On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by \textit{S. cerevisiae}

Dominik Krämer $^*$ Rudibert King $^{**}$

$^*$ Chair of Measurement and Control, Technische Universität Berlin, Germany (e-mail: Dominik.Kraemer@tu-berlin.de).
$^{**}$ (e-mail: Rudibert.King@tu-berlin.de).

Abstract: In the early process development, feeding trajectories and cultivation conditions are altered in order to maximise the desired product. This is in contrast to industrial cultivations, where variations are to be eliminated. In order to fully understand and safely run a new process, real-time information are beneficial. Near-infrared spectrometer that are coupled to a fermenter offer the possibility to gather information on-line. The NIR data can be transformed via partial least squares modeling to estimate substrate and biomass concentrations. At low concentrations and under changed cultivation conditions during process development, however, the estimates may differ from reality. This uncertainty may be reduced by integrating biological and physico-chemical knowledge in the on-line estimation. In this contribution, we present a hybrid approach of near-infrared (NIR) spectroscopy and nonlinear model-based state estimation to enable an improved quality in the on-line estimation of substrates and biomass in a yeast cultivation. As only three cultivations are needed for calibration, on-line state estimation is available in the early stage of development for this process. This approach is compared to the use of both methods separately for estimation of biomass, ammonium, glucose, phosphate and ethanol in cultivations of \textit{S. cerevisiae}.

Keywords: cultivation process, nonlinear state estimation, NIR spectroscopy, hybrid modeling, bio-control, EKF, PLS

1. INTRODUCTION

In cultivation processes, manual measurements are rarely taken and are costly in terms of labor. Using a feedback controller to control the level of substrate concentrations leads to far better results than a simple feedforward scheme but requires the knowledge of the current state. Based on a detailed process model, an on-line estimation can be achieved by nonlinear state estimation employing at-line data shown, e.g. by Kawohl et al. (2007). Beside model-based methods, on-line near-infrared (NIR) spectroscopy can be exploited to characterize the cultivation medium. In recent years, the analytes biomass, glucose, ammonium and ethanol have been measured via NIR or mid-infrared measurements, reviewed by Landgrebe et al. (2010). As the spectra are very complex, multivariate methods are necessary to obtain useful information. These multivariate models are susceptible to overfitting and should not be used in cultivation conditions different to those of the calibration. As the Performance of NIR spectroscopy can only be assessed using experiments, this contribution addresses the application rather than the general concept. The combination of multivariate analysis of NIR spectra and classical model-based estimation offers the chance to obtain more accurate and reliable estimates compared to the use of each single technique alone. The focus of this contribution is on the early application of on-line state estimation that enables model-based control such as on-line optimization as shown, e.g., by Kawohl et al. (2007) and De la Hoz Siegler et al. (2012).

In the first part of this contribution, details to the performed cultivations are given, followed by the presentation of the employed numerical methods that are partial least squares (PLS) modeling and Extended Kalman Filtering (EKF). Afterwards, both the PLS model and the 1st-principle model are presented and discussed. In the last section, a comparison between simple on-line prediction by PLS based on the NIR spectra, an EKF only based on at-line data and a hybrid approach using the on-line PLS data within the EKF framework is drawn.

2. MATERIAL AND METHODS

2.1 Experimental Setup

A genetically modified strain of \textit{S. cerevisiae} that produces the enzyme Endopolygalacturonase (EPG) is cultivated in a 15 L cultivation vessel (Biostat C, Braun Biotech Inc., Germany) using a chemically defined medium. The dissolved oxygen is controlled by the stirrer speed to maintain oxygen saturation of 50\% starting with a minimal stirrer speed of 400 rpm and an air sparging rate of 10 L/min. The pH and temperature are held constant at 6 and 30\°C, respectively. On-line NIR measurements are taken by the
BioPAT Spectro PMD500 (1 mm gap width) made by Sartorius Stedim Biotech GmbH that is directly attached to the reactor. In-line measurements are taken every three seconds. The high sampling rate permits the reduction of noise due to particles and bubbles via averaging.

