decrease again. This indicates the high sensitivity of the estimation values due to the measurement noise variance R and the process noise variance with respect to $\mu_{max,\ G}$, which is Q [4]. The smaller the R and the higher the Q [4], the more the estimated values will rely upon the measurements and as a consequence the filtered values might be changed, if the measured and estimated values deviate from each other. The more glucose is consumed, the larger will be the difference of $\mu_{max,\ G}$ and μ_G, due to the Monod growth kinetics. If the glucose is almost depleted, the extension to the Monod model on ethanol contributes to increasing growth on ethanol. Shortly after 2 h cultivation time, the transition from glucose to ethanol as substrate takes place. The maximum specific growth rate on ethanol $\mu_{max,\ E}$, which has not changed during the growth on glucose starts to increase. According to the typical Monod behaviour, before ethanol is depleted, due to the low substrate concentration, $\mu_{max,E}$ should be almost constant while μ_E should be increasing.
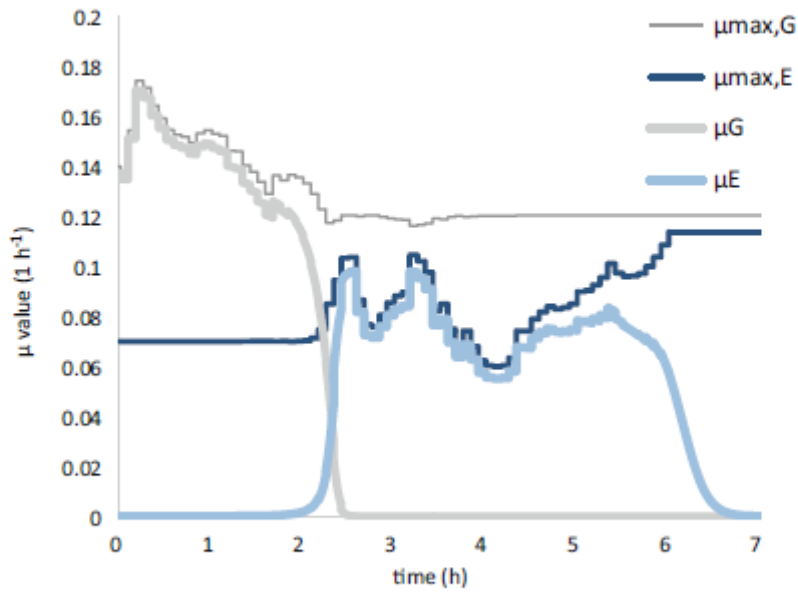
Fig. 6 Estimated maximum specific growth rates with respect to glucose $\mu_{max, G}$ and ethanol $\mu_{max, E}$ as well as the specific growth rates (μ_G and μ_E for glucose and ethanol, respectively)

However this is not observed in Fig. 6 which is due to the fluctuation of the measured and estimated ethanol concentration.

5 Conclusion

In this chapter, the working principles as well as an overview of Kalman filter applications for state and parameter estimation in bioprocesses has been presented. Regarding the type of the Kalman filter, since most biotechnical processes are non-linear, non-linear versions of the Kalman filter, specifically the EKF, are the most applied algorithm among other extensions of the Kalman filter. However the UKF is getting attention in recent years. The results in literature indicate that the UKF algorithms deliver more accurate estimates of the parameters and state variables compared to EKF algorithms.

In spite of the apparent success of Kalman filters for state and parameter estimation in lab-scale bioreactors, the integration of Kalman filters into industrial systems is not very widespread while most of the process models mentioned in literature consider noise-free ideal fermentations, whereas production-scale operations are corrupted by concentration gradients and disturbance. Accordingly, more efforts are required towards performing simulation studies in order to model and validate proper mathematical models associated with complex non-ideal bioprocesses.

Despite the numerous examples on state estimation methods for biotechnological processes in literature, the research on implementing Kalman filters for state

estimation in downstream processing remain rather limited. The advancement in state and parameter estimation methods in downstream processes leads to better knowledge of the location and concentration of the product and key contaminants, which are essential for process optimization and control.

So far most of the Kalman filter algorithms are implemented for monitoring fed-batch cultivations; however more attention is required for real-time implementation of the Kalman filter algorithms for controlling the feed rate and substrate production in these cultivations. Further efforts are also required towards implementation of state estimation methods in batch and continuous cultivations.

From the presented literature, it could be concluded that the non-linear extensions of the Kalman filter are powerful tools for state estimation in bioprocesses; therefore they could be used for digitalization of bioprocesses. Accordingly, in a case study, a digital twin of the baker's yeast batch fermentation process was developed by using a dynamic non-linear model of the process as well as an EKF algorithm. The proposed method gives the possibility to predict glucose, ethanol and biomass concentrations simultaneously from the only available infrequent online measurements of ethanol concentration. The accuracy of the estimated biomass and substrate production are in line with other studies which have also implemented an EKF algorithm for monitoring the baker's yeast cultivation [32, 49]. However, in our application the maximal specific growth rates on glucose and ethanol are also estimated. As a consequence, the rapid and precise estimation of these variables could increase the overall knowledge integration in the digital twin of the process.

Overall, the unique advantage of online monitoring and in general digital twins of bioprocesses is that they could play critical roles in bioprocess development such as supporting problem solving in manufacturing, reducing effort in setting up a control strategy and accelerating process performance by taking corrective actions automatically and in real time.

Appendix

Extended Kalman filter Matlab code: Online state prediction of batch yeast cultivations based on ethanol gas sensors.

```
%Initialization
clear; close all; clc;
sympref('AbbreviateOutput', false);
%Variable and parameter definition
%Symbols for symbolic math calculations
syms G E X P t                      real
syms Y_gx Y_ge Y_ex mu1 mu2 K_M_G K_M_E    real
%Variables / Parameters
initX = [2.5; 6; 0.2; 0.15; 0.08];        % initial state (Biomass,
% Glucose, Ethanol)
```

```
initP  = diag([0.1,0.02, 0.02,0.2,0.02]); % initial process estimation
% covariance matrix
init  = [initX; initP(:)];                % combined initial value
%vector
% for the odesolver
H    = [0 0 1 0 0]                        % observation matrix
H = 1×5
  0   0   1   0   0
Q = diag([0.001,0.001,0.001,0.001,0.001]) % process noise covariance
%matrix

R     = 0.05                              % measurement noise
%covariance matrix
K1    = 0.1;                              % Monod konstant glucose
K2    = 0.1;                              % Monod konstant ethanol

%estimated parameter values
Ygx    =0.15;                             % Yield glucose -> biomass
Yge    =0.34;                             % Yield glucose -> ethanol
Yex    =0.43;                             % Yield ethanol -> biomass
%Process model
%Monod terms
mue1 = mu1*G / (G+K_M_G);
mue2 = mu2*E / (E+K_M_E) * (1 - mue1/mu1);
%Model OD
dS = sym(X * [...
    ( mue1     +mue2)    ;...             % Biomass
     -mue1/Y_gx          ;...             % Glucose
    ( mue1/Y_gx*Y_ge - mue2/Y_ex) ;...    % Ethanol
    0;                                    % mue1
    0;                                    % mue2
    ]);
%Jacobian of Model with respect to state variables
F = jacobian(dS,[X,G,E,mu1,mu2])
P matrix
P = sym('P',[5,5])
dP = F * P + P*F'+Q
%Simulation / State prediction and filtering
%Replace all symbolic parameters with their respective numeric values
F = subs(F, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
dS = subs(dS, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
dP = subs(dP, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
%Assemble all differential equations into a vector of 12 elements
% (3x state, 9x P)
OdeSys = matlabFunction([dS(:);dP(:)],'Vars',{t, [X; G; E; mu1; mu2;
P(:)]});
%load measurement values from file:

load BC2_eth_pred.mat                     % Ethanol sensor
%measurements
load BC2.mat                              % Offline values for

%Simulate the process from one ethanol gas measurement time to the next:
t0 = 0;
```

```matlab
MC = zeros(0,5);                %store filtered states in these variables
SimState = zeros(0,5);
SimTime = [];
for i = 1:numel(timeE)
  tspan = [t0 timeE(i)];
  [T,state] = ode45(OdeSys, tspan, init);       % simulate / solve model

  PS        = state(end,1:5)';                   % predicted state

  MS        = ME(i);                             % measured state
  P         = reshape(state(end,6:end),5,5);     % process covariance
                                                 % matrix
  K         = P*H'/(H*P*H'+R);                   % kalman gain matrix
  FS        = PS + K * (MS-PS(3));               % filtered state
  Pfilt     = P-K*H*P ;                          % filtered process
                                                 % covariance matrix
  init      = [FS; Pfilt(:)];                    % new initial condition
  t0        = timeE(i);                          % new starting time for
                                                 % next iteration
% Save intermediate states for plotting
  MC        = [MC;FS'];
  state(end,1:3) = NaN;
  SimState   = [SimState; state(:,1:5)];
  SimTime    = [SimTime; T];
end
%Results
%Plot the results in a presentable figure and save file to disk
f = figure("Position",[0,0,1600,640]);
subplot(1,2,1);
h = plot([0,time'],[initX(1:3)';M],'.','MarkerSize',20); % Plot
%measurements
set(h, {'color'},{'r'; 'g';'b'}); hold on;
h = plot(SimTime,SimState(:,1:3));                       % Plot simulated values
set(h, {'color'}, {'r'; 'g';'b'});
plot(timeE,ME,'+b','MarkerSize',8); hold off; % Plot ethanol gas sensor
%values
ax = gca;
ax.FontSize = 14;
ax.FontName = 'Times';
ax.Position = [.05 .1 .4 .85];
ax.ActivePositionProperty = 'outerposition';
ax.GridLineStyle =':';
ax.GridAlpha = .7;
xlabel('time $/h$','interpreter','Latex',"FontSize",16);
ylabel('concentration $/\frac{g}{L}
$','interpreter','Latex',"FontSize",16); ylim([0 8]);
grid on; box off; grid(gca, 'minor');
legend('Biomass offline','Glucose offline','Ethanol offline','Biomass
Kalman','Glucose Kalman','Ethanol Kalman','Ethanol gas
sensor','interpreter','Latex',"FontSize",12,"Color",[.9 .9 1 .9]);
subplot(1,2,2);
h = plot(SimTime,SimState(:,4:5));            % Plot mu values over time
ax = gca;
```

```
ax.FontSize = 14;
ax.FontName = 'Times';
ax.Position = [.55 .1 .4 .85];
ax.ActivePositionProperty = 'outerposition';
ax.GridLineStyle =':';
ax.GridAlpha = .7;
ytickformat('%.2f')
set(h,{'color'},{'r'; 'k'});
xlabel('time $/h$','interpreter','Latex',"FontSize",16);
ylabel('$\mu$ value $/\frac{1}{h}
$','interpreter','Latex',"FontSize",16);
grid on; box off; grid(gca,'minor');
legend('$\mu_1$','$\mu_2
$','interpreter','Latex',"FontSize",12,"Color",[.9 .9 1]);
annotation("arrow",[.55 .97],[.1 .1])
annotation("arrow",[.05 .47],[.1 .1])
annotation("arrow",[.05 .05],[.1 .98])
annotation("arrow",[.55 .55],[.1 .98])
saveas(f,'KalmanPred.svg','svg');            % save copy of figure to file

%Calculate Errors
SimTime = SimTime + ((1:numel(SimTime))*1e-10)';
SimValues = interp1(SimTime,SimState(:,1:3),time);
SSE = sum((SimValues - M).^2);
RMSE = sqrt(SSE/numel(time));
SQT = sum((M-mean(M)).^2);
RSq = 1-SSE./SQT;
T1 = table('Size',[3,3],'VariableTypes',
{'double','double','double'},'VariableNames',
{'Biomass','Glucose','Ethanol'},'RowNames',{'RMSEP','Error
%','R²'});
T1(1,:) = num2cell(RMSE);
T1(2,:) = num2cell(SSE./(max(M)-min(M))*100);
T1(3,:) = num2cell(RSq)
```

References

1. Zobel-Roos S, Schmidt A, Mestmäcker F, Mouellef M, Huter M, Uhlenbrock L, Kornecki M, Lohmann L, Ditz R, Strube J (2019) Accelerating biologics manufacturing by modeling or: is approval under the QbD and PAT approaches demanded by authorities acceptable without a digital-twin? PRO 7(2):94

2. Kuchemüller KB, Pörtner R, Möller J (2020) Digital twins and their role in model-assisted design of experiments. In: Advances in biochemical engineering/biotechnology. Springer, Berlin. https://doi.org/10.1007/10_2020_136

3. Nargund S, Guenther K, Mauch K (2019) The move toward biopharma 4.0: In silico biotechnology develops "smart" processes that benefit biomanufacturing through digital twins. Genet Eng Biotechnol News 39(6):53–55

4. Luttmann R, Bracewell DG, Cornelissen G, Gernaey KV, Glassey J, Hass VC, Kaiser C, Preusse C, Striedner G, Mandenius C-F (2012) Soft sensors in bioprocessing: a status report and recommendations. Biotechnol J 7:1040

5. Schügerl K, Bellgardt KH (2012) Bioreaction engineering: modeling and control. Springer, Berlin

6. Schügerl K (2001) Progress in monitoring, modeling and control of bioprocesses during the last 20 years. J Biotechnol 85(2):149–173

7. Narayanan H, Luna MF, von Stosch M, Bournazou MNC, Polotti G, Morbidelli M, Butté A, Sokolov M (2019) Bioprocessing in the digital age - the role of process models. Biotechnol J 761. https://doi.org/10.1002/biot.201900172

8. Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:S.35–S.45

9. Wan EA, van der Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium (Cat. No. 00EX373). IEEE, pp 153–158

10. Matthews M (1990) A state-space approach to adaptive nonlinear filtering using recurrent neural networks. In: Proceedings IASTED Internat. Symp. artificial intelligence application and neural networks

11. Boulet G, Kerr Y, Chehbouni A, Kalma JD (2002) Deriving catchment-scale water and energy balance parameters using data assimilation based on extended Kalman filtering. Hydrol Sci J 47(3):449–467

12. Krämer S, Grum M, Verworn HR, Redder A (2005) Runoff modelling using radar data and flow measurements in a stochastic state space approach. Water Sci Technol 52(5):1–8

13. Williams M, Schwarz PA, Law BE, Irvine J, Kurpius MR (2005) An improved analysis of forest carbon dynamics using data assimilation. Glob Chang Biol 11(1):89–105

14. Annan JD, Hargreaves JC, Edwards NR, Marsh R (2005) Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. Ocean Model 8(1–2):135–154

15. Julier SJ, Uhlmann JK (1997) New extension of the Kalman filter to nonlinear systems. In: Signal processing, sensor fusion, and target recognition VI, vol 3068. International Society for Optics and Photonics, pp 182–193

16. Evensen G (1994) Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. J Geophys 99(C5):10.143–10.162

17. Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman filter technique. Mon Weather Rev 126(3):796–811

18. van der Merwe R (2004) Sigma-point Kalman filters for probabilistic inference in dynamic state-space models. Doctoral dissertation, OGI School of Science and Engineering at OHSU

19. Lee SC, Hwang YB, Chang HN, Chang YK (1991) Adaptive control of dissolved oxygen concentration in a bioreactor. Biotechnol Bioeng 37(7):597–607

20. Ghoul M, Dardenne M, Fonteix C, Marc A (1991) Extended Kalman filtering technique for the on-line control of OKT3 hybridoma cultures. Biotechnol Tech 5(5):367–370

21. Dubach AC, Märkl H (1992) Application of an extended kalman filter method for monitoring high density cultivation of Escherichia coli. J Ferment Bioeng 73(5):396–402

22. Gudi R, Shah S (1993) The role of adaptive multirate Kalman filter as a software sensor and its application to a bioreactor. IFAC Proc 26(2):249–254

23. Gudi R, Gray I, Shah S (1993) Multi-rate estimation and monitoring of process variables in a bioreactor. In: Proceedings of IEEE international conference on control and applications, IEEE

24. Albiol J, Robusté J, Casas C, Poch M (1993) Biomass estimation in plant cell cultures using an extended Kalman filter. Biotechnol Prog 9(2):174–178

25. Gudi RD, Shah SL, Gray MR (1995) Adaptive multirate state and parameter estimation strategies with application to a bioreactor. AICHE J 41(11):2451–2464

26. Petrova M, Georgieva O, Patarinska T (1995) State and time delay estimation of continuous microorganisms cultivation. Bioprocess Eng 12(1–2):103–107

27. Zorzetto L, Wilson J (1996) Monitoring bioprocesses using hybrid models and an extended Kalman filter. Comput Chem Eng 20:S689–S694

28. Holwill IJ, Chard SJ, Flanagan MT, Hoare M (1997) A Kalman filter algorithm and monitoring apparatus for at-line control of fractional protein precipitation. Biotechnol Bioeng 53(1):58–70

29. Hrnčiřík P, Náhlík J, Havlena V (1998) State estimation of Baker's yeast fed-batch cultivation by extended Kalman filter using alternative models. IFAC Proc 31(11):601–606

30. Ganovski L, Bliznakova M, Patarinska T (1999) State estimation of a Uricase production process with Candida utilis. Bioprocess Eng 21(3):273–277

31. Bogaerts P (1999) A hybrid asymptotic-Kalman observer for bioprocesses. Bioprocess Eng 20 (3):249–255

32. Hitzmann B, Broxtermann O, Cha Y-L, Sobieh O, Stärk E, Scheper T (2000) The control of glucose concentration during yeast fed-batch cultivation using a fast measurement complemented by an extended Kalman filter. Bioprocess Eng 23(4):337–341

33. Arndt M, Hitzmann B (2001) Feed forward/feedback control of glucose concentration during cultivation of Escherichia coli. IFAC Proc 34(5):403–407

34. Longhi L, Marcon S, Trierweiler J, Secchi A (2002) State estimation of an experimental bioreactor using the extended Kalman filtering technology. IFAC Proc 35(1):379–382

