

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

Abdolrahim Yousefi-Darani¹, Majharul Islam Babor², Olivier Paquet-Durand², and Bernd Hitzmann³

¹Universität Hohenheim

²University of Hohenheim

³University of Hohenheim, RG Process Analytics and Cereal Sciences

April 28, 2020

Abstract

The ethanol concentration in batch cultivation with the yeast *S. cerevisiae* was predicted on-line using a gas sensor array. Head space samples were pumped past the gas sensors array every five minutes for 10 seconds 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 will be 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 are 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 was below 0.2 g/L and 7 %, respectively. The obtained results indicate that, the model-based calibrated gas sensor array can be a cheap alternative to other tools that are used for monitoring yeast cultivations such as spectroscopy based methods.

Keyword

Gas sensor array, model-based calibration, ethanol prediction, yeast cultivation

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, Bender et al. 2017). *S. cerevisiae* can use glucose as energy and carbon source for aerobic respiration. When a critical glucose concentration is exceeded, this microorganism produces ethanol. This is known as Crabtree effect and results in a lower biomass yield which might be, depending on the goal of the cultivation, undesirable (De Deken 1966). Therefore, in order to achieve high productivity of biomass continues 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, Citterio et al. 1996).

Numerous methods attempt to measure the concentration of volatile organic compounds (VOCs) from the vapor 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 cheap chemical sensors which can be easily integrated in current production processes, thus becoming particularly suitable for the continuous monitoring of microbial fermentation processes (Jiang, Zhang et al. 2015). Recent applications of gas sensor arrays for monitoring fermentation process are reported in literature (Buratti and Benedetti 2016; Hidayat, Nuringtyas et al. 2018; Tan, Xie et al. 2018; Ghosh, Tudu et al. 2017; Li, Yuan et al. 2019; Tan, Balasubramanian et al. 2019). However, only a few works in literature demonstrate the application of gas sensor arrays for monitoring ethanol concentration during *S. cerevisiae* cultivation (Mandenius, Eklöv et al. 1997; Lidén, Mandenius et al. 1998; Bachinger and Mandenius 2001). In order to predict a specific volatile compound with a gas sensor array, chemometric modeling techniques are required. In the previous studies, the calibration methods for the chemometric models are 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 data driven calibration method, which is a time consuming task, is the model-based calibration method. 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. (Lin and Recke 2007) give a systematic approach for development of data-driven soft sensors. Model-based calibration approaches have been implemented on spectroscopy-based monitoring systems. Solle et. al (Solle, Geissler et al. 2003), as well as Paquet-Durand et al. (Paquet-Durand, Assawarajuwan et al. 2017) had used this evaluation technique for the prediction of biomass, glucose, and ethanol during a *S. cerevisiae* cultivation. Furthermore, Paquet-Durand et al. (Paquet-Durand, Ladner et al. 2017 a) applied this method for evaluation of fluorescence measurements during several parallel cultivations of *H. polymorpha* in a microtiter plate.

Based on fluorescence measurements Odman et. al (Odman, Johansen et al. 2009) and Solle et. al (Solle, Geissler et al. 2003) have evaluated yeast cultivations using glucose as substrate and developed chemometric models, one for the glucose consumption phase with concomitant ethanol production and one separate for the ethanol consumption phase (after glucose depletion). They have stated that it was difficult to use one and the same model for both phases. Paquet-Durand et al. (Paquet-Durand, Assawarajuwan et al. 2017) examined artificial neural networks for the correlation between the fluorescence spectra with glucose, biomass and ethanol concentrations. They implemented a model-based training approach for training the neural network with only using a single model. They have reported an accurate prediction of glucose and biomass (error of prediction below 5%) however the prediction error of ethanol was 10 %. This is due to ethanol not being fluorescent and 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 modeling. The main contribution of this paper can be summarized as follow:

- Design and implementation of a gas sensor array and the headspace sampling system in order to achieve accurate prediction of ethanol concentration in the liquid phase during *S. cerevisiae* batch cultivation.
- Implementation of a model-based calibration 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 in the beginning and are also determined during this procedure.