The feeding rates of the fed-batch cultivations are realized by PI-controlled pump-scale-systems. Substrate concentrations in the cultivation broth are determined photometrically using the phosphate FS and glucose GOD FS from DiaSys, Enzytec fluid by Thermo Fisher Scientific for ethanol, and the Berthelot reaction described by Rhine et al. (1998) for ammonium, applying fourfold determination in each case. The quantity of EPG is derived via the enzyme activity using the method described in Glanuche et al. (2016). Processing of the samples, with exception of the optical density and biomass, is conducted after the end of the batch. The biomass is determined trough filtering a defined volume of broth through a pre-weighed 1.5 μm filter, rinsing with water, drying and dividing the difference in filter weight through the filtered volume. The optical density is measured at 600 nm with proper dilution in order to obtain an absorbance from 0.2 to 0.3.

2.2 Multivariate Data Analysis

As there is no specific peak of the pure components at defined wavelengths in the NIR absorbency spectrum, multivariate data analysis is used to extract information from the spectra. Partial least squares regression is a standard tool that is used to correlate the reference measurements with the spectroscopic information in the spectra (Hutchinson et al. (2006)). Out of the various algorithms, the SIMPLS implementation is chosen, which is detailed in Jong (1993) and implemented in MATLAB 2011. The multiple-input multiple-output model predicts the concentration of ammonium, phosphate, glucose, ethanol and the optical density based on the spectral information of the absorption in the NIR region from 1050 to 1650 nm. In order to reduce measurement noise, the spectra are averaged with a sliding window of 100 measurements. The use of first and second derivatives of the spectra was analyzed as well as the scaling of the spectra but did not lead to further improvement in the model quality and is therefore discarded.

2.3 Nonlinear State Estimation

In order to show that even the standard nonlinear state estimation tools are suitable for the hybrid approach, the Extended Kalman Filter (EKF) is selected, see Gelb (1974). As the measurements are usually taken from discrete samples but the model is derived from continuous mass balances, the continuous-discrete formulation of the EKF is used. The equations are given for later reference.

In the time update, when no measurements are available, the states are determined by integration of the model

$$\frac{dx(t)}{dt} = f(\hat{x}(t), u(t), p), \quad \hat{x}(0) = \hat{x}(t_{k}^+)$$

with the states $x \in \mathbb{R}^{n_x}$, the inputs $u \in \mathbb{R}^{n_u}$, the parameters $p \in \mathbb{R}^{n_p}$. Integration is performed from $\hat{x}(t_{k}^+)$ to $\hat{x}(t_{k+1}^+)$. Beside the states, the propagation of uncertainty is of interest, expressed by the differential equation system of the state covariance matrix $P_x$,

$$\dot{P}_x(t) = FP_x + P_x F^T + Q \text{ with } P_{x,0} = P_x(t_{k}^+).$$

which is integrated simultaneously. The process noise spectral density $Q$ is a measure for the model uncertainty and will be specified in the following. Once a measurement $k+1$ is available, the states are corrected in the so-called measurement update

$$P_x(t_{k+1}) = (I - KH)P_x(t_{k+1})(I - KH)^T + KRK,$$

that ensures $P_x(t_{k+1})$ to be positive definite (Gelb (1974)).

Estimation of matrices $Q, P_x(t_0)$ and $R$: The process noise spectral density matrix $Q$ of the states is usually held constant and often estimated via trial and error or must be tuned via optimization to improve the performance of the EKF (Valappil and Georgakis (2000)). In batch processes, the states change exponentially during growth and, therefore, need variable entries in $Q$. As shown by Schneider and Georgakis (2013) the EKF with well chosen initial values $\hat{x}(t_0)$ and $P_x(t_0)$ and dynamically changing $Q$ shows similar results compared to other methods as, e.g., the Moving Horizon State Estimation. In this contribution, we implement the linear method proposed by Valappil and Georgakis (2000) to have a well-designed approximation of the model uncertainty $Q$ without the need for tuning

$$Q = \frac{\partial f}{\partial p}CV_p \frac{\partial f^T}{\partial p}.$$  

Here, $CV_p$ is the parameter covariance matrix which can be approximated by the inverse of the Fisher information matrix $FIM$, that is a lower bound to $CV_p$, see Walter and Pronzato (1997)

$$CV_p \geq FIM^{-1} = \left(\sum_{k=1}^{n_{\text{sample}}} FIM_k\right)^{-1}.$$  

With the Fisher information matrix of one sample $i$ being calculated as

$$FIM_i = \left[\left(\frac{dy}{dp}\right)^T CV_p^{-1} \left(\frac{dy}{dp}\right)\right],$$

$$\frac{dy}{dp} = \frac{\partial y}{\partial x}S(t) + \frac{\partial y}{\partial p} \text{ with } S(t) = \frac{dx}{dp}$$

