# Full Papers

## A New Evaluation Method for 2-D Fluorescence Spectra Based on Theoretical Modeling

By *David Geissler*, *Dörte Solle*, *Eckbert Stärk*, *Thomas Scheper*, *Herbert Märkl*, and *Bernd Hitzmann**

This article presents a new evaluation procedure of 2-D fluorescence spectra obtained during a yeast cultivation without performing a calibration measurement. The 2-D fluorescence spectra are used to predict the process variables biomass, glucose and ethanol. The new calibration procedure uses a theoretical model of these process variables, i.e., differential equations, to replace any calibration measurement. The theoretical model parameters are identified simultaneously during the calculation of the chemometric models. The root mean square error of prediction of the chemometric models with respect to off-line measurements are 1.5 g/L, 0.40 g/L and 0.56 g/L for glucose, biomass and ethanol, respectively.

### 1 Introduction

In bioprocess techniques it is important to monitor the complex biological process in order to be able to run the process as efficiently as possible. There are basically two ways for doing this, these are off-line or on-line measurements.

Off-line measurements require sampling, mostly pretreatments of the sample and measuring. This involves a lot of work and, depending on the analytical methods, a considerable time delay. Furthermore, these methods are invasive and can disturb the current process. On the other hand, off-line measurements can provide detailed analysis of a variety of compounds. The advantage of on-line measurements is the possibility of continuous and in-time measurements without additional work. The pH value, the dissolved oxygen as well as the oxygen and carbon dioxide in exhaust gas are examples, which are typically measured on-line.

However, important process variables (e.g., substrate, product and biomass concentration) are only accessible by complex analytical methods. On the other hand, using chemometric models, these process variables can be predicted from 2-D fluorescence spectra [1,2].

The main disadvantage of data-driven methods, e.g., chemometric models, is that many off-line measurements are required for the calculation of model parameters. Here, a methodology for the calibration of chemometric models is presented, which does not need further off-line measurements of the process variables, i.e., no calibration measurement is needed. The information about the process variables is provided by a theoretical model, whose model parameters will be identified during the determination of the chemometric model.

### 2 Materials and Methods

#### 2.1 Process and Process Model

The process presented here is the cultivation of *Saccharomyces cervevisiae* grown in a 1.5 L bioreactor by using the Schatzmann medium. The general course of such a cultivation run is shown in Fig. 1.

During the first phase of the cultivation the primary substrate (glucose) is consumed. In this phase the cells grow and ethanol is produced. During the second phase, when no more glucose is available, the metabolism of the cells switches to ethanol as substrate. This typical diauxic growth is discussed in detail in the literature [3,4]. The process model used for the

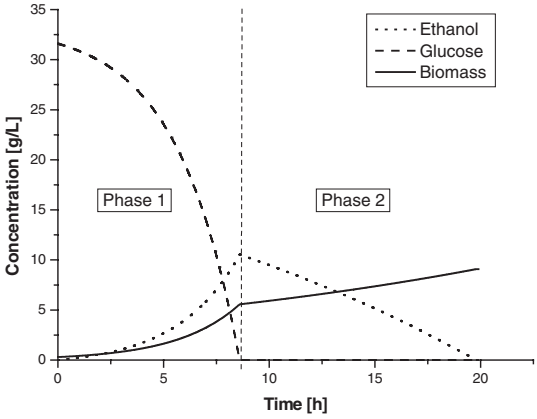


**Figure 1.** Typical concentration profile for a yeast batch cultivation. During phase 1 glucose is consumed and ethanol as well as biomass are produced. In phase 2 ethanol is converted into biomass.

[*] D. Geissler, D. Solle, Dr. E. Stärk, Prof. Dr. T. Scheper, PD Dr. B. Hitzmann (author to whom correspondence should be addressed, e-mail: Hitzmann@IFTC.Uni-Hannover.de), Institut für Technische Chemie, Universität Hannover, Callinstr. 3, D-30167 Hannover, Germany; Prof. H. Märkl, Biotechnologie I, Technische Universität Hamburg-Harburg, Denickestr. 15, D-21071 Hamburg, Germany.