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

The remaining paper is organized 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, 7 g and 10 g of baker's yeast *S. cerevisiae* (fresh baker's yeast, Oma's Ur-Hefe) was used in BC1, BC2, and BC3, respectively. The baker's yeast was suspended into 100 mL Schatzmann medium containing 0.34 g L⁻¹ MgSO₄*7H₂O, 0.42 g L⁻¹ CaCl₂*2H₂O, 4.5 g L⁻¹ (NH₄)₂SO₄, 1.9 g L⁻¹ (NH₄)₂HPO₄, 0.9 g L⁻¹ KCl. After shaking for 10 minutes 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, 7 g/L and 9 g/L glucose for BC1, BC2, and BC3 respectively and 1 mL/L trace elements solution (0.015 g/L FeCl₃*6H₂O, 9 mg/L ZnSO₄*7H₂O, 10.5 mg/L MnSO₄*2H₂O, and 2.4 mg/L CuSO₄ 5H₂O) and 1 mL/L vitamin solution (0.06 g/L myoinositol, 0.03 g/L Ca-pantothenate, 6 mg/L thiamine HCl, 1.5 mg/L pyridoxine HCl, and 0.03 mg/L biotin). All three batch runs were operated at the same conditions, that is, a constant temperature at 30 degC and a maintained pH at 5. The aeration and agitation rates were kept constant at 3.5 L/min 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 let in a drying oven at 103 degC for 24 h. Then they were cooled down for 30 min before weighing. The supernatant of the samples after the centrifugation were examined by HPLC (ProStar, Variant, Walnut Creek, CA, USA) to determine the concentration of ethanol. The supernatant was firstly filtrated with pore size filter, 0.45 µm, polypropylene membrane (VWR, Darmstadt, Germany). Subsequently, it was injected 20 µL into a Rezex ROA-organic acid H+ (8 %) column (Phenomenex, Aschaffenburg, Germany) and operated at 70 °C with 5 mM H₂SO₄ as an eluent at 0.6 mL min⁻¹ flow rate. The concentration of ethanol was calculated by software GalaxieTM Chromatography (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 degC. For this reason, a temperature sensor (DHT22, Aosong Electronics Co., Ltd) as well as two heating elements was 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 (digital to analogue convertor) was used to convert the electrical signals to digital signals from the MOS sensors and the temperature sensor. 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. Fig. 1 illustrates 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 minutes during the cultivation process. Headspace sampling contained the following main stages: exposition stage, purging stage and sensor regeneration stage. In the exposition stage, the headspace of the bioreactor was passed into the sensor chamber for 10 seconds at a flow rate of 400 mL/min with a diaphragm pump (Schwarzer Precision, Essen, Germany). The exposition stage was followed by the purging stage. In this phase clean air was drawn in 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 for 180 seconds. 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 seconds so that the readouts from sensors reached the 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

Analog data obtained from the output voltage responses of the sensors was sampled at 2 s^{-1} frequency during the whole cultivation process. Fig.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_0) can be calculated by the following equation:

$$V_0 = \frac{R_L}{R_L + R_S} * V_C$$

Each sensor reacts differently to volatiles of the gas headspace 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 (Omatu and Yano 2016 ; Kiani, Minaei et al. 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 reduction projects the feature vector onto a lower dimensional space in order to avoid problems associated with high-dimensional, sparse datasets and redundancy (Aguilera, Lozano et al. 2012). Finally, the reduced features should be imported to the prediction model. Fig. 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 and Falasconi 2012). For this reason the following equation was applied.

Where V_i is the response (the voltage of the sensor) of the i^{th} sensor, $V_{0,i}$ 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 seconds, 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: change calculated as the difference in the final value and the baseline value of sensor response in the transient phase.
- Peak area: calculated area from the signal response in the transient phase.

In total, each measurement cycle was characterized by 6 variables (i.e., 3 sensors x 2 features per sensor). In order to quantify the amount of useful information for predicting ethanol concentration from all the variables, principle component analyses (PCA) was performed. The process of PCA is to find a new coordinate system of the mean centered data set, whose axis are perpendicular and have maximal variance in decreasing order. The direction containing the most of the variance of the data is called the first PC. The second PC carries the maximum variance of the rest data and so on. These PCs are statistically unrelated from each other (Otto, 1999; AS, 2006)

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. The process can be simulated by the following differential equations:

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

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

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

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 yield coefficients have been fixed to values which are determined by pre-runs of the process ($Y_{X/G} = 0.175 \text{ g}_X/\text{g}_G$, $Y_{E/G} = 0.473 \text{ g}_E/\text{g}_G$ and $Y_{X/E} = 0.598 \text{ g}_X/\text{g}_E$).