The sensitivities $S(t)$ are calculated by integrating the sensitivity equation system (Biegler (2010)).

$$\frac{dS(t)}{dt} = \frac{\partial f}{\partial x}S(t) + \frac{\partial f}{\partial p} \text{ with } S(t_0) = \frac{dx_0}{dp}.$$
Important for the assessment of uncertainty is the initial state covariance matrix $P_x(t_0)$. It may be assessed based on the difference of the expected and actual measured states as proposed by Valappil and Georgakis (2000)

$$P_x(t_0) = \text{diag} \left( (\hat{x}_0 - \bar{x}_0)(\hat{x}_0 - \bar{x}_0)^T \right). \quad (13)$$

The measurement noise $R$ depends on the measurement procedure. In this case, samples have to be diluted to apply substrate analysis and measure the OD. $R$ therefore varies depending on the level of the measured quantity. Based on past data of analytical assays, linear regression was performed to estimate measurement noise, as described by Heine et al. (2003). Because the different measurement techniques are performed independently, no correlations occur, so that off-elements in $R$ are set to zero.

$$R_{j,j}(k) = (a_j \cdot y_j(k) + b_j)^2 \quad (14)$$

The employed coefficients in (14) for the analytical methods are given in the next section. For the concentrations calculated by the PLS model, the root mean square error of validation (RMSEV) is used as approximation of the standard deviation to deduce the corresponding entries in $R$. Convergence of the filter was first tested in simulation studies as for this nonlinear problem an analytical proof of convergence would be difficult if possible at all.

### 3. MODEL IDENTIFICATION

Four cultivations were performed to study the enzyme production. During three of these cultivations, NIR spectra were collected. The cultivations 1, 2 and 4 served for parameter estimation of the nonlinear process model.

#### 3.1 Multivariate PLS model

The multivariate model predicts the substrate and biomass concentrations based on on-line NIR spectra. The basis of the calibration was cultivation 1 and 2 with three samples added from cultivation 3 in order to account for large values of OD, glucose and ethanol in cultivation 3. As five substances should be modeled, it is expected that at least five factors are necessary for successfully modeling the substances. Here, seven factors were employed with the aim to attain a good compromise between model mismatch and validation error. Unfortunately, the EPG concentration was not high enough to be identified in the NIR spectra.

In fig. 1, the model predictions are plotted over the reference values together with the 2-$\sigma$ confidence interval of the measurement noise. Here, the problematic distribution of the concentration levels of glucose, ethanol and the optical density are apparent. Few measurements are available at moderate levels. This is both due to the lack of sample drawings at night as well as the feeding strategy that was chosen to maximise enzyme production and was therefore not optimal for the calibration of the NIR spectra.

**Model uncertainty:** The root mean square error of validation

$$RMSEV = \sqrt{\sum (y - y_{PLS})^2} / n_v \quad (15)$$

with the number of validation samples $n_v$, serves as respective entry in $R$ for the EKF and are given later in the case study. As can be seen in fig. 1 most substrates and the OD can be matched in the order of the measurement uncertainty. This is in contrast to regions of small glucose concentration and of small OD, where the model mismatch is significantly larger than the measurement error. Knowing that a lot of measurements are taken in this concentration range, no further improvement seems possible with the linear PLS exploiting raw data. Nonlinear calibration models methods could lead to better results concerning low concentrations if more measurements were available. As this is not the case and further calibration effort would be necessary this is not pursued here.