35. Patnaik P (2003) On the performances of noise filters in the restoration of oscillatory behavior in continuous yeast cultures. Biotechnol Lett 25(9):681–685

36. Cha Y-L, Hitzmann B (2004) Ultrasonic measurements and its evaluation for the monitoring of Saccharomyces cerevisiae cultivation. Bioautomation 1:16–29

37. Arndt M, Hitzmann B (2004) Kalman filter based glucose control at small set points during fed-batch cultivation of Saccharomyces cerevisiae. Biotechnol Prog 20(1):377–383

38. Arndt M, Kleist S, Miksch G, Friehs K, Flaschel E, Trierweiler J, Hitzmann B (2005) A feedforward–feedback substrate controller based on a Kalman filter for a fed-batch cultivation of Escherichia coli producing phytase. Comput Chem Eng 29(5):1113–1120

39. Patnaik PR (2005) The extended Kalman filter as a noise modulator for continuous yeast cultures under monotonic, oscillating and chaotic conditions. Chem Eng J 108(1–2):91–99

40. Rocha I, Veloso AC, Ferreira E (2006) Design of estimators for specific growth rate control in a fed-batch E. coli fermentation

41. Henry O, Kamen A, Perrier M (2007) Monitoring the physiological state of mammalian cell perfusion processes by on-line estimation of intracellular fluxes. J Process Control 17 (3):241–251

42. Soons Z, Shi J, Van der Pol L, Van Straten G, Van Boxtel A (2007) Biomass growth and kLa estimation using online and offline measurements. IFAC Proc 40(4):85–90

43. Klockow C, Hüll D, Hitzmann B (2008) Model based substrate set point control of yeast cultivation processes based on FIA measurements. Anal Chim Acta 623(1):30–37

44. Veloso AC, Rocha I, Ferreira E (2009) Monitoring of fed-batch E. coli fermentations with software sensors. Bioprocess Biosyst Eng 32(3):381–388

45. Jianlin W, Liqiang Z, Tao Y (2010) On-line estimation in fed-batch fermentation process using state space model and unscented Kalman filter. Chin J Chem Eng 18(2):258–264

46. Jianlin W, Xuying F, Liqiang Z, Tao Y (2010) Unscented transformation based robust kalman filter and its applications in fermentation process. Chin J Chem Eng 18(3):412–418

47. Bavdekar VA, Prakash J, Patwardhan SC, Shah SL (2011) Moving window ensemble Kalman filter for delayed and multi-rate measurements. IFAC Proc 44(1):11997–12002

48. Dewasme L, Goffaux G, Hantson A-L, Wouwer AV (2013) Experimental validation of an Extended Kalman Filter estimating acetate concentration in E coli cultures. J Process Control 232(2013):148–157

49. Popova S, Ignatova M, Lyubenova V (2013) State and parameters estimation by extended Kalman filter for studying inhomogeneous dynamics in industrial bioreactors

50. Sbarciog M, Coutinho D, Wouwer AV (2014) A simple output-feedback strategy for the control of perfused mammalian cell cultures. Control Eng Pract 32:123–135

51. Fernandes S, Richelle A, Amribt Z, Dewasme L, Bogaerts P, Wouwer AV (2015) Extended and unscented Kalman filter design for hybridoma cell fed-batch and continuous cultures. IFAC-Papers 48(8):1108–1113
52. Dewasme L, Fernandes S, Amribt Z, Santos LO, Bogaerts P, Wouwer AV (2015) State estimation and predictive control of fed-batch cultures of hybridoma cells. J Process Control 30:50–57
53. Zhao L, Wang J, Yu T, Chen K, Liu T (2015) Nonlinear state estimation for fermentation process using cubature Kalman filter to incorporate delayed measurements. Chin J Chem Eng 23(11):1801–1810
54. Krämer D, King R (2016) On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by S. cerevisiae. IFAC-Papers 49(7):609–614
55. Simutis R, Lübbert A (2017) Hybrid approach to state estimation for bioprocess control. Bioengineering 4(1):21
56. Krishna VV, Pappa N, Rani SJV (2018) Implementation of embedded soft sensor for bioreactor on Zynq processing system. In: 2018 international conference on recent trends in electrical, control and communication (RTECC), IEEE
57. Krämer D, King R (2019) A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. J Process Control 82:91–104
58. Ritschel TK, Boiroux D, Nielsen MK, Huusom JK, Jørgensen SB, Jørgensen JB (2019) The extended Kalman filter for nonlinear state estimation in a U-loop bioreactor. In: 2019 IEEE conference on control technology and applications (CCTA), IEEE
59. Feidl F, Garbellini S, Luna MF, Vogg S, Souquet J, Broly H, Butté A (2019) Combining mechanistic modeling and Raman spectroscopy for monitoring antibody chromatographic purification. PRO 7(10):683
60. Lisci S, Grosso M, Tronci S (2020) A geometric observer-assisted approach to tailor state estimation in a bioreactor for ethanol production. PRO 8(4):480
61. Sonnleitner B (2013) Automated measurement and monitoring of bioprocesses: key elements of the M 3 C strategy. In: Measurement, monitoring, modelling and control of bioprocesses. Springer, Berlin, pp 1–33
62. Biechele P, Busse C, Solle D, Scheper T, Reardon K (2015) Sensor systems for bioprocess monitoring. Eng Life Sci 15(5):469–488
63. Vojinović V, Cabral JMS, Fonseca LP (2006) Real-time bioprocess monitoring: part I: in situ sensors. Sensors Actuators B Chem 114(2):1083–1091
64. Chhatre S (2012) Modelling approaches for bio-manufacturing operations. In: Measurement, monitoring, modelling and control of bioprocesses. Springer, Berlin, pp 85–107
65. Monod J (1949) The growth of bacterial cultures. Annu Rev Microbiol 3(1):371–394
66. Henson MA, Seborg DE (1992) Nonlinear control strategies for continuous fermenters. Chem Eng Sci 47(4):821–835
67. Jones KD, Kompala DS (1999) Cybernetic model of the growth dynamics of Saccharomyces cerevisiae in batch and continuous cultures. J Biotechnol 71(1–3):105–131
69. Contois DE (1959) Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures. Microbiology 21(1):40–50
69. Yousefi-Darani A, Paquet-Durand O, Babor M, Hitzmann B (2020) Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in *Saccharomyces cerevisiae batch cultivation*. Biosyst Eng 198(2020):198–209
70. Galvanauskas V, Simutis R, Levišauskas D, Urniežius R (2019) Practical solutions for specific growth rate control systems in industrial bioreactors. PRO 7(10):693
71. Oisiovici RM, Cruz SL (2000) State estimation of batch distillation columns using an extended Kalman filter. Chem Eng Sci 55(20):4667–4680
72. Hashemi R, Engell S (2016) Effect of sampling rate on the divergence of the extended Kalman filter for a continuous polymerization reactor in comparison with particle filtering. IFAC-Papers 49(7):365–370

4. 3. Parameter and state estimation of backers yeast cultivation with a gas sensor array and unscented Kalman filter.

By Abdolrahimahim Yousefi-Darani, Olivier Paquet-Durand, Jörg Hinrichs, Bernd Hitzmann

Engineering
in Life Sciences

RESEARCH ARTICLE

Parameter and state estimation of backers yeast cultivation with a gas sensor array and unscented Kalman filter

Abdolrahimahim Yousefi-Darani[1] | Olivier Paquet-Durand[1] | Jörg Hinrichs[2] |
Bernd Hitzmann[1]

[1] Department of Process Analytics and Cereal Science, University of Hohenheim, Stuttgart, Germany

[2] Department of Soft Matter Science and Dairy Technology, University of Hohenheim, Stuttgart, Germany

Correspondence
Abdolrahim Yousefi-Darani, Department of process analytics and cereal science, University of Hohenheim, Garbenstr. 23, 70599, Stuttgart, Germany.
Email: rahim@uni-hohenheim.de

This article is dedicated to Professor Thomas Bley on the occasion of his 70th birthday.

Abstract

Real-time information about the concentrations of substrates and biomass is the key to accurate monitoring and control of bioprocess. However, on-line measurement of these variables is a challenging task and new measurement systems are still required. An alternative are software sensors, which can be used for state and parameter estimation in bioprocesses. The software sensors predict the state of the process by using mathematical models as well as data from measured variables. The Kalman filter is a type of such sensors.

In this paper, we have used the Unscented Kalman Filter (UKF) which is a nonlinear extension of the Kalman filter for on-line estimation of biomass, glucose and ethanol concentration as well as for estimating the growth rate parameters in *S. cerevisiae* batch cultivation, based on infrequent ethanol measurements. The UKF algorithm was validated on three different cultivations with variability of the substrate concentrations and the estimated values were compared to the off-line values.

The results obtained showed that the UKF algorithm provides satisfactory results with respect to estimation of concentrations of substrates and biomass as well as the growth rate parameters during the batch cultivation.

KEYWORDS
batch cultivation, bioprocess supervision, ethanol, state estimation, Unscented Kalman filter

1 | INTRODUCTION

The ability to measure primary process variables, such as biomass, substrate and product concentrations is essential in order to guarantee the successful operation and automatic control of bioprocesses at their optimal state. But direct on-line measurements of these biological state variables are often not possible due to the lack of cheap or reliable measuring devices. In fact, in many practical applications, only some of the state variables involved are available for on-line measurement. Therefore, the development of methodologies, namely software sensors which can provide accurate estimation of process variables that are not measurable in real time, is of great interest [2–4].

Abbreviations: EKF, Extended Kalman filter; *S. cerevisiae*, *Saccharomyces cerevisiae*; UKF, unscented Kalman filter

76

The classical Kalman filter and its nonlinear extensions are a type of such software sensors, which have received a huge interest for state estimation of bioprocesses. In general, the Kalman filters combine the information of a process model and the available measurements for state and parameter estimation. As Harvey [5] pointed out, in the classical Kalman filter, both the process model and the measurement equation of the state space model are linear. However, most bioprocesses are highly nonlinear, therefore the classical Kalman filter cannot be used for state estimation in such processes. Different nonlinear extensions of the Kalman filter are available which mostly differ in the approximation of the prediction uncertainty. The extended Kalman filter (EKF) is a standard nonlinear estimation technique which can handle nonlinearity's by local first order Taylor approximations of the non-linear functions in the model. Implementation of EKF algorithms for state and parameter estimation in bioprocess have been reported in numerous studies. Lisci and Tronci et al. [6] have implemented an extended Kalman filter for state estimation in fed-batch cultivation of *S. cerevisiae* based on temperature, dissolved oxygen and substrate concentration measurements. Krishna et al. [7] have implemented an EKF algorithm for estimation of lactose concentration in fed-batch cultivation of *Kluyveromyces marxianus* based on dissolved oxygen measurements; Krämer and King [8] used an EKF for estimation of substrate and biomass concentration in fed-batch cultivation of *S. cerevisiae*; Lee et al. [9] applied an EKF algorithm for noise filtering from the dissolved oxygen measurements during batch cultivation of *E. coli* and Hitzmann et al. [10] implemented an EKF complemented by a special flow-injection analysis system for glucose measurements during fed-batch cultivation of *S. cerevisiae*. Based on the estimation, a feed forward PI-control with a set point of 0.5 g/L was carried out. The mean deviation of the set point and the estimated value as well as the set point and the measured value were 0.05 and 0.11 g/L respectively. A similar approach is discussed by Klockow et al. [11] where the time delay of the measurements was compensated by a ring-buffer. They showed that a set point of 0.007 g/L can be realized reliably.

Usually the well-known EKF shows good prediction results. Nevertheless, in spite of the reported satisfactory results, it has some disadvantages. It is reliable for systems which are almost linear on the time scale of the update intervals; it requires the calculation of Jacobians at each time step, which may be difficult to obtain for higher order systems; it does linear approximations of the system at a given time instant, which may introduce errors in the estimation, leading to a state divergence over time [12, 13].

The unscented Kalman filter (UKF) is another nonlinear extension of the classical Kalman filter which is very similar to the EKF, but instead of approximating the non-linear

PRACTICAL APPLICATION

In a previous study we have designed and implemented a model-based calibrated gas sensor array for on-line measurement of ethanol concentration in batch cultivation with the yeast S. cerevisiae [1]. The obtained results indicate that the gas sensor array was able to predict ethanol concentration with high accuracy. However the predicted values are only available every five minutes. Therefore in this work, in order to have continues values of ethanol concentration as well as the values of biomass, glucose and the growth rates, we have implemented an unscented Kalman filter (UKF) algorithm.

The obtained results indicate, that accurate continues concentrations of the state variables as well as the growth rates can be obtained by the UKF algorithm. No off-line measurements for calibration are required in the proposed algorithm.

The proposed method is a cheap alternative to other tools that are used for monitoring yeast cultivations such as spectroscopy based methods.

process model by calculating the Jacobian of the dynamics for the determination of the estimation error variance, the transformed probability distributions are approximated directly. This is done by representing the distribution by a set of chosen sample points, transforming these points by the non-linear model function, and then approximating the mean and variance of the transformed distribution by the mean and variance of the transformed points [14]. In recent years, a number of authors have demonstrate the successful application of UKF algorithms for state and parameter estimation in bioprocesses. For instance, Jianlin et al. [15] have implimented an UKF algorithm for biomass and substrate prediction based on dissolved oxygen and carbon dioxide measurements in a fed-batch cultivation of *S. cerevisiae*. Using the same microorganism, Simutis and Lübbert [16] have applied an UKF algorithm for estimation of biomass and its specific growth rate based on oxygen uptake and CO_2 formation rate measurements in a fed-batch cultivation. Furthermore, Krämer and King [17] have used an UKF algorithm for filtering out noise from measured state variables which were predicted with a near infra-red spectrometer in a fed-batch cultivation of *S. cerevisiae*.

As it can be seen, previous studies have exclusively focused on implementing UKF algorithms in fed-batch cultivations, however on-line monitoring and estimation

of state variables in batch cultivations is also crucial in order to achieve high productivity over the process. Therefore, in this contribution an UKF algorithm is designed for on-line estimation of biomass, glucose and ethanol concentration as well as for estimating the growth rate parameters in *S. cerevisiae* batch cultivation, based on the infrequently available ethanol measurements. In order to evaluate the reliability of the UKF, the proposed algorithm was validated on three different cultivations with variability of the substrate concentrations.

This paper is organized as follows: In the coming section of this work, the experimental setup and the cultivation conditions, the dynamic model of *S. cerevisiae* batch cultivation and a brief description of the on-line ethanol measurement method as well as the unscented Kalman filter are described. In section 3 results and discussion is presented, and section 4 concludes this paper.

2 | MATERIALS AND METHODS

2.1 | Batch cultivation process

Three batch cultivations of *S. cerevisiae*, named BC1, BC2, and BC3, were performed. *S. cerevisiae* (fresh baker's yeast, Oma's Ur-Hefe) was pre-cultivated before fermentation. 5 g of baker's yeast was used in all cultivations. The baker's yeast was inoculated into 100 mL Schatzmann medium [18] and after shaking for 10 min, they were added into the stainless steel tank bioreactor (Minifors, Infors HT, Bottmingen, Switzerland). The medium used for batch cultivations was the same as for the pre-culture, but with 9 g/L, 5 g/L and 2.85 g/L glucose for BC1, BC2, and BC3 respectively and 1 mL/L trace elements solution. All three batch runs were operated at the same conditions, that is, a constant temperature at 30°C and a maintained pH at 5. The aeration and agitation rates were kept constant at 3.5 L/min and 450 rpm, respectively. Detailed experimental conditions of the cultivations are described by Yousefi-Darani et al. [1].

2.2 | Nonlinear process model

For modelling the process, an ideal stirred tank reactor in batch mode has been assumed with a cell growth kinetic approximated by the Monod model, where the substrate glucose as well as ethanol (when glucose is depleted) are the single growth-limiting factor. According to the mass balance, the dynamic process model consists of the following equations [19]:

$$\frac{dX}{dt} = \mu_G X + \mu_E X \tag{1}$$

$$\frac{dG}{dt} = -\frac{\mu_G X}{Y_{X/G}} \tag{2}$$

$$\frac{dE}{dt} = \frac{\mu_G X}{Y_{E/G}} - \frac{\mu_E X}{Y_{X/E}} \tag{3}$$

were G, E and X are the glucose, ethanol and the biomass concentrations, respectively. $Y_{X/G}$, $Y_{E/G}$ and $Y_{X/E}$ are the yield coefficients with respect to the conversion from glucose to biomass, glucose to ethanol and ethanol to biomass, respectively. μ_G and μ_E are the specific growth rates on glucose and ethanol, respectively and are calculated as

$$\mu_G = \frac{\mu_{max,\,G} \cdot G}{K_G + G} \tag{4}$$

$$\mu_E = \frac{\mu_{max,\,E} \cdot E}{K_E + E} \cdot \left(1 - \frac{\mu_G}{\mu_{max,\,G}}\right)^2 \tag{5}$$

$\mu_{max,\,G}$ and $\mu_{max,\,E}$ are the maximum specific growth rates on glucose and on ethanol, respectively. The values for $\mu_{max,\,G}$ and $\mu_{max,\,E}$ are estimated with the UKF filtering.

Therefore they are described with two additional differential terms, which do not alter the values:

$$\frac{d\mu_{max,\,G}}{dt} = 0 \tag{6}$$

$$\frac{d\mu_{max,\,E}}{dt} = 0 \tag{7}$$

In this contribution, the yield coefficients as well as the Monod equation's constants (K_E and K_G) have been fixed to values which were chosen from literature [1, 20].