 **DOI:** 10.1002/elsc.200300019  

bioprocess describes the concentration of biomass, glucose and ethanol[1)].

$$\frac{dX(t)}{dt} = \mu_1 \cdot X(t) + \mu_2 \cdot X(t) \quad (1)$$

$$\frac{dS(t)}{dt} = -\mu_1 \cdot X(t) \cdot \frac{1}{Y_{XS}} \quad (2)$$

$$\frac{dE(t)}{dt} = \mu_1 \cdot X(t) \cdot Y_{ES} - \mu_2 \cdot X(t) \cdot Y_{XE} \quad (3)$$

Differential equation (1) describes the growth of biomass $X(t)$ on glucose $S(t)$ in phase 1 and on ethanol $E(t)$ in phase 2. Therefore, two specific growth rates $\mu_1$ and $\mu_2$ must be considered. Eq. (2) describes the consumption of glucose, which correlates with the growth of the biomass in phase 1. Eq. (3) describes the production as well as consumption of ethanol, which is produced in phase 1 and consumed in phase 2. $Y_{XS}$, $Y_{ES}$ and $Y_{XE}$ are the yield coefficients with respect to the conversion between glucose, ethanol and biomass. The specific diauxic growth is considered through the fact, that $\mu_1$ is only greater then zero while glucose is present and $\mu_2$ is only greater than zero while ethanol is present. Therefore, no growth on ethanol can be observed while glucose is not completely consumed.

This bioprocess model has five unknown parameters. However, it is not possible to identify all of them through the 2-D fluorescence spectra. The changes in the spectra can be correlated with changes in the concentrations, but the prediction of absolute values cannot be made without any reference. So, the yield coefficients have to be determined by preruns of the cultivation. Furthermore, when the cultivation is started the concentrations of the process variables are known. With these requirements the calculation of the chemometric model as well as the identification of the kinetic parameters $\mu_1$ and $\mu_2$ through the 2-D fluorescence spectra can be performed.

Based on the differential equations (1) to (3) a Simulink model was created in MATLAB (Ver.6.0.0.88 R12 and Simulink 4.0, The MathWorks, Inc., Natik, USA). This model solves the given system of differential equations numerically using a fourth-order Runge-Kutta method and simulates the process variables depending on the given kinetic parameters.

### 2.2 Fluorescence Spectrometer

During the cultivation 201 spectra were recorded at an interval of 6 minutes by a 2-D fluorescence spectrometer (BioView, Delta Light & Optics, Denmark). A spectrum consists of 150 measured excitation/emission wavelength combinations with excitation wavelengths from 270 nm up to 550 nm and measured emissions from 310 nm up to 590 nm, with 20 nm intervals respectively [5]. For further computing the spectra were complemented (using the value 'NaN' in MATLAB) in order to obtain a 16 × 16 matrix for each spectrum. Furthermore, the spectra were smoothed by adjacent averaging because measurement noise has been identified in the spectra.

### 2.3 Simultaneous Identification of Kinetic Model Parameters and Chemometric Model Determination

In order to calculate the bioprocess variables biomass, glucose and ethanol from the 2-D fluorescence spectra, chemometric modeling is required. The chemometric modeling and the identification of the parameters were performed simultaneously in several steps:

1. During the first step the process variables glucose, ethanol and biomass were simulated using arbitrary but sensible start parameters $\mu_1 = 0.4\ h^{-1}$ and $\mu_2 = 0.04\ h^{-1}$. These values of the parameters are determined roughly by preruns of the process.
2. Based on the simulation, the time was determined at which the glucose is completely consumed. This marks the end of phase 1 and the beginning of phase 2. Additionally, the end of the process was determined, i.e. the time at which ethanol is completely consumed. Now, the process is divided into two phases and the spectra were divided into two sets corresponding to these phases in order to calculate independent chemometric models for these phases. Spectra, recorded after the time determined as the end of the process, were not considered at all.
3. The spectra of each phase were respectively subdivided into a calibration and a validation set. Then the spectra, just as the simulated process variables, were centered. Based on the calibration spectra and the simulated process variables, n-dimensional partial least square regressions (n-PLS) (N-way Toolbox 1.04 for MATLAB, C. A. Andersson and R. Bro, http://www.models.kvl.dk/source) [6] were performed for each variable and each phase individually. A previously performed principle component analysis had shown that two principle components are sufficient. The calculated models were applied to the validation spectra for prediction with the restriction, that the values can not be less than zero. So, we obtained five individual models, three models for phase 1 (glucose ethanol and biomass) and two models for phase 2 (ethanol and biomass), because the glucose concentration is zero and has not to be predicted.
4. The so obtained predictions were compared with the simulation. The differences were calculated and divided by the number of totally used spectra in both phases in order to consider their varying number. Then, they were squared and summed up. The obtained error value was minimized using a Nelder-Mead simplex algorithm by optimizing the parameters $\mu_1$ and $\mu_2$ (max 100 iterations).
5. At last, the complete process was predicted using all spectra and the five calculated chemometric models in their

1) List of symbols at the end of the paper.

  

corresponding phases. For validation the result was compared with off-line measurements and the root-mean-square error of prediction between prediction and off-line measurement was calculated. The simulation and the off-line measurements were evaluated in the same way.

## 3 Results

After the optimization procedure of the model parameters this evaluation procedure makes simultaneously five chemometric models to predict the biomass, glucose and ethanol concentration from the 2-D fluorescence spectra. Their prediction is shown in Fig. 2. The comparison with the off-line measurements is shown in Tab. 1. The concentrations of glucose, ethanol and biomass are well described by the models. The prediction error for glucose comes to 5 % with respect to the maximal occurring value (31.6 g/L). The errors for biomass and ethanol are 4 % and 5 % with respect to their maximal values (10 g/L, 10.7 g/L respectively).

For a final evaluation of the process parameters $\mu_1$ und $\mu_2$, the simulation based on the identified parameter combination was compared to the off-line measurements. This is shown in Fig. 3 and the errors are listed in Tab. 2. Corresponding to the result of the chemometric prediction one can see that the cultivation is well described by the simulation and the different phases are correctly identified. The error is about the same as for the prediction, this means ca. 5 % with respect to the highest occurring concentration. Therefore, the specific growth rates for the two phases are $\mu_1 = 0.352\ h^{-1}$ and $\mu_2 = 0.0504\ h^{-1}$.

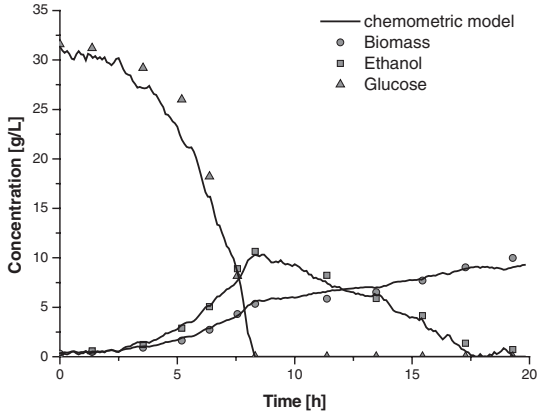


**Figure 2.** Prediction of the process variables using 2-D fluorescence spectra as well as off-line measurements.

**Table 1.** Error (root-mean-square error of prediction) of the chemometric model prediction with respect to off-line measurements.

| Process variable | RMSEP |
|---|---|
| Glucose | 1.5 g/L |
| Biomass | 0.40 g/L |
| Ethanol | 0.56 g/L |