#### 3.2 Nonlinear Process Model

The unstructured process model consists of mass balances for ammonium (Am), glucose (Glc), phosphate (Ph), and ethanol (Et), enzyme EPG (E) and biomass (X), see (16) to (21). In addition, the broth volume that changes due to feeding $u$, pH-compensator fluids $u_{\text{ph-Comp}}$ and sampling $u_{\text{out}}$ is incorporated based on a mass balance assuming constant density, see (22). The values of the parameters $p$ are determined with a modified maximum likelihood estimation that handles different measurement noise depending on the measurement range as explained in Heine et al. (2003) using the optimizer SNOPT7 documented in Gill et al. (2008). With this approach it cannot be guaranteed that a global optimum is found.

$$\begin{align*}
\dot{\text{Am}} &= \text{Glc} - \text{Ph} - \text{Et} - \text{X} \\
\dot{\text{Glc}} &= \text{Am} + \text{Ph} - \text{Et} - \text{X} \\
\dot{\text{Ph}} &= \text{Am} + \text{Ph} - \text{Et} - \text{X} \\
\dot{\text{Et}} &= \text{Am} + \text{Ph} - \text{Et} - \text{X} \\
\dot{\text{X}} &= \text{Am} + \text{Ph} + \text{Glc} + \text{Ph-Comp} - \text{out} \\
V &= \text{Am} + \text{Ph} + \text{Glc} + \text{Ph-Comp} - \text{out}
\end{align*} \quad (16)$$

All states except the volume are measured as concentrations so that the measurement equations consist of terms $c_i = \frac{n_i}{V} \quad (23)$

Additionally, the optical density $OD$ can be used as at-line measurement of the biomass concentration.
It has to be noted that this linear approximation is only valid in the range from 1 to 25 g/L. 

The reaction rates are given in the following. For enzyme production, low concentrations of ammonium and phosphate are to be avoided. No influence of these substrates on growth could be observed in the conducted experiments. With respect to glucose, the simple and often used Monod kinetic is employed (Bastin and Dochain (1990)) to describe the limiting effects of glucose on biomass production.

\[
\frac{r}{c} = \frac{\mu_{\text{max},X} c_{\text{Glc}}}{c_{\text{Glc}} + K_{\text{Glc},X}} 
\]

(25)

Without the need for parameter estimation of λ and \(K_{\text{Glc},\text{Et},P}\) it is assumed that at high glucose concentrations, ethanol is produced with the rate

\[
\frac{r}{c} = \frac{\mu_{\text{max},\text{Et},P} c_{\text{Glc}}}{c_{\text{Glc}} + K_{\text{Glc},\text{Et},P}} 
\]

(26)

and degraded once the glucose concentration is low, with

\[
\frac{r}{c} = \frac{\mu_{\text{max},\text{Et},X} c_{\text{Et}}}{c_{\text{Et}} + K_{\text{Et},\text{Et}}} \cdot \frac{K_{\text{Glc},\text{Et}}}{c_{\text{Glc}} + K_{\text{Glc},\text{Et}}} 
\]

(27)

Finally, based on preliminary studies, a growth associated production of enzyme is proposed. The maximum production rate is achieved at moderate growth rates, which leads to

\[
\frac{r}{c} = \mu_{\text{max},E} \cdot \exp \left( -\frac{1}{2} \cdot \frac{r_{X} - r_{X,\text{opt}}}{k_{E}} \right)^{2} \]  

(28)

Both the measurements based on the PLS model and the manual measurements were used for parameter estimation. Identified parameters and their corresponding standard deviation are given in tab. 1.