2.3 | On-line ethanol measurements

The on-line ethanol measurements were performed in a self-developed measurement system which contains two main parts, namely the headspace sampling system and the measurement chamber. Headspace sampling procedure consisted of an automated sequence of internal operations. First the head space samples of the bioreactor are pumped past the measurement chamber for 10 s at a flow rate of 400 mL min⁻¹ with a diaphragm pump (Schwarzer Precision, Essen, Germany) every five minutes. The measurement chamber has a volume of 250 mL and contains a gas sensor array which is equipped with commercially available metal oxide semiconductor (MOS) gas sensors (TGS 822, TGS 813 and MQ3). In the next step the chamber is flushed by pure oxygen for regeneration. By preforming these steps, a peak shaped measurment signal is obtained.
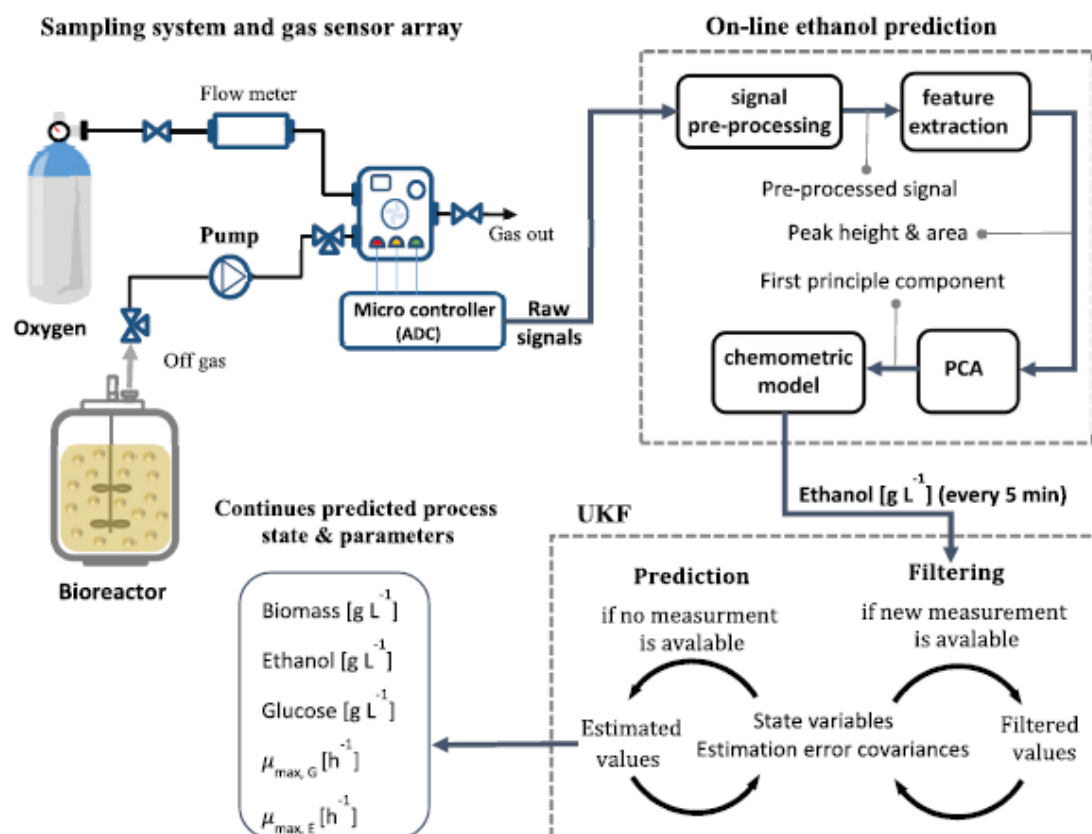
Sampling system and gas sensor array

On-line ethanol prediction

FIGURE 1 The operation scheme of the on-line ethanol measurement system and the UKF algorithm for continuous state variables and parameter estimation

Ethanol concnetration is obtained from the raw signals by implementing signal processing methods and a chemometric model, which is described in detail in the literature [1]. Using the ethanol measurement in the gas phase, the ethanol concentration in the liquid phase of the cultivation broth is determined. Every 5 min a new ethanol measurement value is sent to the UKF algorithm. As measurement model the identity is used, i. e. the ethanol value itself. The operation scheme of the on-line ethanol measurement system and the UKF algorithm is presented in Figure 1.

2.4 | Unscented Kalman filter

In this work, an UKF is implemented to estimate continuous ethanol, glucose and biomass concentrations as well as the maximal growth rates during *S. cerevisiae* batch cultivation. As all Kalman filter approaches, the UKF calculates the most probable system state by appropriately weighing

model predictions and actual measurements according to model uncertainty and measurement error, respectively. The UKF was chosen here over other extensions of the Kalman filter since it accurately calculates the statistical distributions of even nonlinear systems.

The UKF algorithms consist of two steps namely the prediction step (time update) and the filtering step (measurement update) which are summarized as follows:

- **Prediction step (time update):**

    Using the last known state $\hat{X}_k$ the process model is used to predict the state variables X_k until the next measurement z_k is available.

    For the Kalman filtering to work, the state covariance P_k must also be estimated somehow from the last known state covariance $\hat{P}_k$. In the UKF the unscented transformation is used to estimate this state covariance.

The idea is to use a collection (2n+1, where n is the number of state variable) of well-chosen system states [22], based on the last known state and state covariance and then propagate/predict the system state from each of those so called sigma points. The predicted system state X_k is then the weighted average of these 2n+1 predictions. The estimated covariance P_k is essentially the weighted covariance of the same predictions.

- **Correction step (measurement update):**

Whenever a measurement is available, the predicted state values X_k are combined with the measured values z_k to provide corrected or filtered state estimates $\hat{X}_k$. For this reason the Kalman gain matrix K_k must be calculated:

$$K_k = \frac{P_k \cdot H^T}{H \cdot P_k \cdot H^T + R} \qquad (10)$$

Here R is the measurement error, H is the Jacobian matrix of the measurement model h ( ) and P_k is the already mentioned estimated state covariance matrix of the prediction. The corrected or filtered system state $\hat{X}_k$ is essentially the weighted average of the predicted system state and the measurement with the Kalman gain as weight:

$$\hat{X}_k = X_k + K_k \cdot (z_k - h(X_k)) \qquad (11)$$

The filtered state covariance $\hat{P}_k$ is updated based on the estimated state covariance P_k, the Kalman gain K_k and the process noise covariance matrix Q:

$$\hat{P}_k = P_k - (K_k \cdot H \cdot P_k) + Q \cdot \Delta t - (K_k \cdot H \cdot Q \cdot \Delta t) \quad (12)$$

Δt is the time difference to the last known state/ measurement.

The filtered values $\hat{X}_k$ and $\hat{P}_k$ can then be used as initial conditions for the next prediction/ simulation of the process as well as for the estimation of the covariance until the next measurement is obtained and everything repeats again.

The reliability and quality of a Kalman filter estimator can be evaluated by observabilty analyses, the theory of which has been established previously [21, 22]. Observabilty analysis provides an assessment of the theoretical possibility of estimating the state variables or parameters of the system from the available measurements (sensors) [23]. Due to Salau et al. [24], if the idea is to use the smallest number of sensors in order to simultaneously estimate the state and parameters of a system, using the Kalman

filter makes it impossible to find complete system observability. For instance, in the process considered here, the two additional differential equations for parameter estimation of $\mu_{max,G}$ and $\mu_{max,E}$ produces a Jacobian matrix with corresponding row elements equal to null and therefore the observability is not given. However, due to the chosen process model with low correlation of the state variables, the observabilty of the process is guaranteed, since in this case, diagonal time-invariant matrices Q and R can be successfully applied, which makes the UKF tuning considerably simple. A more comprehensive description about implementation of the UKF algorithm as well as the observabilty analysis can be found in [25–28].

In this study, a continuous-discrete UKF is used, i.e., a continuous time update and a discrete-time measurement update. Table 1 presents the initial values for the UKF filter as well as the parameters of the model.

The UKF was implemented using the software Matlab R2019a (version 9.6.0); for all calculations a normal office PC (Intel CoreR i5 8500 with 8 GiB of RAM) with Window 10 was used. For the simulation, the system of in total 5 differential equations was solved numerically using the explicit, Runge-Kutta based ode45 method from Matlab.

2.5 | Off-line measurements

Samples for analysing the concentrations of biomass, glucose, and ethanol were regularly taken from the bioreactor and put into pre-weighed and pre-dried micro-centrifuge tubes. Cell dry weight was determined by centrifugation (Universal 16 R, Hettich Zentrifugen GmbH & Co. KG, Tuttlingen, Germany) of a sample with 1.5 mL (2 times) at 14,000 rpm for 10 min at 4°C. The wet cells were let in a drying oven at 103°C for 24 h. Subsequently, they were cooled down for 30 min before weighing. The supernatant of the samples after the centrifugation was examined by HPLC (ProStar, Variant, Walnut Creek, CA, USA) to determine the glucose and ethanol concentrations. Firstly, the supernatant was filtrated with pore size filter, 0.45 μm, polypropylene membrane (VWR, Darmstadt, Germany), then 20 μL was injected into a Rezex ROA-organic acid H+ (8%) column (Phenomenex, Aschaffenburg, Germany) and operated at 70°C with 5 mM H_2SO_4 as an eluent at 0.6 mL/min flow rate. The concentrations of glucose and ethanol were calculated by Galaxie software (Varian, Walnut Creek, CA, USA).

In order to evaluate the performance of the UKF algorithm with respect to the accuracy of predicting ethanol concentrations, the root-mean square error (RMSE) between the predicted ethanol concentration with the

TABLE 1 Initial conditions for the mathematical model as well as the unscented Kalman filter

Parameter	Description	Value					
$X_{t=0}$	Initial biomass concentration	BC1	2.48 g/L	BC2	2.44 g/L	BC3	2.6 g/L
$G_{t=0}$	Initial glucose concentration	BC1	9 g/L	BC2	5 g/L	BC3	2.85 g/L
$E_{t=0}$	Initial ethanol concentration	BC1	0.1 g/L	BC2	0.18 g/L	BC3	0.25 g/L
$\mu_{max,G}$	Initial maximal growth rate on glucose	BC1	$0.16\ h^{-1}$	BC2	$0.18\ h^{-1}$	BC3	0.14 g/h
$\mu_{max,E}$	Initial maximal growth rate on ethanol	BC1	$0.007\ h^{-1}$	BC2	$0.004\ h^{-1}$	BC3	$0.008\ h^{-1}$
$Y_{X/G}$	conversion from glucose to biomass	$0.175\ g_X\ g_G^{-1}$					
$Y_{E/G}$	conversion from ethanol to glucose	$0.473\ g_E\ g_G^{-1}$					
$Y_{X/E}$	conversion from biomass to ethanol	$0.598\ g_X\ g_E^{-1}$					
K_G	Monod constant for glucose	0.01 g/L					
K_E	Monod constant for ethanol	0.01 g/L					
α	Unscented transformation constant	$1e^{-3}$					
n	Unscented transformation constant	0					
β	Unscented transformation constant	2					
R	Measurement noise variance	$0.0225\ g^2 L^{-2}$					
Q	Process noise covariance matrix	$\text{diag}\,(0.001\ g^2 L^2 h^{-1}\ 0.001\ g^2 L^2 h^{-1}\ 0.001\ g^2 L^2 h^{-1}\ 0.005\ h^{-3}\ 0.005\ h^{-3})$					
P_0	Initial process estimation covariance matrix	diag (0 0 0 0 0)					
H	Observation matrix	(0 0 1 0 0)					

UKF and the off-line ethanol concentrations was calculated and compared to the RMSE between the on-line measured ethanol concentration and off-line ethanol concentrations. Furthermore, the percentages standard error (SE) with respect to the maximum ethanol concentration, for the predicted ethanol concentrations from the UKF as well as the on-line ethanol concentrations were calculated. RMSE and SE are calculated according to the following equations:

$$MSE = \sqrt{\sum_{i=1}^{N} \frac{(\hat{Y}_i - Y_i)^2}{N}} \quad (8)$$

$$SE\ (\%) = \frac{\sqrt{\sum_{i=1}^{N} \frac{(\hat{Y}_i - Y_i)^2}{N}}}{Y_{max}} \cdot 100\% \quad (9)$$

$\hat{Y}_i$ represents the predicted ethanol concentration from the UKF algorithm or the measured on-line ethanol concentration, Y_i is the concentration determined by the off-line values, N stands for the measurement count and Y_{max} is the maximum ethanol concentration in the corresponding off-line data.

To evaluate the accuracy of the estimated biomass and glucose concentrations from the UKF algorithm, the RMSE between the predicted values and the off-line values as well as the SE with respect to the maximum off-line value were calculated.

3 | RESULTS AND DISCUSSION

The UKF presented in section 2 is used for continuous estimation of ethanol concentration on the basis of infrequent on-line ethanol measurements from the gas sensor array. The UKF algorithm was also used for estimation of biomass and glucose as well as estimating the maximum growth rate on glucose and ethanol. The algorithm was validated on three cultivations with different initial conditions. Figure 2 presents the performance of the UKF for the estimation of ethanol concentration during three cultivation runs.

Figure 2 shows the estimates of the concentrations of ethanol in the bioreactor computed using the UKF (blue curve) together with the on-line measured ethanol concentrations (grey dots) and the HPLC off-line ethanol concentrations (black squares). Note that the HPLC off-line values were not used during the estimation of the state variables, and are only taken to show that the estimates are accurate.

When a deviation of the on-line measured and estimated values are present, the estimated values are shifting to the measured ones, indicating that the measured data are not dominating the estimation. For instance, in BC2 between 2 and 3 h cultivation time, there is a significant difference between the measured values and the off-line values which can be caused by several factors including temperature fluctuations in the ethanol measurement chamber, an inappropriate determination of the base line, or electrical noise in the sensor circuit. Due to the fact that the process model fits to the off-line values, it could be
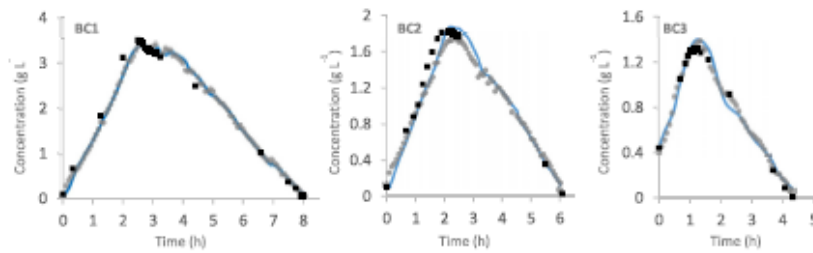
FIGURE 2 Measured ethanol concentration with the gas sensor array (grey dots), estimated ethanol concentration with UKF (blue curve) and off-line ethanol concentration (black squares) during all three cultivations (BC1 – BC3)

stated that the on-line measured values are not accurate. In this case the UKF predicted values are a compromise of the process model and the on-line measured values which leads to much more accurate predictions. The accuracy of the UKF regarding ethanol prediction was evaluated by comparing the RMSEP and SEP of the estimated (filtered) ethanol concentration with the non-filter ones (the measured ethanol concentration), and the results are presented in Table 2.

Table 1 shows the performance of the UKF compared with the case where the ethanol concentrations were measured off-line. In BC2 the RMSEP and SEP of the off-line measured ethanol concentration is considerably larger compared to the filtered ones. In BC1 and BC3 the UKF slightly increased the accuracy of the ethanol concentrations. However, the standard error of estimated ethanol concentration with the UKF during all cultivations is below 5% which is a decent value.

TABLE 2 Comparison of off-line measured values and UKF estimated ethanol concentration

Cultivation	Off-line measured values		UKF estimated values	
	RMSEP	SEP	RMSEP	SEP
BC1	0.63 g/L	5.5%	0.15 g/L	4%
BC2	0.16 g/L	9.5%	0.08 g/L	4.5%
BC3	0.08 g/L	6.5%	0.09 g/L	4.5%

The UKF is also able to accurately estimate the concentrations of biomass and glucose during the cultivations. Figure 3 presents the estimated values as well as the off-line values of biomass and glucose for BC1 – BC3. In BC1 the estimates of the biomass concentration between 6 and 8 h cultivation time, deviate from the off-line values, however their evaluation is almost the same. This might be due to faulty sample handling or measurement errors.
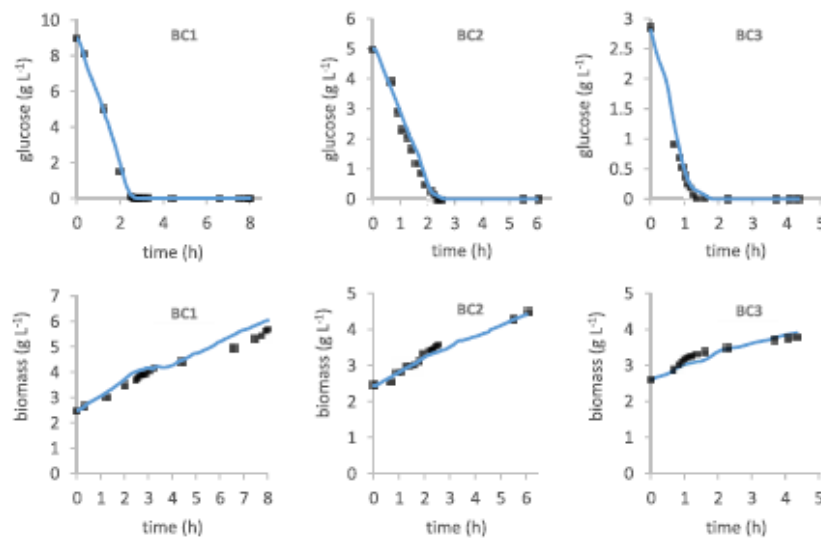
FIGURE 3 Estimated biomass and glucose concentrations with UKF (solid line) and off-line biomass and glucose concentrations (black squares) during all three cultivations (BC1 – BC3)

TABLE 3 RMSEP and SEP for glucose and biomass

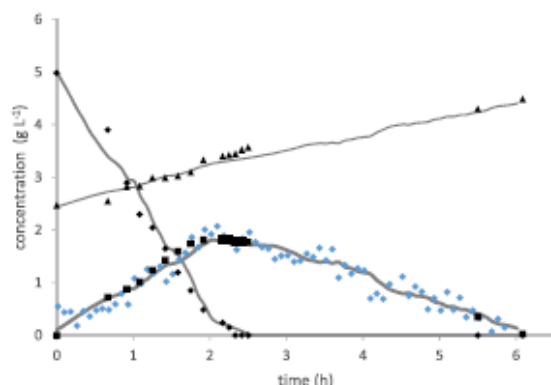| Cultivation	Biomass		Glucose	
	RMSEP	SEP	RMSEP	SEP
BC1	0.29 g/L	9%	0.13 g/L	1.7%
BC2	0.09 g/L	5%	0.16 g/L	4%
BC3	0.1 g/L	5%	0.16 g/L	4%

FIGURE 4 Artificially distorted measured ethanol concentration with random noise (●), estimated ethanol concentration with UKF (—), off-line ethanol concentration (black squares), estimated glucose concentration with UKF (—), off-line glucose concentration (black diamonds), estimated biomass concentration with UKF (—), off-line biomass concentration (black triangles) in BC2

The data in Figure 3 indicates that the typical diauxic growth pattern of baker's yeast on glucose is observed. First the glucose is consumed and biomass as well as ethanol are produced, then ethanol is converted to biomass. The off-line measurements and its corresponding estimated values fit quite well together as can be seen in Table 3.