The specific growth rates (μ_{G0} and μ_{E0}) which are the kinetic parameters of the process model are determined during the model-based calibration procedure.

2. 6. Model - based calibration procedure

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

$$c_E = p_0 + (p_1 \times \text{PC}_1) + (p_2 \times \text{PC}_1^2)$$

Where c_E is the predicted ethanol concentration, PC_1 is the first principle component of the gas sensor array data and p_0 , p_1 and p_2 are the parameters of the 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 that 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 minimized by implementing an optimization 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 model-based calibration procedure is presented in Fig. 4.

The optimization method which was used to minimize the error of prediction is a particle swarm optimization 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 (Wang, Gandomi et al. 2014).

By applying this model-based calibration 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 which 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 characterized by the root-mean-squared error of calibration (RMSEC), standard error of calibration with respect to the maximum ethanol concentration (SEC), root-mean-squared error of prediction (RMSEP) and standard error of prediction with respect to the maximum ethanol concentration (SEP).

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:

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

\hat{Y}_p 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), 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:

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

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 discussions

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 emission from the nutrient components of the medium (Bachinger and 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 bioreactor was filled with 1 L of the medium and a glucose solution (to a final concentration in bioreactor of 7 g/L) and five sampling cycles were performed from the headspace of the bioreactor. Exactly before the 6th sampling cycle, the suspended cell was added and the sampling cycles were performed every 5 minutes until the end of the cultivation process (Fig.5).

Fig.5 (a) indicates the off-line as well as the simulated evolution of glucose, ethanol and biomass during the cultivation (BC2). The parameters of the model were acquired by least squares fitting to the off-line measurements. 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 behavior 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 not so sensitive to the other components of the cultivation medium.

3.2. Calibration models

For the model based-based calibration approach (MBC), 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 minimized by the particle swarm optimization 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 classical calibration method (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 minimized. 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.

The data in Table 1. reveals 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, therefore the parameter estimation method can be considered reliable.

As a method of assessing the fit of the calibration models to the data, the correlation plots were prepared (Fig. 6). In 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 root-mean-square error of calibration (RMSEC) and standard error of calibration (SEC) was chosen as the numerical tool for the accuracy assessment of the calibration models. The values are given in Table 2.

The results if Table 2 indicates 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 analyzed 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).

3.2. 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. Fig. 7 (a) and Fig.7 (b) presents 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 BC3 respectively. Fig. 7 (c) and Fig.7 (d) presents 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 BC1 and BC3 respectively. Fig. 7 (e) and Fig.7 (f) presents 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 BC1 and BC2 respectively.

Fig. 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 3.

Table. 3 reveals that, by using the MBC approach, the SEP of prediction is below 7 % in all cases besides when BC2 is used for calibration and the data set from BC1 is used for validation (SEP = 9.3 %). However, even though in the model-based calibration approach, during the calibration procedure of the chemometric model no off-line measurement were used, the prediction corresponds very well with the simulated values. Furthermore, in comparison with the predictions from using the classical calibration method, lower prediction errors are obtained.

The larger errors of prediction from the classical calibration method 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 highly depends on the kinetic parameters of the simulation model which are obtained from the optimization algorithm. If they are close to the real values, the process model will describe the process sufficiently accurate. Therefore, improving the optimization process can even 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 are below 10 % in all three cultivations 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 toward 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 minimizing the prediction error by optimizing the kinetic parameter values of the simulation model as well as the parameter values of the chemometric model.

The proposed model-based calibration 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 mechanical stress on the sensor as well as the number of interfering substances is smaller in the gas phase compared to the liquid phase and, furthermore, the so-called sensor fouling by cell adhesion is not a problem here, because a sterile barrier in the form of a mass filter can be easily introduced into the gas stream, this gas sensor arrays have a big potential. Compared to HPLC measurements the time delay is much shorter. Besides these advantages of using this method for ethanol detection, the proposed method is also inexpensive to implement and has just a few maintenance requirements. Therefore the gas sensor array seems to be a useful non-invasive tool for continues monitoring and might also be used for closed loop controlling of processes involving ethanol measurements.