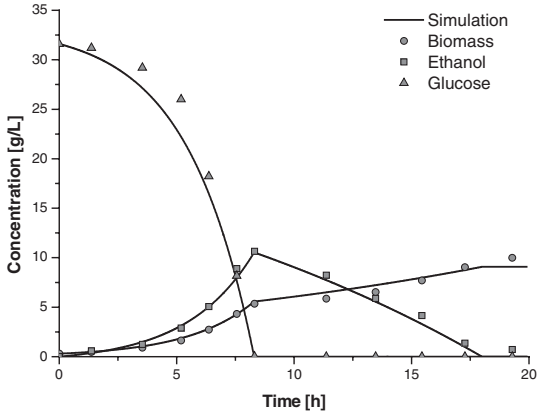


**Figure 3.** Simulation of the process variables based on the identified process parameters as well as off-line measurements

**Table 2.** Error (root-mean-square error of prediction) of the simulation with respect to off-line measurements.

| Process variable | RMSEP |
|---|---|
| Glucose | 1.4 g/L |
| Biomass | 0.40 g/L |
| Ethanol | 0.48 g/L |

## 4 Summary

With the method presented here the kinetic parameters of a theoretical process model were successfully identified, and chemometric models for the prediction of glucose, ethanol and biomass were calculated. This was achieved without any off-line measurements. All that was needed are 2-D fluorescence spectra of a cultivation that were interpreted by a MATLAB program based on an appropriate process model. Apart from the start concentration the off-line measurements can be reduced to a minimum and only serve for the verification of the chemometric models. The moment when the metabolism switches over from glucose to ethanol can be identified more precisely by spectroscopy because of the numerous measurements. Furthermore, errors of measurement and noise can be filtered more effectively to ensure correct results. Although the results in Fig. 2 are based on smoothed spectra the noise can still be seen. But in this case spectra were recorded only every six minutes. The BioView spectrometer is capable of recording one spectrum per minute. This makes it possible to better smooth in order to enhance the quality of the prediction and identify the phases more precisely.

Received: March 26, 2003 [F 019]

## Symbols used

| | | |
|---|---|---|
| $X(t)$ | [g/L] | biomass concentration |
| $S(t)$ | [g/L] | glucose concentration |

  

| | | |
|---|---|---|
| $E(t)$ | [g/L] | ethanol concentration |
| $\mu_1$ | [$h^{-1}$] | specific growth rate growing on glucose |
| $\mu_2$ | [$h^{-1}$] | specific growth rate growing on ethanol |
| $Y_{XS}$ | [g/g] | yield coefficient glucose to biomass |
| $Y_{ES}$ | [g/g] | yield coefficient glucose to ethanol |
| $Y_{XE}$ | [g/g] | yield coefficient ethanol to biomass |

## References

[1] B. Hitzmann, T. Pekeler, C. Lindemann, S. Marose, T. Scheper, Computer Applications in Biotechnology, CAB7 – 7th Int. Conf., Osaka **1998**.

[2] D. Solle, C. Protsch, R. Ulber, K. Seja, B. Willke, R. Faurie, B. Hitzmann, T. Scheper, *Optimierung der industriellen Tryptophanproduktion, Sensorik*, Spektrum Akademischer Verlag, Heidelberg **2001**.

[3] S.-O. Enfors, J. Hedenberg, K. Olsson, *Bioprocess Eng.* **1990**, *5*, 191.

[4] K. Schügerl, K.-H. Bellgardt, *Bioreaction Engeneering – Modeling and Control*, Springer-Verlag, Berlin Heidelberg **2000**.

[5] B. Hitzmann, S. Marose, C. Lindemann, T. Scheper, Computer Applications in Biotechnology, CAB7 – 7th Int. Conf., Osaka **1998**.

[6] R. Bro, *Ph. D. Thesis*, Royal Veterinary and Agricultural University, Kopenhagen **1998**.

*This paper was also published in German in Chem. Ing. Tech.* ***2002****, 74 (8), 1167.*

  