As illustrated above, the UKF is able to accurately predict the state variables in all three cultivations, based on the available infrequent on-line ethanol measurements. However, as indicated in Figure 2 and Table 2, the measured ethanol concentration is not noisy, therefore an accurate estimation of the state variables by the UKF is not far from expectation. Therefore, in order to check the prediction ability of the UKF algorithm when the measurement data are noisy, the measured ethanol values were artificially distorted by random noise and the UKF algorithm was performed.

Figure 4 shows the estimated as well as the off-line values of state variables of BC2, by the UKF with considering noisy ethanol measurements. As it can be seen, even if the measurement values are artificially distorted by random noise, the UKF does not show much different results in predicting the state variables (SEP for all state variables is below 6%). The results for the other two cultivation data records are qualitatively the same, and are thus not repeated here.

As already stated previously, the UKF algorithm was also used for estimating the maximal specific growth rates. To prove the capability of the UKF for estimating the maximal specific growth rates, different starting values were chosen for these parameters in each cultivation. These values are chosen according to a rough estimate of these parameters.

Figure 5 presents the estimated maximum specific growth rates with respect to glucose $\mu_{\max,G}$ and ethanol $\mu_{\max,E}$ as well as specific growth rates itself for glucose μ_G and ethanol μ_E across all three cultivations.

In BC1 the $\mu_{\max,G}$ and μ_G are increasing sharply short after the inoculation starts, this indicates that the chosen starting values are lower than the actual values, therefore the UKF algorithm converges to the true values. When the glucose is almost depleted, the transition from glucose to ethanol as substrate takes place, therefore $\mu_{\max,E}$ and μ_E would start to increase. However, shortly before glucose is completely depleted, $\mu_{\max,G}$ increases which results in the decrease of μ_E, therefore the UKF increases the $\mu_{\max,E}$ and μ_E to compensate the under estimation of μ_E. According to the typical Monod behaviour, before ethanol is depleted, due to the low substrate concentration, $\mu_{\max,E}$ should be almost constant while μ_E should be increasing. However this is not observed in BC1 which is due to the fluctuation of the measured and estimated ethanol concentration which can be seen between 4 and 7 h of cultivation time.

In BC2, after inoculation, the specific growth rates and its maximum values with respect to glucose are increasing slightly. This indicates that the chosen starting values are not far from the actual values. Accordingly, in BC3, the specific growth rates and its maximum values with respect to glucose are decreasing after the inoculation and shortly thereafter they increase again. This indicates that the chosen starting values are lower than the actual values, nevertheless the UKF algorithm converges them to reasonable values.

The high sensitivity of the estimated values due to the measurement noise variance and the process noise variance can be observed by comparing the estimated growth rates from BC1 to BC3. In BC2 and BC3, the measured and estimated ethanol concentrations shows less fluctuation compared to the ones from BC2, therefore the estimated values for the specific growth rates show less fluctuating compared to the ones from BC1.

In Figure 6 the estimation error variances of the process variables and the maximal specific growth rate with respect to glucose and ethanol are presented. The initial values of the variances are all set to zero. All values seem to be very much reproducible with respect to all three cultivation runs. During the glucose phase the
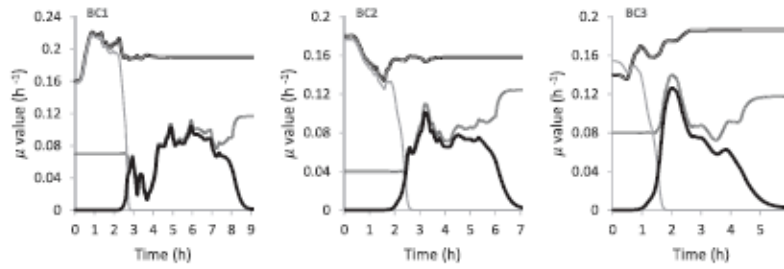
FIGURE 5 Estimated maximum specific growth rates with respect to glucose $\mu_{max,G}$ (■) and ethanol $\mu_{max,E}$(—) as well as the specific growth rates μ_G (■) and μ_E (—) for glucose and ethanol respectively

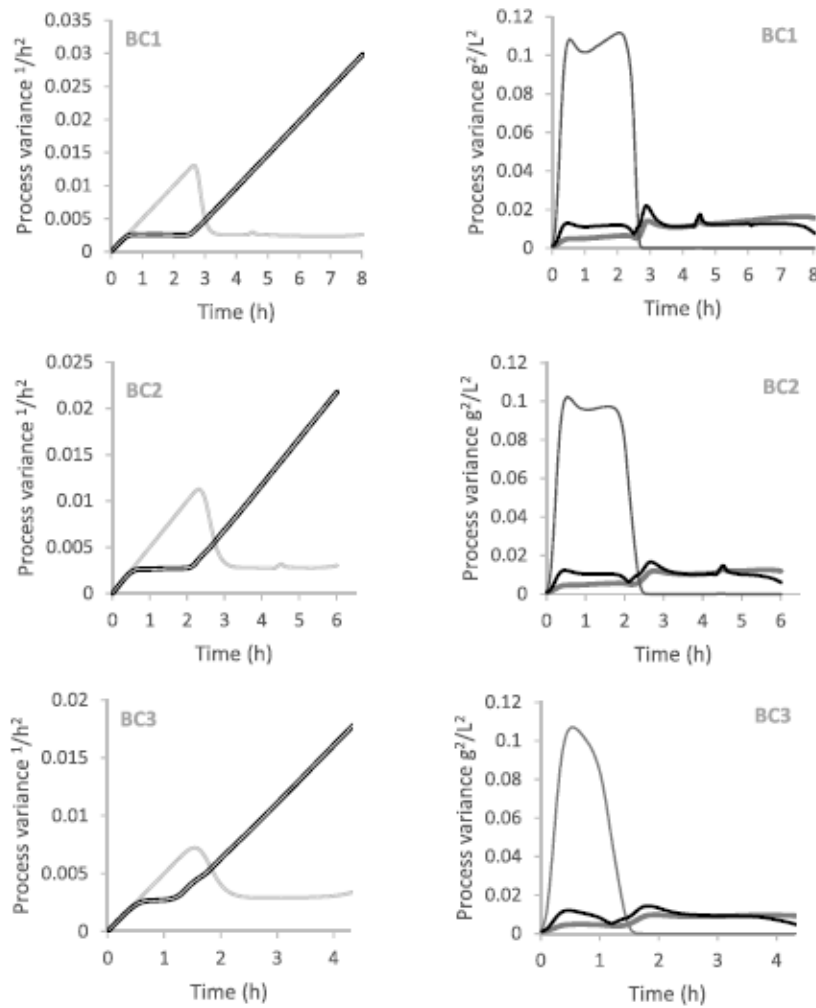**FIGURE 6** Estimated error variance of the maximal specific growth rates ($\mu_{max,G}$ (■) and $\mu_{max,E}$ (—)) and estimated error variance of the process variables (biomass (—), glucose (—) and ethanol (—))

estimation error variance values of glucose are much higher compared to the ones of biomass and ethanol. The values are roughly one order of magnitude higher. The reason is, that the ration of the change of ethanol and glucose with time during the glucose phase is the ration of the yield coefficients Y_{GX}/Y_{GE} whose value is 0.37 g g^{-1}. Therefore, the change in glucose is much higher (2.7 times) which causes a larger error and a larger variance. If one considers the variance of glucose as 0.1 g$_2$ L^{-2} and ethanol as 0.01 g$_2$ L^{-2} then the square root is 0.316 g/L and 0.1 g/L respectively. If one calculate the ration of error of ethanol by the error of glucose 0.1/0.316 then almost the same values is obtained as the ration of the yield coefficients. During the ethanol phase the estimation error variance of glucose become small, because no increase in ethanol can be detected and therefore no glucose is present. The variance of biomass and ethanol are stable throughout the cultivation. The variances of the of maximal growth rate on glucose are increasing fast to a constant value during the glucose phase and are increasing constantly during ethanol phase. The corresponding values with respect to ethanol behave in the same manner but inverted. If no measurement information of growth on the substrate is present, the estimation error variance is just increasing constantly, however the variance decreased clearly when growth occurred on that substrate. The constant values as well as the slope during increasing are the same.

4 | CONCLUDING REMARKS

The design of monitoring and control algorithms to improve the performance of bioprocesses is of major importance. However, it is often difficult to find inexpensive and robust commercially available sensors that allow real-time monitoring of important process variables, such as the biomass and substrate concentrations as well as the growth rates. Therefore, the development of software sensors is of paramount importance.

In this work, a dynamic nonlinear model was used and an unscented Kalman filter algorithm was implemented for parameter and state estimation during *S. cerevisiae* batch cultivaton. The proposed UKF algorithm only requires on-line data from infrequent ethanol measurements together with the yield coefficients of the process model.

Three batch cultivations with different initial conditions were conducted in order to analyse the behaviour and performance of the UKF. The result obtained showed that with the proposed UKF algorithms, it was possible to estimate the specific growth rates as well as continuous ethanol, glucose and biomass concentrations with great

accuracy, during the cultivation process. In order to check the quality margin for estimation with respect to presented noise in the measured on-line ethanol values, a simulation was preformed; as a conclusion, the UKF algorithm is still able to predict the parameters and state variables, if the noise is about less than 10%.

The proposed UKF algorithm can be used for comprehensive monitoring of the baker's yeast batch fermentation process as well as for design and implementation of control strategies for the fed-batch fermentation process of the baker's yeast.

CONFLICT OF INTEREST
The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT
The data used to support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Olivier Paquet-Durand https://orcid.org/0000-0002-2479-543X

REFERENCES
1. Yousefi-Darani, A., Babor, M., Paquet-Durand, O., Hitzmann, B., Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in Saccharomyces cerevisiae batch cultivation. Biosystems Engineering. 2020, 198, 198–209.
2. Dochain, D., Bastin, G., Adaptive control of fedbatch bioreactors. Chemical Engineering Communications. 1990, 87, 67–85.
3. Dochain, D., State and parameter estimation in chemical and biochemical processes: a tutorial. Journal of process control. 2003, 13, 801–818.
4. Bernard, O., Chachuat, B., Hélias, A., Rodriguez, J., Can we assess the model complexity for a bioprocess: theory and example of the anaerobic digestion process. Water science and technology. 2006, 53, 85–92.
5. Harvey, A.C., Forecasting, structural time series models and the Kalman filter. Cambridge university press, 1990.
6. Lisci, S., Grosso, M., Tronci, S., A Geometric Observer-Assisted Approach to Tailor State Estimation in a Bioreactor for Ethanol Production. Processes. 2020, 8, 480.
7. Krishna, V. V., Pappa, N., Rani, S. J. V., Implementation of Embedded Soft Sensor for Bioreactor on Zynq Processing System. 2018 International Conference on Recent Trends in Electrical, Control and Communication (RTECC), IEEE.
8. Krämer, D., King, R., On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by S. cerevisiae. IFAC-PapersOnLine. 2016, 49, 609–614.

9. Lee, S.C., Hwang, Y.B., Chang, H.N., Chang, Y.K., Adaptive control of dissolved oxygen concentration in a bioreactor. Biotechnology and bioengineering. 1991, 37, 597–607.

10. Hitzmann, B., Broxtermann, O., Cha, Y.-L., Sobieh, O. et al., The control of glucose concentration during yeast fed-batch cultivation using a fast measurement complemented by an extended Kalman filter. Bioprocess Engineering. 2000, 23, 337–341.

11. Klockow, C., Hüll, D., Hitzmann, B., Model based substrate set point control of yeast cultivation processes based on FIA measurements. Analytica chimica acta. 2008, 623, 30–37.

12. Julier, S. J., Uhlmann, J. K., New extension of the Kalman filter to nonlinear systems. In Signal processing, sensor fusion, and target recognition VI. 1997, (Vol. 3068, pp. 182–193). International Society for Optics and Photonics.

13. Van Der Merwe, R., *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models. 2004*, (Doctoral dissertation, OGI School of Science & Engineering at OHSU).

14. Julier, S., Uhlmann, J., Durrant-Whyte, H. F., A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Transactions on automatic control. 2000, 45, 477–482.

15. Jianlin, W., Xuying, F., Liqiang, Z., Tao, Y., Unscented transformation based robust kalman filter and its applications in fermentation process. Chinese Journal of Chemical Engineering. 2010, 18, 412–418.

16. Simutis, R., Lübbert, A., Hybrid approach to state estimation for bioprocess control. Bioengineering. 2017, 4, 21.

17. Krämer, D., King, R., A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. Journal of Process Control. 2019, 82, 91–104.

18. Schatzmann, H., *Anaerobes Wachstum von Saccharomyces cerevisiae: regulatorische Aspekte des glycolytischen und respirativen Stoffwechsels. 1975*, (Doctoral dissertation, ETH Zurich).

19. Solle, D., Geissler, D., Stärk, E., Scheper, T. et al., Chemometric modelling based on 2D-fluorescence spectra without a calibration measurement. Bioinformatics. 2003, 19, 173–177.

20. Paquet-Durand, O., Assawarajuwan, S., Hitzmann, B. (2017). Artificial neural network for bioprocess monitoring based on fluorescence measurements: Training without offline measurements. Eng. Life Sci., 17, 874–880.

21. Kou, S. R., Elliott, D. L., Tarn, T. J., Observability of nonlinear systems. Information and Control. 1973, 22, 89–99

22. Hermann, R. and Krener, A., Nonlinear controllability and observability. IEEE Transactions on automatic control. 1977, 22, 728–740.

23. Golabgir, A., Hoch, T., Zhariy, M., Herwig, C., Observability analysis of biochemical process models as a valuable tool for the development of mechanistic soft sensors. Biotechnology progress. 2015, 31, 1703–1715

24. Salau, N. P., Trierweiler, J. O., Secchi, A. R., Observability analysis and model formulation for nonlinear state estimation. Applied Mathematical Modelling. 2014, 38, 5407–5420.

25. Julier, S. J. (2002, May). The scaled unscented transformation. In *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)* (Vol. 6, pp. 4555–4559). IEEE.

26. Kandepu, R., Foss, B., Imsland, L., Applying the unscented Kalman filter for nonlinear state estimation. Journal of process control. 2008, 18, 753–768.

27. Bogaerts, P., Wouwer, A. V., Parameter identification for state estimation—application to bioprocess software sensors. Chemical engineering science. 2004, 59, 2465–2476.

28. Goffaux, G., Wouwer, A. V., Bioprocess state estimation: some classical and less classical approaches. Control and observer design for nonlinear finite and infinite dimensional systems, pp. 111–128. Springer, Berlin, Heidelberg, 2005.

4. 4. Application of fuzzy logic control for the dough proofing process.

By Abdolrahimahim Yousefi-Darani, Olivier Paquet-Durand, Bernd Hitzmann. Published in Food and Bioproducts Processing. Volume 115, pages 36-46, February 2019.

ELSEVIER

Application of fuzzy logic control for the dough proofing process

Abdolrahimahim Yousefi-Darani*, Olivier Paquet-Durand, Bernd Hitzmann

Department of Process Analytics and Cereal Science, University of Hohenheim, Stuttgart, Germany

ABSTRACT

A fuzzy logic control system is designed and applied to a proofing system. The controller is experimentally evaluated and the performance is compared to the ones from a PID controller. Dough pieces with different amounts of yeast added in the ingredients and in different temperature starting states are prepared and proofed with the supervision of the fuzzy control system. The controller is designed to maintain the volume of the dough pieces similar to volume expansion of a dough piece in standard conditions during the proofing process. Controllers are evaluated by means of performance criteria and the final volume of the dough samples. It is demonstrated that the fuzzy logic controller can provide significant better control, does not require a mathematical model and has better disturbance rejection properties.

© 2019 Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

1. Introduction

The proofing (fermentation) process of dough is one of the quality-determining steps in the production of baking goods. Beside the fluffiness, whose fundaments are established during fermentation, the flavour of the final product is influenced very much during this production stage. One of the prime aspects in industrial baking enterprises is to control the proofing process in order to achieve a predictable final volume of dough pieces at the end of this process step.

Popular strategies for process control consist of classical control techniques and advance control techniques. Classical control techniques are mainly represented by the proportional-integral-derivative (PID) controller. The greatest hindrance to the widespread use of classical control strategies in the process industry is the need for a detailed mathematical model for the control system and the difficulty to copy its realization from one application to another. This makes the introduction of such strategies quite expensive in such

fields as the food industry (Lahtinen, 2001). Application of a PID controller for controlling the fermentation process in bread making has been previously studied by the authors (Yousefi-Darani et al., 2018), however the complexity of the mathematical model of the fermentation process, illustrates the elaborateness of implementing the controller.