Conflicts of interests:

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

References:

- Aguilera, T., et al. (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. doi: org/10.3390/s120608055.
- AS, C. P. (2006). "The Unscrambler Tutorials. On line at: <http://www.camo.com/downloads> U 9.
- Bachinger, T. and C.-F. Mandenius (2000). Searching for process information in the aroma of cell cultures. *Trends in biotechnology*, 18(12): 494-500. doi: org/10.1016/s0167-7799(00)01512-2.
- Bachinger, T. and C. F. Mandenius (2001). Physiologically motivated monitoring of fermentation processes by means of an electronic nose. *Engineering in life sciences*, 1(1): 33-42. doi: org/10.1002/1618-2863(200107)1:13.0.CO;2-9.
- De Deken, R. (1966). The Crabtree effect: a regulatory system in yeast. *Microbiology* 44(2): 149-156. doi: org/10.1099/00221287-44-2-149.
- Buratti, S. and S. Benedetti (2016). *Alcoholic Fermentation Using Electronic Nose and Electronic Tongue. Electronic noses and tongues in food science*, Elsevier: 291-299.
- Ghosh, S., B. Tudu, N. Bhattacharyya and R. Bandyopadhyay (2017). A recurrent Elman network in conjunction with an electronic nose for fast prediction of optimum fermentation time of black tea. *Neural Computing and Applications* 31(2): 1165-1171. doi: org/10.1007/s00521-017-3072-y.
- Hidayat, S. N., T. R. Nuringtyas and K. Triyana (2018). Electronic Nose Coupled with Chemometrics for Monitoring of Tempeh Fermentation Process. 2018 4th International Conference on Science and Technology (ICST), IEEE.

doi: [org/10.1109/ICSTC.2018.8528580](https://doi.org/10.1109/ICSTC.2018.8528580).

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

Kiani, S., et al. (2016). A portable electronic nose as an expert system for aroma-based classification of saffron. *Chemometrics and Intelligent Laboratory Systems* 156: 148-156. doi: [org/10.1016/j.chemolab.2016.05.013](https://doi.org/10.1016/j.chemolab.2016.05.013).

Lidén, H., C.-F. Mandenius, L. Gorton, N. Q. Meinander, I. Lundström and F. Winquist (1998). On-line monitoring of a cultivation using an electronic nose. *Analytica chimica, acta* 361(3): 223-231. doi: [org/10.1016/S0003-2670\(98\)00035-X](https://doi.org/10.1016/S0003-2670(98)00035-X).

Li, G., L. Yuan, X. Wang, Y. Meng, J. Li, Y. Zhao and Y. Peng (2019). Rapid Quantification Analysis of Alcohol During the Green Jujube Wine Fermentation by Electronic Nose. *IOP Conference Series: Earth and Environmental Science*, IOP Publishing. doi: [org/10.1088/1755-1315/330/5/052046](https://doi.org/10.1088/1755-1315/330/5/052046).

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

doi: [org/10.1016/j.compchemeng.2006.05.030](https://doi.org/10.1016/j.compchemeng.2006.05.030).

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

Mandenius, C. F., et al. (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.

doi: [org/10.1002/\(SICI\)1097-0290\(19970720\)55:2<427::AID-BIT20>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0290(19970720)55:2<427::AID-BIT20>3.0.CO;2-C)

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

Otto, M (1999), *Chemometrics: Statistics and Computer Application in Analytical Chemistry*.

Odman, P., C. L. Johansen, L. Olsson, K. V. Gernaey and A. E. Lantz (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. doi: [org/10.1016/j.jbiotec.2009.08.018](https://doi.org/10.1016/j.jbiotec.2009.08.018).

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

doi: [org/10.1002/elsc.201700044](https://doi.org/10.1002/elsc.201700044)

Paquet-Durand, O., T. Ladner, J. Buchs, B. Hitzmann (2017 a). 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. doi: [org/10.1016/j.chemolab.2017.10.003](https://doi.org/10.1016/j.chemolab.2017.10.003)

Solle, D., D. Geissler, E. Stark, T. Scheper and B. Hitzmann (2003). Chemometric modelling based on 2D-fluorescence spectra without a calibration measurement. *Bioinformatics* 19(2): 173-177. doi: [10.1093/bioinformatics/19.2.173](https://doi.org/10.1093/bioinformatics/19.2.173).