On the other hand, advanced process control techniques target at nonlinearities and high-order dynamics which are ubiquitous in many processes in the food industry. Thus, it is often necessary to employ advanced food process control techniques for these processes (Kondakci and Zhou, 2017).

Advanced control techniques might be classified into the following three main categories: model-based control, fuzzy logic control, and artificial neural network-based control (Jean-Pierre., 2004; Romagnoli and Palazoglu, 2005).

Fuzzy logic controller is a rule-based approach which uses variable formulated rules such as IF (condition) and THEN (conclusion). Fuzzy logic controller brings a systematic methodology to convert expert knowledge into a heuristic control algorithm (Perez-Correa and Zaror, 1993). Therefore, it becomes very beneficial for those food processes where human experts exists and no mathematical models are avail-

able. Recent applications of fuzzy logic controllers on food process are reported in literature (Amiryousefi et al., 2017; Chung et al., 2010; Rahman et al., 2012; Li et al., 2010; Wu et al., 2018; Ioannou et al., 2006; Sui and Tong., 2016; Mahadevappa et al., 2017; Al-Mahasneh et al., 2012; Wu et al., 2017; Schmitz et al., 2014; Villanueva et al., 2015; Sui et al., 2018). The variety of fuzzy control applications indicates that this technique is becoming an important tool for complex processes. The main reason for the popularity of fuzzy controllers is the adequate control over nonlinear processes. The nonlinearity of the dough proofing process and volume expansion of dough pieces during the proofing process have been explained by Stanke et al. (2014). Due to its nonlinearity and complexity, dough proofing process offers attractive and challenging problems concerning control.

Deviation of controlled variable from the target values, high temperature fluctuation during the process and high settling time are the drawbacks of the existing PID control system for the dough proofing process. Therefore, in this contribution fuzzy logic control techniques were applied in order to achieve more efficient control over the proofing process in bread making and to overcome the drawbacks of the existing PID control system. This is the first time that a proofing process is controlled by a fuzzy logic system. Efficiency of the proofing process is achieved by using the output parameter provided by the fuzzy control system such as increasing the proofing chambers temperature due to the smaller size of the dough piece compared to the standard sample at the same time. The main contributions of this paper can be summarized as follows.

1) Design and implementation of a fuzzy logic control system based on image processing for controlling the proofing process in bread making.
2) Comparison of the performance of the designed control system with a PID controller in order to find the most efficient control system for the proofing process in bread making.

The remaining paper is organized as follows. Section 2 provides the materials and methods which were applied in this study. The process of controller design is presented in Section 3. Section 4 provides the performance of the controller and Section 5 concludes this paper.

2. Materials and methods

2.1. Monitoring system and image processing

The imaging system for online measurements of volume expansion of the dough pieces during proofing was adapted from Yousefi-Darani et al. (2018). Briefly, a digital camera (DFK 31BU03.H, The Imaging Source Europe GmbH, Germany) was installed at the side of a proofing chamber (Klimaschrank VC 4033, Vötsch Industrietechnik GmbH, Germany) and a LED light was used in the chamber for better illumination. The digital image processing algorithm which was used to evaluate fermentation process is carried out by software programmed in Matlab (R2016a, the Mathworks Inc., Natic, Massachuset USA). Fig. 1 shows the image processing steps which were used in this study. The image processing starts with cropping the image in order to eliminate the edges and frame an image containing one dough piece. The next step is converting the
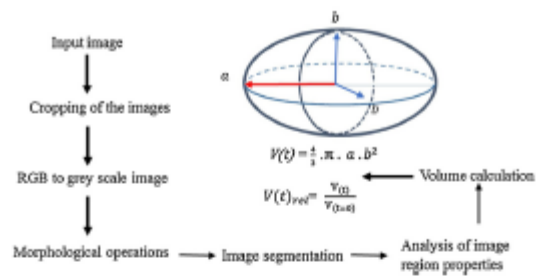

Fig. 1 – Image processing steps and relative volume measurement of the doughs.

image to grey scale image which is followed by some morphological operations such as dilation and erosion. In order to identify objects in the digital image a thresholding operation was applied and for calculating the volume of the dough piece, the bounding box surrounding the binary image was determined (Zettel et al., 2016). The volume of the dough piece was approximated by an ellipsoid shape with the same width and height. The relative volume of the dough piece was then calculated according to Eqs. (1) and (2).

$$V(t) = \frac{4}{3} . \pi . Lenght(t) . Height(t)^2 \qquad (1)$$

$$V(t)_{rel} = \frac{V(t)}{V(t=0)} \qquad (2)$$

where $V(t)$ is the volume of the dough, $V(t)_{rel}$ is relative volume at time t.

2.2. Reference volume measurements and data set for control

A reference data set (target values) for the process was prepared which presents the volume expansion of a dough piece obtained under standard conditions. In order to provide the reference data, a method for the production of bread was set up; a dough piece with standard recipe was prepared and proofed in the proofing chamber for one hour. During the whole proofing process, the temperature and moisture of the proofing chamber were kept constant at standard conditions (30 °C and 80% relative humidity). An image was captured every minute, the relative volume of the dough was calculated with the image processing system, and this value was considered as reference value for the corresponding point which the image was captured.

2.3. Pore size distribution in baked breads

Dough pieces which were fermented for one hour in the proofing chamber were baked for 25 min at 220 °C. Baked breads were cooled for 30 min at room temperature and cut into slices and then scanned with a flat scanner (HP SCANJET 8300). Images were processed using a dedicated software called "Gebäckanalyse" (Germany).

The pores areas are classified in five different classes: small ($0.10–2.00\,mm^2$), relative small ($2.01–3.00\,mm^2$), medium ($3.01–6.00\,mm^2$), relative big ($6.01–10.00\,mm^2$) and big ($10.01–11.00\,mm^2$). All measurements were performed in triplicate.
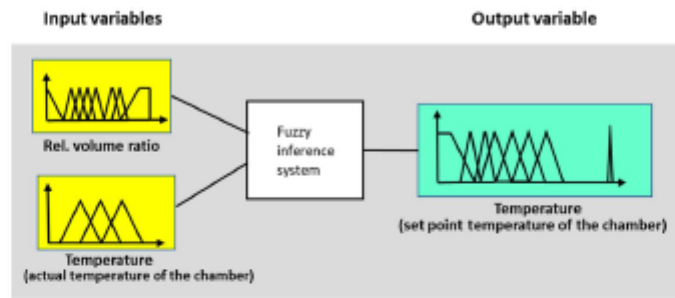
Fig. 2 – Input and output variables of the fuzzy logic dough proofing system.

2.4. Sample preparation

The dough preparation method was similar to the pervious method for the PID controller. Three different compositions of dough (standard recipe, more yeast and less yeast) were studied. The standard recipe dough formulation was 1000 g wheat flour type 550 (Rettenmeier Mühle GmbH, Germany), 600 g water, 20 g salt (Siede-Speisesalz, Südsalz GmbH, Germany), 35 g baking malt (Ulmer Goldmalz, CSM Deutschland GmbH, Germany), 10 g margarine (Meister Goldback, CSM Deutschland GmbH, Germany) and 10 g dry yeast (Trockenhefe, Safe-instand Red, France). The mixing was done in a spiral mixer (DIOSNA Dierks & Söhne GmbH, Germany) with 1 min at low speed (25 rpm) followed by 5 min at high speed (50 rpm). The dough rest was carried out for 20 minutes at 30 °C. The dough was then laminated and shaped in a shaping–moulding machine so that dough pieces for bread rolls are obtained. Dough pieces with more yeast and less yeast were prepared with the same recipe as the standard recipe, except that the amount of yeast was 7.5 g for less and 12.5 g for more yeast. In order to prepare frozen dough pieces, the dough pieces were frozen in a freezer (T = −19 °C) for more than six hours. Thawed dough pieces were prepared by thawing the frozen dough pieces in a refrigerator for more than six hours (T = +7 °C).

2.5. Temperature profile of the frozen dough pieces

In order to evaluate the temperature at the dough surface (DST) and dough centre (DCT) of the frozen dough pieces during fermentation, a thermometer (Almemo 2590–3S-4S) was inserted in the centre and surface of the frozen dough pieces before freezing.

3. Controller design

3.1. The fuzzy system

A typical fuzzy system is running in three steps. During the first step crisp data are converted into fuzzy data or membership functions (fuzzification). In the next step membership functions are combined with the control rules to derive the fuzzy output (fuzzy inference). The last step is called defuzzification where each associated output is calculated. Keeping the same framework, a fuzzy system with two inputs and one output was developed in the present study. Fig. 2 presents an overview of the fuzzy system.

3.1.1. Parameter selection for the fuzzy system

Several parameters are influencing the fermentation time of bread dough: temperature, quality and quantity of the yeast,

type of flour, amount of salt, water content, improvers, or process conditions for dough preparation such as mixing time and speed. The impacts of these parameters on the proving process have been widely studied by various authors (Shehzad et al., 2010; Chevallier et al., 2012; Chiotellis and Campbell, 2003; Sahlström et al., 2004). However, the main controlling factors which can be changed during the proofing process are temperature and humidity. Here the proofing chambers humidity is set constant. Therefore, only temperature was chosen as control variable.

In the present study the controlled variable is the volume of the dough piece during the proofing process, therefore the value obtained from dividing the measured relative volume of the dough sample by the relative volume from the reference data at the same time points is used as one input variable (V_{rr}), the second input is the actual temperature of the proofing chamber (T_m), and the output is the temperature set point to be adjusted at the proofing chamber (T_{sp}).

3.1.2. Fuzzy input and output sets

The fuzzy system was programmed using the software MATLAB (Fuzzy Logic Toolbox).The classical notion of membership functions (Zadeh, 1965) was used to evaluate the input and output parameters as fuzzy variables. The membership functions are in the form of triangular shape and trapezoidal functions. The triangular shape is a function of a vector, x, and depends on three scalar parameters a, b, and c, as given by Eq. (3) and the trapezoidal function is a function of vector y, which depends on four scalar parameters a, b, c, and d, as given by Eq. (4).

$$f(x \; ; a, b, c) = \begin{cases} 0, & x \le a \\ \dfrac{c-x}{c-b} & b \le x \le c \\ \dfrac{c-x}{c-b} & b \le x \le c \\ 0, & c \le x \end{cases} \tag{3}$$

$$f(y; a, b, c, d) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a} & b \le x \le c \le c \\ 1, & a \le x \le b \\ \dfrac{d-x}{d-c} & c \le x \le d \\ 0, & d \le x \end{cases} \tag{4}$$

In Eq. (3) the parameters a and c locate the feet of the triangle and the parameter b locates the peak and in Eq. (4) the
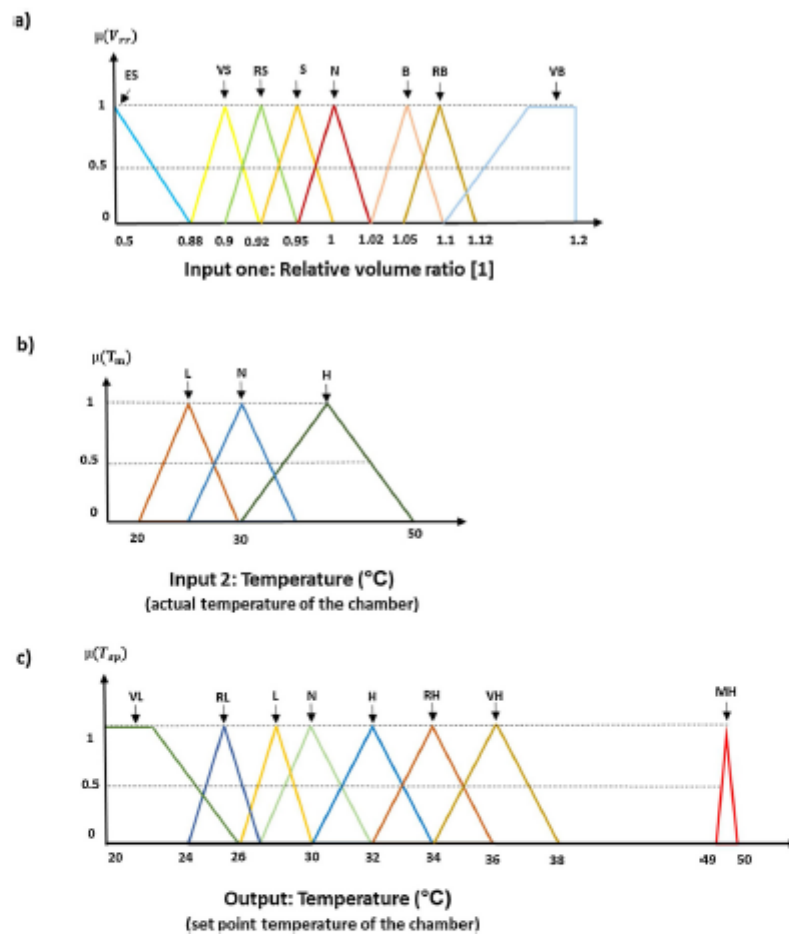
Fig. 3 – Fuzzy sets and membership functions for the input and output variables of the dough proofing controller: (a) fuzzy sets for relative volume ratio, (b) fuzzy sets for temperature (input), (c) fuzzy sets for set point temperature (output).

parameters a and d locate the feet of the trapezoid and the parameters b and c locate the shoulders.

The output of the system (set point temperature of the proofing chamber temperature control unit) is responsible for controlling the convergence speed of the system, but due to the sensitivity of dough pieces to high temperatures, special care should be taken in to account when it comes to choosing the membership functions and their appropriate ranges. Therefore, in order to define the membership functions, several experiments were conducted. For the first input, based on the relative volume ratio, eight membership functions were defined: extremely small (ES), very small (VS), relative small (RS), small (S), normal (N), big (B), relative big (RB) and very big (VB). For the second input variable (actual proofing chambers temperature), three linguistic terms were selected and expressed by the fuzzy sets, low (L), normal (N) and high (H).

For the output variable (set point temperature of the proofing chamber (T_{sp})) eight membership functions were defined and expressed as: very low (VL), relative low (RL), low (L), normal (N), high (H), relative high (RH) very high (VH) and maximum high (MH).

The membership functions of the fuzzy sets for each of the two input variables and the output variable are shown in Fig. 3(a)–(c).

3.1.3. Fuzzy rule base

After setting the input and output parameters, the next step is matching them with IF-THEN rules and aggregation of the levels of matching, so that the inference can be made. Mamdani's fuzzy inference method was used here implemented in MATLAB. All rules used for controlling the proofing process are presented in a matrix form in Fig. 4. This matrix configuration allows seeing a large number of rules and their outcomes at a glance.

All rules are evaluated in parallel, and the order of the rules is unimportant. Below are some examples how the rules are evaluated.

IF dough volume is much larger compared to the standard sample and the actual temperature of the proofing chamber is high, then it implies that the dough needs to be cooled down in order to decrease the volume expansion. This can be represented by the following rule.

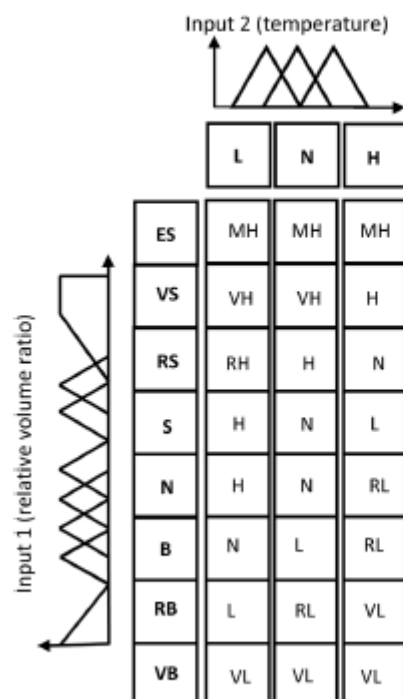IF (V_{rr} — VB) and (T_m — H) THEN (T_{sp} — VL)

Fig. 4 – Fuzzy logic rule matrix of the control system for dough proofing.

When dough volume is smaller than the standard sample and the temperature of the proofing chamber is low, it indicates that higher temperatures should be applied in order to accelerate volume expansion of the dough. This can be represented by the following rule.

IF $(V_{rr} - VS)$ and $(T_m - L)$ THEN $(T_{sp} - VH)$

Once the reference data and the sample have the same size there is no further necessity to either decrease or increase the temperature. Therefore, the temperature is set to 30 °C. This can be represented by the following rule.

IF $(V_{rr} - N)$ and $(T_m - N)$ THEN $(T_{sp} - N)$

Fig. 5 illustrates an example of fuzzification and application of one rule in the proposed fuzzy control method for the dough proofing control.

In Fig. 5 an example of the application of rule number 4 is presented where the actual values are as follows: the relative volume ratio is $V_{rr} = 1.07$ and the measured temperature of the proofing chamber is $T_m = 33\,°C$. Each input triggers two membership functions, which are for the relative volume ratio B (Big) with the value of the membership function mf-V_{rr}(B[1.07]) = 0.2 and RB (Relative Big) with mf-V_{rr}(RB[1.07]) = 0.8; and for the temperature the membership functions N (Normal) mf-T_m(N[33 °C]) = 0.3 and H (High) with mf-T_m(H[33 °C]) = 0.7. Thus, the conditional parts of four rules are true and the overall conclusion must be drawn from rules 1–4 in Fig. 6.

The conditional part is composed out of two conditions which are connected logically by an AND function. Therefore, the outcome will be calculated by the minimal value of the corresponding membership function values. For rule 4 it will

be the minimum of 0.2 and 0.3. As a consequence the area of the membership function L of T_{sp} below 0.2 will contribute to the final result. This will be illustrated in Fig. 6, where the contributions coming from the other 3 rules can be seen.

3.1.4. Defuzzification
Defuzzification converts the fuzzy values into a crisp value. In this contribution the method of centre of gravity (COG) was used according to Eq. (5).

$$T_{sp} = \frac{\int_{T_{sp1}}^{T_{sp7}} T_{sp} \cdot \mu\left(T_{sp}\right) dT_{sp}}{\int_{T_{sp1}}^{T_{sp7}} \mu\left(T_{sp}\right) dT_{sp}} \tag{5}$$

where T_{sp} is the crisp output value of the fuzzy controller, T_{spi} is the i th bound of integration and $\mu\left(T_{sp}\right)$ represents its membership value, whose top is cut as explained in Fig. 7.

3.2. Closed loop control system

The designed fuzzy logic system as well as the image processing system are implemented in a feedback control system. The complete outlook of the setup for controlling the dough proofing process in bread making is depicted in Fig. 8.

The monitoring system is implemented in the loop to continuously measure the volume of the dough pieces. The starting temperature of the proofing chamber is always 30 °C. The control system starts with capturing an image from the dough sample. The image is sent to the image evaluation algorithm, relative volume is calculated. The ratio of the relative volume of the actual sample to the relative volume from the reference data set among with the actual temperature of the proofing chamber is sent to the fuzzy logic controller. The output signal (temperature set point of the proofing chamber) is obtained based on the fuzzy decision making rules. This loop is repeated every 5 min for one hour.

4. Results and discussion

4.1. Controller performance

Dough pieces in different temperature starting states with variation in amount of yeast were prepared and fermented in the proofing chamber using the controller. Volume evolutions of a dough pieces during the proofing processes are shown in Fig. 9.

The optimum results for dough fermentation is considered to be the same as the final volume of the standard dough sample. The analysis of volume variation as function of time reveals that similar to the PID controller, regardless to the actual starting state of the dough and any alteration in amount of yeast, the fuzzy controller is able to yield comparable final volumes compared to the standard samples after one hour.

Volume evaluation curves presented in Fig. 9 depicts a significant decrease in volume deviation of samples which were proofed by the fuzzy logic controller from target values compared to the ones proofed by the PID controller.

The performance index of a controller can be defined as the root mean squared error (RMSE) of the controlled process output and the standard output. Furthermore, deviations of final volume of the controlled dough pieces from standard dough pieces (DFS) as well as the settling time (the time required for the response curve to reach and stay within a range of certain percentage of the final value) were used as controller
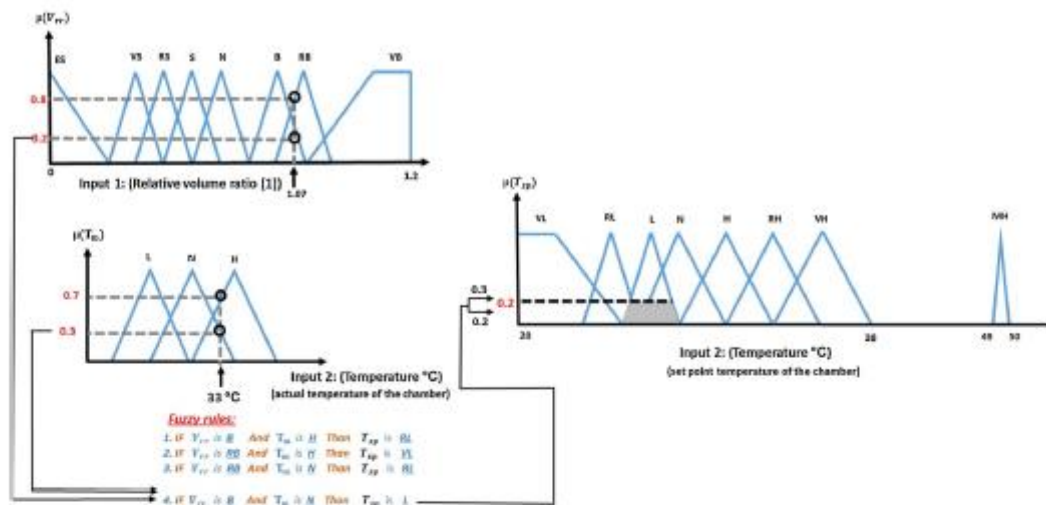
Fig. 5 – Example of fuzzification and application of rule number 4 in the developed dough proofing fuzzy controller (see text for a detailed description).
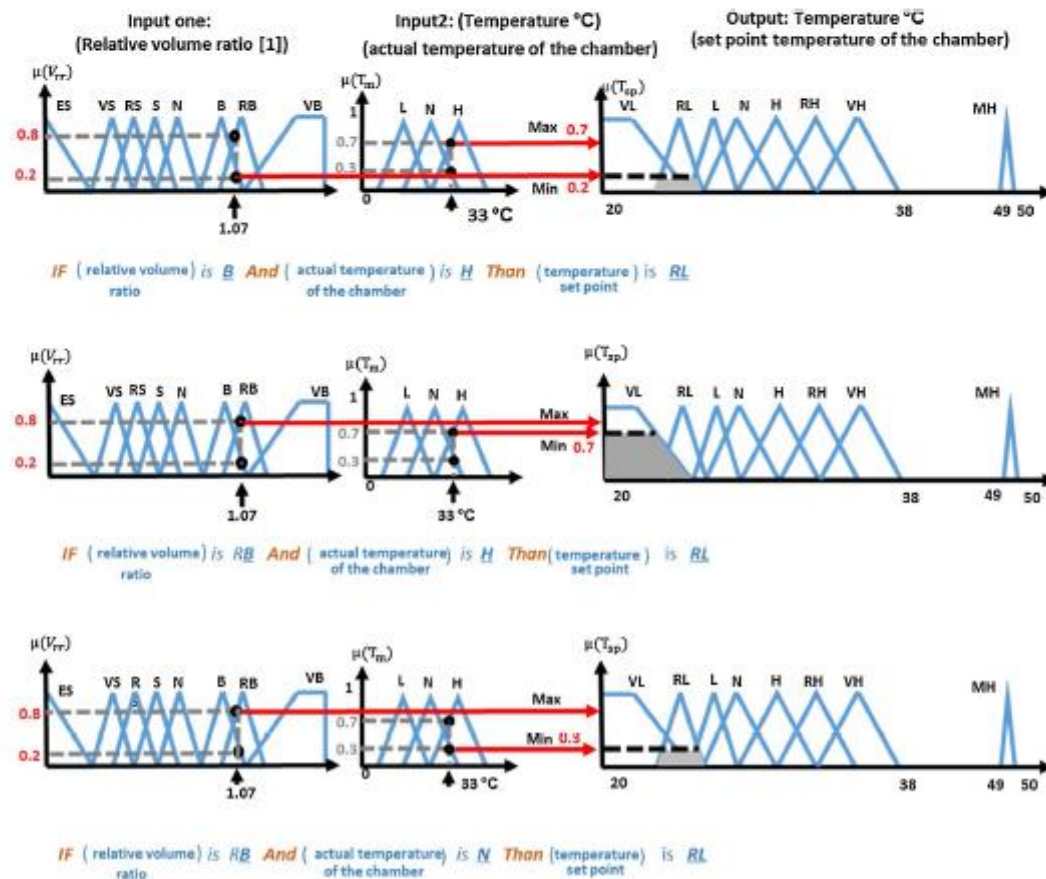
Fig. 6 – Contributions from the other three rules for the determination of T_{sp}, (b).

93

$$\text{Output temperature} = \frac{\int_{T_{sp_1}}^{T_{sp_7}} T_{sp} \cdot \mu(T_{sp}) dT_{sp}}{\int_{T_{sp_1}}^{T_{sp_7}} \mu(T_{sp}) dT_{sp}} = 23.6\,°C$$
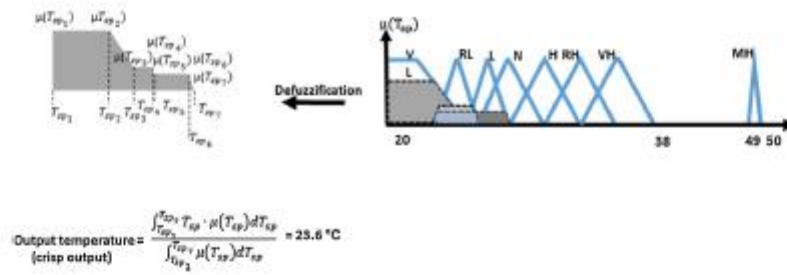
Fig. 7 – Illustration of the defuzzification of the results to obtain the crisp value for T_{sp}.
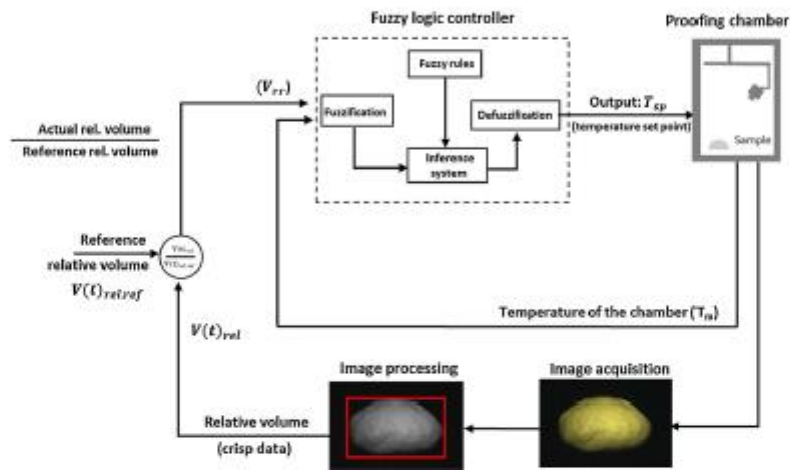
Fig. 8 – Schematic diagram of the fuzzy logic closed loop control system for dough proofing.

Fig. 9 – Relative volume evolutions for dough pieces in different temperature starting states and with different amounts of yeast obtained with the standard, fuzzy control and PID control. Normal state-more yeast (A), normal state-less yeast (B), thawed dough-more yeast (C), thawed dough-less yeast (D), frozen state-more yeast (E), frozen state-less yeast (F).

94

Table 1 – Performance criteria of fuzzy and PID controller for the proofing process.

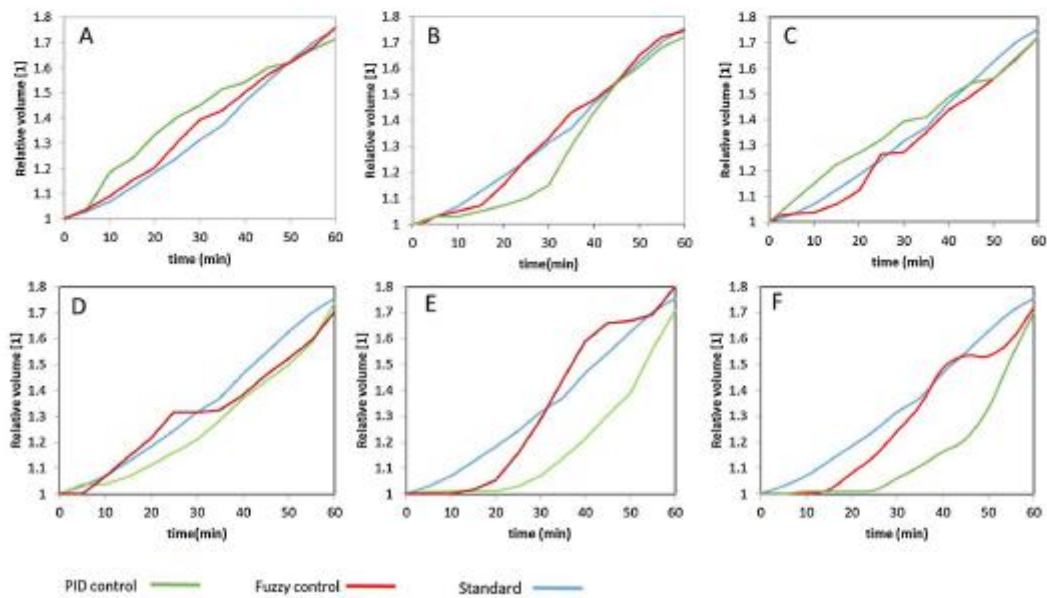| State of the dough	Controller	RMSE	Deviation of final volume from standard	Settling time (min)
Normal state-more yeast	Fuzzy control	0.03	0 %	37
	PID control	0.09	2.28 % (lower)	47
Normal state-less yeast	Fuzzy control	0.02	0.57 % (lower)	20
	PID control	0.05	1.71 % (lower)	38
Thawed dough-more yeast	Fuzzy control	0.04	1.70 % (lower)	21
	PID control	0.06	1.70 % (lower)	37
Thawed dough-less yeast	Fuzzy control	0.06	2.85 % (lower)	43
	PID control	0.08	1.14 % (lower)	43
Frozen state-more yeast	Fuzzy control	0.07	2.28 % (higher)	51
	PID control	0.17	2.31 % (lower)	60
Frozen state-less yeast	Fuzzy control	0.07	2.28 % (lower)	60
	PID control	0.20	2.85 % (lower)	60

performance index for the volume evaluation curves (Table 1) and IAE (integral absolute error) as performance index for the time–temperature curves. RMSE, DFS and IAE are defined as Eqs. (6)–(8) respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(V_s - V_c)^2}{n}} \qquad (6)$$

$$DFS\% = \frac{V_{sf} - V_{cf}}{100} \qquad (7)$$

$$IAE = \int (V_s - V_c)\, dt \qquad (8)$$

where V_s is the volume of the controlled dough piece, V_c is the volume of the standard sample, V_{sf} is the final volume of the controlled dou gh piece and V_{cf} is the final volume of a standard dough piece.

Data in Table 1 reveal that the response of the fuzzy logic controller gives more satisfactory performance with respect to RMSE and final volume of the dough pieces at the end of the proofing process. Furthermore, the settling time for the fuzzy controller is significantly shorter than the settling time for the PID controller.

The settling time of both controllers (60 min) for proofing dough pieces in frozen state are relatively high which is due to the longer time required for thawing frozen dough pieces. Freezing presents a significant effect as disturbance to the proofing process, causing the volume of the samples to deviate from the set value during the proofing process. Consequently, high temperatures should be applied to rapidly minimize the deviation. The highest temperature applied for thawing frozen dough pieces is 50 °C by both controllers. Increasing the thawing temperature beyond 50 °C would lead to shorter settling time. But on the contrary, high temperatures result in denaturation of proteins, deactivation of the yeast cells and higher enzyme activity (protease and amylase). These higher enzyme rates would result in weaker doughs and substantially less oven spring (Siffring and Bruinsma, 1993). Therefore, with considering the final volume of the frozen dough pieces at the end of the process which is comparable with the standard samples, the long settling time could be neglect able.

Data in Fig. 10 presents the time–temperature profile of the proofing chamber during the proofing process for both controllers as well as the total IAE (the sum of areas above and below the set point) of the time–temperature graphs.

A lower IAE represent less temperature fluctuation during the proofing process. Temperature fluctuation may result in defects such as small and compact final products with a dense crumb structure, poor retention of CO_2 and inhibiting expansion during proofing.

However, the final volume of the dough pieces proofed with any of the controllers are in the same range. The data in Fig. 9 reveals that the total IAE for dough samples in normal state and thawed state which were proofed with the fuzzy controller is significantly lower compared to the ones proofed with the PID controller. However for the other three cases it is higher.

4.2. Temperature profile of the frozen dough

Time-temperature profile recorded at the centre of the frozen dough pieces (DCT) and surface of the frozen dough pieces (DST) during the proofing process as well as the variations of the air temperature in the proofing chamber (PCT) are shown in Fig. 10. Data in Fig. 10 illustrates when the fuzzy controller is applied the temperature fluctuation between DST and PCT is lower compared to when the PID controller is applied. However, temperature fluctuation between DST and DCT are lower when the PID controller is applied.

According to literature (Cauvain, 2012; Karnel and Stauffer, 1993), optimal dough temperature at the end of the proofing process is 28–35 °C. Temperatures below this range results in small final products with dense and crumb structure and temperatures higher than that can produce unpleasant flavours through production of organic acids. The time-temperature data in Fig. 11 reveals that when the fuzzy logic controller is applied, the temperature at the surface and centre of the frozen dough pieces reach to 30 °C after approximately 40 min and stays constant until the end of the proofing process. On the contrary, when the PID controller is applied, the temperature at the surface and centre of the dough pieces are never in equilibrium during the process and the surface temperature of the dough at the end of the process is slightly higher than 40 °C. Therefore, the fuzzy logic controller shows better results over controlling the temperature of the frozen dough pieces during the proofing process.