Spinelle, L., M. Gerboles, G. Kok, S. Persijn and T. Sauerwald (2017). Review of portable and low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds. *Sensors* 17(7): 1520. doi: [org/10.3390/s17071520](https://doi.org/10.3390/s17071520).

Tan, C., D. Xie, Y. Liu, W. Peng, X. Li, L. Ai, C. Wu, C. Wen, X. Huang and J. Guo (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. doi: org/10.1021/acs.analchem.7b05189.

Tan, J., B. Balasubramanian, D. Sukha, S. Ramkissoon and P. Umaharan (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. doi: org/10.1111/jfpe.13175.

Wang, G.-G., A. H. Gandomi, X.-S. Yang and A. H. Alavi (2014). A novel improved accelerated particle swarm optimization algorithm for global numerical optimization. *Engineering Computations*. doi: org/10.1108/EC-10-2012-0232.

Wild, R., D. Citterio, J. Spichiger and U. E. Spichiger (1996). Continuous monitoring of ethanol for bioprocess control by a chemical sensor. *Journal of biotechnology* 50(1): 37-46. doi.org/10.1016/0168-1656(96)01547-7.

Tables:

Table 1. Predicted values of the specific growth rates and the parameters of the PCR models using MBC and CCM calibration approaches.

Table 2. RMSEC and SEC for both calibration methods.	Table 2. RMSEC and SEC for both calibration methods.	Table 2. RMSEC and SEC for both calibration methods.
Cultivation	Model-based calibration	Classical calibration method
BC1	0.01 g/L 3.3 %	0.17 g/L 6.0 %
BC2	0.04 g/L 2.5 %	0.09 g/L 7.6 %
BC3	0.03 g/L 2.4 %	0.05 g/L 9.0 %

Table 3. The RMSEP and SEP obtained with the MBC and CCM approach.

Using BC1 for the calibration model	Using BC1 for the calibration model	Using BC1 for the calibration model
Validation with	Model-based calibration	Classical calibration method
BC2	0.16 g/L 9.3 %	0.1 g/L 6.0 %
BC3	0.08 g/L 6.3 %	0.1 g/L 7.6 %
Using BC2 for the calibration model	Using BC2 for the calibration model	Using BC2 for the calibration model
Validation With	Model-based calibration	Classical calibration method
BC1	0.63 g/L 5.6 %	0.34 g/L 9.7 %
BC3	0.06 g/L 5.0 %	0.11 g/L 9.0 %
Using BC3 for the calibration model	Using BC3 for the calibration model	Using BC3 for the calibration model
Validation With	Model-based calibration	Classical calibration method
BC1	0.24 g/L 6.9 %	0.29 g/L 8.5 %
BC2	0.11 g/L 6.3 %	0.13 g/L 7.2 %

Figure legends:

Figure 1. Schematic diagram of the measurement system.

Figure 2. Basic gas sensor interface circuit diagram

Figure. 3. Block diagram of the proposed ethanol prediction algorithm.

Figure 4. Flowchart of the model-based calibration procedure to get optimal parameters for the process model as well as the calibration model.

Figure 5. The characteristics of the yeast batch cultivation BC2 (a) and the raw signal response of one of the sensors (TGS 822) to the cultivation (b).

Figure 6. Predicted versus simulated ethanol concentrations using the MBC approach for all three cultivations (BC1 - BC3) as well as the predicted versus off-line ethanol concentrations using the CCM approach for all three cultivations (BC1 - BC3).

Figure 7. Predicted versus simulated ethanol concentrations using the MBC approach for all three cultivations (BC1 - BC3) as well as the predicted versus off-line ethanol concentrations using the CCM approach for all three cultivations (BC1 - BC3)

Hosted file

Figures.docx available at <https://authorea.com/users/314481/articles/444828-model-based-calibration-of-a-gas-sensor-array-for-on-line-monitoring-of-ethanol-concentration-in-saccharomyces-cerevisiae-batch-cultivation>