4.3. Pore size distribution

Fig. 12 illustrates the pore size distribution for breads obtained from dough pieces proofed with the fuzzy controller and the PID controller. Pore size distribution of the bread was plotted between number of bubbles (%) and pore area. In Fig. 12, S stands for standard, N stands for normal, f stands for frozen, M stands for more yeast, L stands for less yeast, P stands for PID controller and F stands for Fuzzy controller. Sapirstein (1999) reported that pore size in the baked bread ranged between 0.08 and 8 mm equivalent diameters, which correspond to areas of
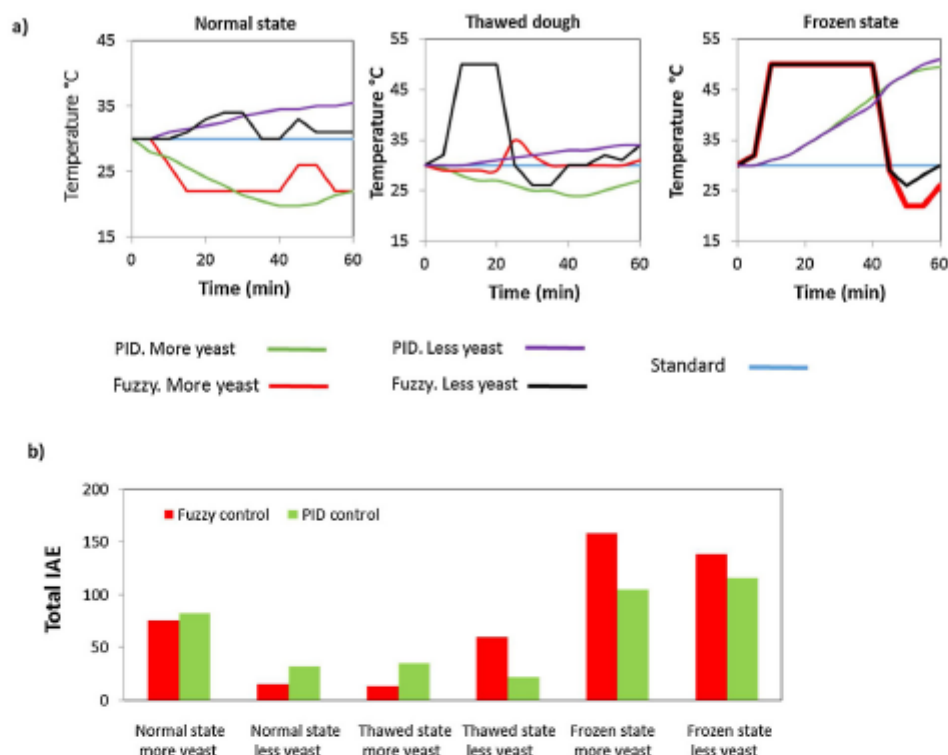
Fig. 10 – Time-temperature profile of the proofing chamber during the proofing process (a) and the total IAE of the time–temperature graphs (b).
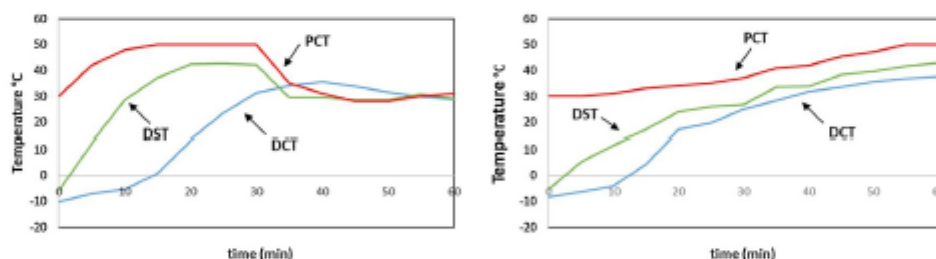

Fig. 11 – Time–temperature evolution in the proofing chamber (PCT) and at centre (DCT) and surface (DST) of dough piece during the proofing process, which were frozen at the proofing start.

0.02–$200\,mm^2$. The pores size distribution in baked bread in present study ranged between 0.1 and $11\,mm^2$ pore areas.

A general trend shown in the data is difference in pore size distribution between the standard sample and the samples which were proofed by the controllers.

In all samples, smaller pores are presented in greater numbers. Applying any of the two controllers increased the number of pores classified as small. Furthermore, for all samples beside the ones in frozen state with less yeast and thawed state with more yeast, when the fuzzy logic controller was applied, the number of small pores increased more compering to when the PID controller was applied.

Comparing the pore population classified as relative small, medium, and relative big, a non-significant difference is observed when different controller is applied. Moreover, no significant difference is observed between the mentioned pore population from the standard sample and the samples proofed by the controllers. On the other hand, a significant difference ($p < 0.05$) in percentage of pores classified as big is observed between the standard samples and the samples which were proofed by the controllers. These observations can be supported by the fact that in bread, pore sizes distribution is related to a balance between the rate of heating and the time needed for the outer region of the dough to develop some rigidity (Datta et al., 2007). Therefore, the difference in pore distribution is due to the rate of heating which each of the controllers apply during the proofing process.

5. Conclusion

A fuzzy logic control system was successfully designed and applied to the proofing process in bread making. Experiments

96

Fig. 12 – Pore size distribution in bread.
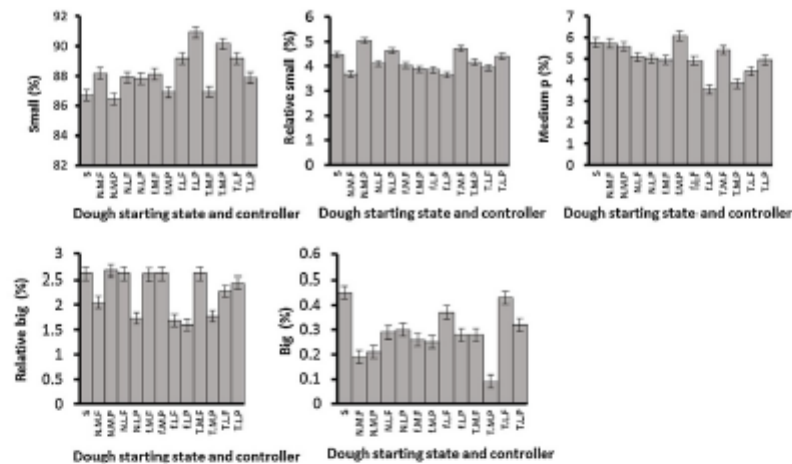
were performed in order to proof dough pieces in different starting states and with different amount of yeast added in the ingredients. The efficiency of the controller was evaluated and compared to maintain the volume of the dough pieces similar to the target values during the proofing process. The results were compared with the performance of a PID controller from a previous study.

The obtained results indicate that average deviations of the volume of the dough pieces with different amount of yeast and in different temperature starting states at the end of the proofing process was less than 2.9% compared to the target values. The obtained performance of the system is very satisfactory with respect to volume control and set point deviation compared to the PID controller. Additionally, the fuzzy logic controller performance with respect to overall IAE, RSME, settling time, surface and central temperature of the frozen dough pieces and pore size distribution in baked breads exhibited satisfactory performance compared with the PID controller.

Furthermore, fuzzy logic algorithms can be constructed in a user-friendly way and, due to the fact that fuzzy logic controllers are closer to human thinking and perception, by elaborating a set of rules the system can be controlled without requiring the computation of any mathematical model. As a consequence the implementation of the fuzzy controller was faster. Therefore, the fuzzy controller represents a viable alternative for controlling the process with.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgment

References

Al-Mahasneh, M., Amer, M.B., Rababah, T., 2012. Modeling moisture sorption isotherms in roasted green wheat using least square regression and neural-fuzzy techniques. Food Bioprod. Process. 90 (2), 165–170.

Amiryousefi, M.R., Mohebbi, M., Golmohammadzadeh, S., Koocheki, A., Baghbani, F., 2017. Fuzzy logic application to model caffeine release from hydrogel colloidosomes. J. Food Eng. 167, 35–42.

Cauvain, S.P. (Ed.), 2012. Bread Making, Improving Quality. Woodhead Publ., Cambridge, UK.

Chevallier, S., Zúñiga, R., Le-Bail, A., 2012. Assessment of bread dough expansion during fermentation. Food Bioprocess Technol. 5 (2), 609–617.

Chung, C.-C., Chen, H.-H., Ting, C.-H., 2010. Grey prediction fuzzy control for pH processes in the food industry. J. Food Eng. 96 (4), 575–582.

Datta, A.K., Sahin, S., Sumnu, G., Keskin, S.O., 2007. Porous media characterization of breads baked using novel heating modes. J. Food Eng. 79 (1), 106–116.

Jean-Pierre, C., 2004. Process Control: Theory and Applications. Springer, London.

Kamel, B.S., Stauffer, C.E., 1993. Advances in Baking Technology. Chapman & Hall, London, UK.

Kondakci, T., Zhou, W., 2017. Recent applications of advanced control techniques in food industry. Food Bioprocess Technol. 10 (3), 522–542.

Lahtinen, S., 2001. Identification of fuzzy controller for use with a falling-film evaporator. Food Control 12 (3), 175–180.

Li, Z., Raghavan, G.V., Wang, N., 2010. Apple volatiles monitoring and control in microwave drying. LWT Food Sci. Technol. 43 (4), 684–689.

Mahadevappa, J., Groß, F., Delgado, A., 2017. Fuzzy logic based process control strategy for effective sheeting of wheat dough in small and medium-sized enterprises. J. Food Eng. 199, 93–99.

Perez-Correa, J., Zaror, C., 1993. Recent advances in process control and their potential applications to food processing. Food Control 4 (4), 202–209.

Rahman, M.S., Rashid, M., Hussain, M., 2012. Thermal conductivity prediction of foods by neural network and fuzzy (ANFIS) modeling techniques. Food Bioprod. Process. 90 (2), 333–340.

Romagnoli, J.A., Palazoglu, A., 2005. Introduction to Process Control. CRC Press.

Sahlström, S., Park, W., Shelton, D.R., 2004. Factors influencing yeast fermentation and the effect of LMW sugars and yeast fermentation on hearth bread quality. Cereal Chem. 81 (3), 328–335.

Sapirstein, H., 1999. The imaging and measurement of bubbles in bread. Bubbles in Food. GM.

Schmitz, J., Silva, F., Neves Filho, L., Fileti, A., Silveira, V., 2014. Multivariable fuzzy control strategy for an experimental chiller system. J. Food Process Eng. 37 (2), 160–168.

Shehzad, A., Chiron, H., Della Valle, G., Kansou, K., Ndiaye, A., Réguerre, A., 2010. Porosity and stability of bread dough during proofing determined by video image analysis for different compositions and mixing conditions. Food Res. Int. 43 (8), 1999–2005.

Siffring, K., Bruinsma, B.L., 1993. Effects of proof temperature on the quality of pan bread. Cereal Chem. 70 (3), 351–353.

Stanke, M., Zettel, V., Schütze, S., Hitzmann, B., 2014. Measurement and mathematical modeling of the relative volume of wheat dough during proofing. J. Food Eng. 131, 58–64.

Sui, S., Tong, S., 2016. Fuzzy adaptive quantized output feedback tracking control for switched nonlinear systems with input quantization. Fuzzy Sets Syst. 290, 56–78.

Sui, S., Chen, C.L.P., Tong, S.C., 2018. Fuzzy adaptive finite-time control design for non-triangular stochastic nonlinear systems. IEEE Trans. Fuzzy Syst., http://dx.doi.org/10.1109/TFUZZ.2018.2882167.

Villanueva, D., Posada, R., González, I., García, Á., Martínez, A., 2015. Monitoring of a sugar crystallization process with fuzzy logic and digital image processing. J. Food Process Eng. 38 (1), 19–30.

Wu, C., Liu, J., Jing, X., Li, H., Wu, L., 2017. Adaptive fuzzy control for nonlinear networked control systems. IEEE Trans. Syst. Man Cybern. Syst. 47 (8), 2420–2430.

Wu, C., et al., 2018. Observer-based adaptive fault-tolerant tracking control of nonlinear nonstrict-feedback systems. IEEE Trans. Neural Netw. Learn. Syst. 29 (7), 3022–3033.

Yousefi-Darani, A., Paquet-Durand, O., Zettel, V., Hitzmann, B., 2018. Closed loop control system for dough fermentation based on image processing. J. Food Process Eng., e12801.

Zadeh, L.A., 1965. Information and control. Fuzzy Sets 8 (3), 338–353.

Zettel, Viktoria, et al., 2016. Image analysis and mathematical modelling for the supervision of the dough fermentation process. In: AIP Conference Proceedings. AIP Publishing, Vol. 1769. No. 1.

5. Discussion

A bottleneck in bioprocess monitoring and control is often caused by the lack of reliable sensors. That is why estimation techniques issued from control theory have been applied to on-line estimation of bioprocess variables. That induced the development of software sensors, which associate a sensor (hardware) and an estimation algorithm (software) in order to provide on-line estimates of the unmeasurable variables and kinetic parameters (Nicoletti et al., 2009).

Within this thesis software sensors were developed for monitoring and control of the fermentation of baker's yeast, three times the suspension (baker's yeast cultivation process) and once the solid-state fermentation (dough fermentation) was investigated. Both processes are of high importance for industries.

The designed software sensors for the cultivation of the baker's yeast process are based on metal oxide gas sensor array. Application of gas sensor arrays for on-line monitoring of yeast cultivation has been previously reported in literature (Lidén et al., 1998). However, lack of stability over time (sensor drift) and the high cost of recalibration are factors which had limited the widespread adoption of gas sensor arrays in real industrial setups (Di Carlo and Falasconi, 2012).

According to literature, there are several different methods for drift compensation; these methods can be classified into three main categories. The first is the search for new materials that can reversibly interact with the relevant gas, so that the detected molecules unbind from the sensor material as soon as the gas has been purged from the sensor surface (Yamazoe, 191; Roth et al., 1994). The second is the dynamical characterization of the sensor response and additionally, the third is the use of appropriate signal processing techniques, including feature extraction and pattern recognition methods (Vergara et al., 2007; Liu et al., 2013). Accordingly, the last two methods were applied in this work for sensor drift compensation. In addition, in order to overcome the time consuming and expensive calibration issue of the gas sensor array, a model-based calibration procedure was performed. For this procedure, the only requirement is the process mathematical model (simulation model) and the response of the gas sensor array from a single cultivation run. Then the parameters of the simulation model can be calculated by minimising the prediction error by optimising the kinetic parameter values of the simulation model as well as the parameter values of the chemometric model.

The designed gas sensor array and the model-based calibration approach was used for on-line prediction of ethanol concentration in three different cultivations. The experiments carried out show that the proposed method provides comparable results to the reference ethanol concentration values obtained by HPLC. Thus, the gas sensor array has great potential in on-line monitoring of ethanol concentration during yeast cultivation. However, due to frequent flushing of the sensor array with oxygen flow, predicted ethanol concentrations are available only every five minutes. Therefore, in order to have continues ethanol concentrations as well as the concentration of other important process variables such as biomass, glucose and the growth rates, it was crucial to implement a software sensor. For this reason, the nonlinear extensions of the Kalman filter namely the Extended Kalman filter (EKF) and the Unscented Kalman filter (UKF) were separately applied and used as state estimation methods. The designed software sensors give the possibility to predict glucose, ethanol and biomass concentrations simultaneously from the only available infrequent on-line measurements of ethanol concentration.

By applying the EKF as the state estimator, the accuracy of the estimated biomass and substrate production were in line with other studies which have also implemented an EKF algorithm for monitoring the baker's yeast cultivation (Hitzmann et al., 2000; Popova et al., 2013). However, in this application the maximal specific growth rates on glucose and ethanol were also estimated. As a consequence, the rapid and precise estimation of these variables could increase the overall knowledge of the process.

In spite of the satisfactory results obtained by the EKF, it has some disadvantages such as: it requires the calculation of Jacobians at each time step, which may be difficult to obtain for higher order systems; it does linear approximations of the system at a given time instant, which may introduce errors in the estimation, leading to a state divergence over time. Therefore, another software sensor based on the the Unscented Kalman filter was designed and implemented for continuous state and parameter estimation of the baker's yeast fermentation process. The results were compared with the ones optioned from the EKF (from cultivations with similar initial conditions).

The obtained results indicate that the application of both techniques delivers satisfactory estimates of the important state variables and the product concentrations. However, the errors using UKF is comparatively less than the errors using EKF, which means the UKF is more adaptive than EKF in state estimation. The decisive advantage of the Unscented Kalman filter

is that the dynamic process model and the model that relates the state variables to the quantities that are measured on-line can be used in their original forms. Only the description of the uncertainty of the states is approximated (Simutis and Lübbert, 2017).

It can be concluded that the developed software sensors expand the application range of the gas sensor array for the nonlinear state and parameter estimation of the baker's yeast fermentation process. Consequently, it is promising to build up a compact and economical version of such measurement systems. Such sensors would be cost-effective and miniaturized devices for routine analysis, which could be advantageous to real-time bioprocess monitoring and control.

Since in solid-state fermentation (dough fermentation), the concentration of ethanol production is low, it was not possible to apply the gas sensor array for on-line monitoring of the process. Therefore, another software sensor was designed based on image analysis for predicting the volume of the dough pieces. The measurement system was complimented with a controller and was used for controlling the fermentation process. The controller consists of a fuzzy logic control system which changes the proofing chamber's temperature according to the volume of the dough pieces. The controller maintains the volume of the dough pieces similar to the volume expansion of a dough piece in standard conditions during the proofing process. The controller was experimentally evaluated by preparing dough pieces with different amount of yeast added in the ingredients and in different starting states and proofing them with the supervision of the fuzzy control system. The performance of the controller was compared to the performance of a PID controller from another study (Yousefi-Darani et al., 2018).

The obtained results indicates that average deviation of the volume of the dough pieces (at the end of the proofing process) with different amount of yeast added in the ingredients and in different starting states was less than 2.9 % compared to the target values. The obtained performance of the system is very satisfactory with respect to volume control and set point deviation compared to the PID controller. Additionally, the fuzzy logic controller performance with respect to RMSE, settling time, surface and central temperature of the frozen dough pieces and pore size distribution in baked breads exhibited satisfactory performance compared with the PID controller. Furthermore, fuzzy logic algorithms can be constructed in a user-friendly way and, due to the fact that fuzzy logic controllers are closer to human thinking and perception, by elaborating a set of rules the system can be controlled without

requiring the computation of any mathematical model. As a consequence, the implementation of the fuzzy controller is faster. Therefore, the fuzzy controller represents a viable alternative for controlling the process.

Bibliography

Adesanya, V. O., M. P. Davey, S. A. Scott and A. G. Smith (2014). "Kinetic modelling of growth and storage molecule production in microalgae under mixotrophic and autotrophic conditions." Bioresource technology **157**: 293-304.

Ali, J. M., N. H. Hoang, M. A. Hussain and D. Dochain (2015). "Review and classification of recent observers applied in chemical process systems." Computers & Chemical Engineering **76**: 27-41.

Alonso, A. A., I. G. Kevrekidis, J. R. Banga and C. E. Frouzakis (2004). "Optimal sensor location and reduced order observer design for distributed process systems." Computers & chemical engineering **28**(1-2): 27-35.

Autio, K. and T. Laurikainen (1997). "Relationships between flour/dough microstructure and dough handling and baking properties." Trends in Food Science & Technology **6**(8): 181-185.

Babin, P., G. Della Valle, H. Chiron, P. Cloetens, J. Hoszowska, P. Pernot, A.-L. Réguerre, L. Salvo and R. Dendievel (2006). "Fast X-ray tomography analysis of bubble growth and foam setting during breadmaking." Journal of Cereal Science **43**(3): 393-397.

Barker, R. and M. Williamson (1992). Assessment of VOC emissions and their control from baker's yeast manufacturing facilities. Final report, Midwest Research Inst., Cary, NC (United States).

Bastin, G. and D. Dochain (1991). On-line estimation and adaptive control of bioreactors: Elsevier, Amsterdam, 1990 (ISBN 0-444-88430-0). xiv+ 379 pp. Price US $146.25/Dfl. 285.00, Elsevier.

Baughman, D. R. and Y. A. Liu (2014). Neural networks in bioprocessing and chemical engineering, Academic press.

Bejarano, F., A. Poznyak and L. Fridman (2007). "Hierarchical second-order sliding-mode observer for linear time invariant systems with unknown inputs." International Journal of Systems Science **38**(10): 793-802.

Biechele, P., C. Busse, D. Solle, T. Scheper and K. Reardon (2015). "Sensor systems for bioprocess monitoring." Engineering in Life sciences **15**(5): 469-488.

Bitzer, M. and M. Zeitz (2002). "Design of a nonlinear distributed parameter observer for a pressure swing adsorption plant." Journal of process control **12**(4): 533-543.

Bogaerts, P. (1999). "A hybrid asymptotic-Kalman observer for bioprocesses." Bioprocess Engineering **20**(3): 249-255.

Bogaerts, P. and R. Hanus (2001). "On-line state estimation of bioprocesses with full horizon observers." Mathematics and computers in simulation **56**(4-5): 425-441.

Bogaerts, P. and A. V. Wouwer (2003). "Software sensors for bioprocesses." ISA transactions **42**(4): 547-558.

Bonny, J.-M., J. Rouille, G. Della Valle, M.-F. Devaux, J.-P. Douliez and J.-P. Renou (2004). "Dynamic magnetic resonance microscopy of flour dough fermentation." Magnetic resonance imaging **22**(3): 395-401.

Carmel, L., S. Levy, D. Lancet and D. Harel (2003). "A feature extraction method for chemical sensors in electronic noses." Sensors and Actuators B: Chemical **93**(1-3): 67-76.

Chen, Z., Z. Chen, Z. Song, W. Ye and Z. Fan (2019). "Smart gas sensor arrays powered by artificial intelligence." Journal of Semiconductors **40**(11): 111601.

Dewasme, L., G. Goffaux, A.-L. Hantson and A. V. Wouwer (2013). Experimental validation of an Extended Kalman Filter estimating acetate concentration in E. coli cultures." Journal of Process Control 23.2 (2013): 148-157.

Del Rio-Chanona, E. A., J. L. Wagner, H. Ali, F. Fiorelli, D. Zhang and K. Hellgardt (2019). "Deep learning-based surrogate modeling and optimization for microalgal biofuel production and photobioreactor design." AIChE Journal **65**(3): 915-923.

Di Carlo, S., and M. Falasconi. Drift correction methods for gas chemical sensors in artificial olfaction systems: techniques and challenges. INTECH Open Access Publisher, 2012.

Dochain, D. (2003). "State and parameter estimation in chemical and biochemical processes: a tutorial." Journal of process control **13**(8): 801-818.

Fissore, D., R. Pisano and A. A. Barresi (2007). "Observer design for the Selective Catalytic Reduction of NOx in a loop reactor." Chemical Engineering Journal **128**(2-3): 181-189.

Fouchard, S., J. Pruvost, B. Degrenne, M. Titica and J. Legrand (2009). "Kinetic modeling of light limitation and sulfur deprivation effects in the induction of hydrogen production with Chlamydomonas reinhardtii: Part I. Model development and parameter identification." Biotechnology and bioengineering **102**(1): 232-245.

Gan, Z., P. Ellis and J. Schofield (1995). "Gas cell stabilisation and gas retention in wheat bread dough." Journal of Cereal Science **21**(3): 215-230.

Gardner, J. W. and J. E. Taylor (2009). "Novel convolution-based signal processing techniques for an artificial olfactory mucosa." IEEE Sensors Journal **9**(8): 929-935.

Gauthier, J.-P., H. Hammouri and S. Othman (1992). "A simple observer for nonlinear systems applications to bioreactors." IEEE Transactions on automatic control **37**(6): 875-880.

Gélinas, P. (2012). "In search of perfect growth media for Baker's yeast production: Mapping patents." Comprehensive Reviews In Food science and food safety **11**(1): 13-33.

Gélinas, P. (2014). "Fermentation control in baker's yeast production: mapping patents." Comprehensive Reviews in Food Science and Food Safety **13**(6): 1141-1164.

Georgieva, P. and S. de Azevedo (2009). "Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control."

Hartyáni, P., I. Dalmadi and D. Knorr (2013). "Electronic nose investigation of Alicyclobacillus acidoterrestris inoculated apple and orange juice treated by high hydrostatic pressure." Food Control **32**(1): 262-269.

Hitzmann B, Broxtermann O, Cha Y-L, Sobieh O, Stärk E, Scheper T (2000) The control of glucose concentration during yeast fed-batch cultivation using a fast measurement complemented by an extended Kalman filter. Bioprocess Eng 23(4):337–341.

Hines, E., E. Llobet and J. Gardner (1999). "Electronic noses: a review of signal processing techniques." IEE Proceedings-Circuits, Devices and Systems **146**(6): 297-310.

Jianlin, W., F. Xuying, Z. Liqiang and Y. Tao (2010). Unscented transformation based robust kalman filter and its applications in fermentation process. Chinese Journal of Chemical Engineering 18(3): 412-418.

Julier, S., J. Uhlmann and H. F. Durrant-Whyte (2000). "A new method for the nonlinear transformation of means and covariances in filters and estimators." IEEE Transactions on automatic control **45**(3): 477-482.

Julier, S. J. and J. K. Uhlmann (1997). New extension of the Kalman filter to nonlinear systems. Signal processing, sensor fusion, and target recognition VI, International Society for Optics and Photonics.

Kadlec, P., B. Gabrys and S. Strandt (2009). "Data-driven soft sensors in the process industry." Computers & chemical engineering **33**(4): 795-814.

Kilborn, R. H., & Preston, K. R. (1981). to the Study of Dough Characteristics'. *Cereal Chem*, *58*(3), 198-201.

Kristensen RN (2002). Fed-Batch Process Modelling for State Estimation and Optimal Control. Copenhagen, Denmark: Nørhaven Digital.

Kroll, P., A. Hofer, S. Ulonska, J. Kager and C. Herwig (2017). "Model-based methods in the biopharmaceutical process lifecycle." Pharmaceutical Research **34**(12): 2596-2613.

Lidén, Helena, Carl-Fredrik Mandenius, Lo Gorton, Nina Q. Meinander, Ingemar Lundström, and Fredrik Winquist. "On-line monitoring of a cultivation using an electronic nose." Analytica chimica acta 361, no. 3 (1998): 223-231.

Lisci, S., Grosso, M., & Tronci, S. (2020). A Geometric Observer-Assisted Approach to Tailor State Estimation in a Bioreactor for Ethanol Production. Processes, 8(4), 480.

Liu, Hang, and Zhenan Tang. "Metal oxide gas sensor drift compensation using a dynamic classifier ensemble based on fitting." Sensors 13, no. 7 (2013): 9160-9173.

Lourenço, N. D., Lopes, J. A., Almeida, C. F., Sarraguça, M. C., & Pinheiro, H. M. (2012). Bioreactor monitoring with spectroscopy and chemometrics: a review. *Analytical and bioanalytical chemistry*, *404*(4), 1211-1237.

Llobet, E., J. Brezmes, R. Ionescu, X. Vilanova, S. Al-Khalifa, J. Gardner, N. Bârsan and X. Correig (2002). "Wavelet transform and fuzzy ARTMAP-based pattern recognition for fast gas identification using a micro-hotplate gas sensor." Sensors and Actuators B: Chemical **83**(1-3): 238-244.

Llobet, E., J. Brezmes, X. Vilanova, J. E. Sueiras and X. Correig (1997). "Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array." Sensors and Actuators B: Chemical **41**(1-3): 13-21.

Luenberger, D. G. (1964). "Observing the state of a linear system." IEEE transactions on military electronics **8**(2): 74-80.

Luttmann, R., D. G. Bracewell, G. Cornelissen, K. V. Gernaey, J. Glassey, V. C. Hass, C. Kaiser, C. Preusse, G. Striedner and C. F. Mandenius (2012). "Soft sensors in bioprocessing: a status report and recommendations." Biotechnology journal **7**(8): 1040-1048.

Musatov, V. Y., V. Sysoev, M. Sommer and I. Kiselev (2010). "Assessment of meat freshness with metal oxide sensor microarray electronic nose: A practical approach." Sensors and Actuators B: Chemical **144**(1): 99-103.

Nicoletti, Maria Carmo, Lakhmi C. Jain, and R. C. Giordano. "Computational intelligence techniques as tools for bioprocess modelling, optimization, supervision and control." In Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control, pp. 1-23. Springer, Berlin, Heidelberg, 2009.

Patnaik, P. (1997). "Artificial intelligence as a tool for automatic state estimation and control of bioreactors." Laboratory Robotics and automation **9**(6): 297-304.

Peris, M. and L. Escuder-Gilabert (2009). "A 21st century technique for food control: Electronic noses." Analytica chimica acta **638**(1): 1-15.

Persaud, K. and G. Dodd (1982). "Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose." Nature **299**(5881): 352-355.

Popova S, Ignatova M, Lyubenova V (2013) State and parameters estimation by extended Kalman filter for studying inhomogeneous dynamics in industrial bioreactors.

Reilly, M. T. and M. Charles (1990). "The use of on-line sensors in bioprocess control." Sensors in Bioprocess Control: 243-291.

Reyman, G. (1992). "Modelling and control of fed-batch fermentation of bakers' yeast." Food Control **3**(1): 33-44.

Romano, A., G. Toraldo, S. Cavella and P. Masi (2007). "Description of leavening of bread dough with mathematical modelling." Journal of food engineering **83**(2): 142-148.

Roth, M.; Hartinger, R.; Faul, R.; Endres, H.E. Drift reduction of organic coated gas-sensors by temperature modulation. Sens. Actuators B Chem. 1996, 36, 358–362.

Sahlström, S., W. Park and D. R. Shelton (2004). "Factors influencing yeast fermentation and the effect of LMW sugars and yeast fermentation on hearth bread quality." Cereal chemistry **81**(3): 328-335.

Scott, S. M., D. James and Z. Ali (2006). "Data analysis for electronic nose systems." Microchimica Acta **156**(3-4): 183-207.

Shankar, P. and J. B. B. Rayappan (2015). "Gas sensing mechanism of metal oxides: The role of ambient atmosphere, type of semiconductor and gases-A review." Sci. Lett. J **4**(4): 126.

Simutis, Rimvydas, and Andreas Lübbert. "Hybrid approach to state estimation for bioprocess control." Bioengineering 4, no. 1 (2017): 21.

Skaf, A., G. Nassar, F. Lefebvre and B. Nongaillard (2009). "A new acoustic technique to monitor bread dough during the fermentation phase." Journal of Food Engineering **93**(3): 365-378.

Solle, D., D. Geissler, E. Stärk, T. Scheper and B. Hitzmann (2003). Chemometric modelling based on 2D-fluorescence spectra without a calibration measurement. Bioinformatics 19(2): 173-177.

Stanke, M., V. Zettel, S. Schütze and B. Hitzmann (2014). "Measurement and mathematical modeling of the relative volume of wheat dough during proofing." Journal of Food Engineering **131**: 58-64.

Struyf, N., E. Van der Maelen, S. Hemdane, J. Verspreet, K. J. Verstrepen and C. M. Courtin (2017). "Bread dough and baker's yeast: An uplifting synergy." Comprehensive Reviews in Food Science and Food Safety **16**(5): 850-867.

Studer, R., V. R. Benjamins and D. Fensel (1998). "Knowledge engineering: principles and methods." Data & knowledge engineering **25**(1-2): 161-197.
Thibault, J., V. Van Breusegem and A. Chéruy (1990). "On-line prediction of fermentation variables using neural networks." Biotechnology and Bioengineering **36**(10): 1041-1048.

Trivedi, N. B., G. K. Jacobson, W. Tesch and J. P. Friend (1986). "Bakers' yeast." Critical Reviews in Biotechnology **4**(1): 75-109.

Tronci, S., F. Bezzo, M. Barolo and R. Baratti (2005). "Geometric observer for a distillation column: development and experimental testing." Industrial & engineering chemistry research **44**(26): 9884-9893.

Vatcheva, I., H. De Jong, O. Bernard and N. J. Mars (2006). "Experiment selection for the discrimination of semi-quantitative models of dynamical systems." Artificial Intelligence **170**(4-5): 472-506.

Venkateswarlu, C. (2005). "Advances in monitoring and state estimation of bioreactors."

Vergara, A.; Llobet, E.; Brezmes, J.; Ivanov, P.; Cane, C.; Gr ´ acia, I.; Vilanova, X.; Correig, X. ` Quantitative gas mixture analysis using temperature-modulated micro-hotplate gas sensors: Selection and validation of the optimal modulating frequencies. Sens. Actuators B Chem. 2007, 123, 1002–1016.

Vries, D., K. J. Keesman and H. Zwart (2010). "Luenberger boundary observer synthesis for Sturm–Liouville systems." International journal of control **83**(7): 1504-1514.

Veloso, A. C., I. Rocha and E. Ferreira (2009). Monitoring of fed-batch E. coli fermentations with software sensors. Bioprocess and biosystems engineering 32(3): 381-388.

Veloso, A. C., & Ferreira, E. C. (2017). Online analysis for industrial bioprocesses: broth analysis. In Current Developments in Biotechnology and Bioengineering (pp. 679-704). Elsevier.

Wang, J.-W., Y.-Q. Liu and C.-Y. Sun (2016). Luenberger observer design for state estimation of a linear parabolic distributed parameter system with discrete measurement sensors. 2016 12th World Congress on Intelligent Control and Automation (WCICA), IEEE.

Yan, J., X. Guo, S. Duan, P. Jia, L. Wang, C. Peng and S. Zhang (2015). "Electronic nose feature extraction methods: A review." Sensors **15**(11): 27804-27831.

Yamazoe, N. New approaches for improving semiconductor gas sensors. Sens. Actuators B 1991, 5, 7–19.

Yousefi-Darani, Abdolrahim, Olivier Paquet-Durand, Viktoria Zettel, and Bernd Hitzmann. "Closed loop control system for dough fermentation based on image processing." Journal of Food Process Engineering 41, no. 5 (2018): e12801.

Zadeh, L. A. (1965). "Fuzzy sets." Information and control **8**(3): 338-353.

Zhang, Z., J. Scharer and M. Moo-Young (1997). Mathematical model for aerobic culture of a recombinant yeast. Bioprocess engineering **17**(4): 235-240.

Zhang, D., E. A. Del Rio-Chanona, P. Petsagkourakis and J. Wagner (2019). "Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization." Biotechnology and bioengineering **116**(11): 2919-2930.

Zhang, J., Z. Deng, K.-S. Choi and S. Wang (2017). "Data-driven elastic fuzzy logic system modeling: Constructing a concise system with human-like inference mechanism." IEEE Transactions on Fuzzy Systems **26**(4): 2160-2173.

Zhao, Y. (1996). Studies on modeling and control of continuous biotechnical processes, Norwegian University of Science and Technology.

Appendices

Appendix A. Eidesstattliche Versicherung

Anlage 3

Eidesstattliche Versicherung über die eigenständig erbrachte Leistung

gemäß § 18 Absatz 3 Satz 5 der Promotionsordnung der Universität Hohenheim für die Fakultäten Agrar-, Natur- sowie Wirtschafts- und Sozialwissenschaften

1. Bei der eingereichten Dissertation zum Thema

Development of software sensors for on-line monitoring of baker's yeast fermentation process

....

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Ich habe nicht die Hilfe einer kommerziellen Promotionsvermittlung oder -beratung in Anspruch genommen.

4. Die Bedeutung der eidesstattlichen Versicherung und der strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Die Richtigkeit der vorstehenden Erklärung bestätige ich. Ich versichere an Eides Statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

25.02.21 _____   _Abdolrahim Yousefidarani_____

Ort, Datum   Unterschrift

Appendix B. Curriculum vitae

AbdolRahim Yousefi-Darani

Seemuhlenweg,15; Stuttgart; Germany
rahim_yousefi2002@yahoo.com
017647694264

Education	**Phd candidate**	**03. 2018 - Current**
	University of Hohenheim	
	Phd thesis: Development of software sensors for on-line monitoring of baker's yeast fermentation process (working title)	
	Master of science	**09. 2015 - 08. 2017**
	University of Hohenhiem	
	Food science and engineering	
	Bachelor of science	**09. 2005 - 08. 2009**
	Azad University of Khorasgan, Iran	
	Food science and engineering	

Working experience

Wissenschaftlischer mitarbeiter   **08. 2017 - Current**

University of Hohenheim

- Supervised bachelor and Master students for their thesis
- Running microbial cultivation process
- Supervised students during laboratory courses.

Quality Management Specialist   **07. 2010 - 08. 2013**

Pooran Powder, Isfahan, Iran

- Performing chemical, microbial and physical tests during all production
- Quality and hygiene control of all sections involved in the production in accordance with HACCP and food safety standards
- Reporting and documentation.

Languages skills

Farsi: Native

English: Fluent

German: Basic

Computer & programming skills

Basic knowledge of programing in Matlab and Python

Basic knowledge of HTML

Proficient in Microsoft excel

Reference

Prof. Dr. Bernd Hitzmann   Head of Process Analytics and Cereal Science department
at university of Hohenheim
Bernd.hitzmann@uni-hohenheim.de

25.01.21