Table 1. Model parameters with uncertainties

| Parameter | Name | Unit | Value | Deviation |
|-----------|------|------|-------|-----------|
| \(r_{X,\text{max}}\) | \(r_{X,\text{max}}\) | 1/h | 0.187 | 0.0023 |
| \(K_{\text{Glc},X}\) | \(K_{\text{Glc},X}\) | g/L | 0.0041 | 0.0015 |
| \(r_{\text{Et},\text{max},\text{Et},P}\) | \(r_{\text{Et},\text{max},\text{Et},P}\) | 1/h | 0.8908 | 0.0199 |
| \(r_{\text{Et},\text{max},\text{Et},X}\) | \(r_{\text{Et},\text{max},\text{Et},X}\) | 1/h | 0.0193 | 0.0004 |
| \(K_{\text{Et},\text{Et}}\) | \(K_{\text{Et},\text{Et}}\) | g/L | 0.0001 | 0.0026 |
| \(K_{\text{Glc},\text{Et}}\) | \(K_{\text{Glc},\text{Et}}\) | g/L | 0.0300 | 0.02925 |
| \(Y_{\text{Am},X}\) | \(Y_{\text{Am},X}\) | g/g | 0.0544 | 0.0008 |
| \(Y_{\text{Ph},X}\) | \(Y_{\text{Ph},X}\) | g/g | 0.0208 | 0.0005 |
| \(Y_{\text{Glc},\text{Et}}\) | \(Y_{\text{Glc},\text{Et}}\) | g/g | 2.5615 | 0.0095 |
| \(Y_{\text{Et},X}\) | \(Y_{\text{Et},X}\) | g/g | 1.8657 | 0.0094 |
| \(\lambda\) | \(\lambda\) | g/L | 1.402 | 0.0262 |
| \(K_{\text{Glc},\text{Et},P}\) | \(K_{\text{Glc},\text{Et},P}\) | g/L | 1 | - |

The parameters that describe the consumption of ethanol are associated with a high uncertainty and indicate overfitting. As (26) and (27) are based on biological reasoning, the high standard deviation is accepted here. All other parameters concerning growth and substrate consumption are significant enough. As the inverse Fisher matrix is only a lower bound for the parameter uncertainty, the true parameter uncertainty may be higher than calculated. In order to have more realistic predictions of uncertainty in the EKF a relative uncertainty of 2 percentage points is added to all parameter standard deviations. Please note that due to confidentiality, no scaling in the graph and no parameter values for the production of EPG are given.

3.3 Comparison of model quality

The evolutions of feeding rates and concentrations of one out of the four calibration cultivations are shown in fig. 2. Together with the analytical reference values the course of the fitted nonlinear model is given. In the same graph, the values generated by the PLS model based on the NIRSpectra are displayed. Due to errors on the data collecting computer, not the whole time course of PLS prediction is available in fig. 2. The physical model lacks exact description of the growth process and the ammonium consumption at the end of the cultivation.

Table 2. Coefficients for measurement noise as defined in (14)

| Name | a (\(-\)) | b (unit) | unit |
|------|-----------|---------|------|
| Ammonium | 0.06 | 0.01 | g/L |
| Phosphat | 0.07 | 0.01 | g/L |
| Glucose | 0.06 | 0.25 | g/L |
| Ethanol | 0.05 | 0.25 | g/L |
| Biomass | 0.04 | 0.03 | g/L |
| OD | 0.07 | 0.2 | - |

Fig. 2. Comparison of experimental cultivation data (black circles with errorbar), simulated concentrations (grey lines) and PLS results (black lines)

The PLS prediction of Ethanol is largely noisy towards the end of the cultivation and differs significantly from the offline measurements. As the concentrations of ammonium and phosphate do not change very much, the PLS model may lack predictive capacity.

All in all one may conclude that both the PLS model and the physical model need a broader data base and a more complex model structure in order to generate more reliable predictions if used on their own. In spite of low model quality a hybrid approach allows more reliable online state estimation as will be pointed out in the following section.

4. ASSESMENT OF HYBRID APPROACH

In the previous section, off-line manual measurements were used to fit both models. As these data are not available on-line, the real-time performance of the hybrid approach will be tested and evaluated using only at-line and on-line
data. Here, three scenarios are investigated. In Scenario A, a PLS model is applied to the on-line NIR data to predict the course of the cultivation. In scenario B, the model-based state estimator is used, employing only the at-line measurement of OD. In Scenario C, a combination of both methods is applied, in which the EKF is supplied with the PLS predictions, respecting the relatively high measurement noise variance of the PLS model. Please note that beside cultivation 1 and 2, three samples of cultivation 3 were used for identification of the PLS model.

The feeding trajectories and the development of the volume of the investigated cultivation are given in fig. 3 and apply for all scenarios. The feeding concentrations were 80 g/L ammonium, 40 g/L phosphate and 450 g/L glucose.

4.1 Prediction by PLS - Scenario A

The PLS predictions are given in fig. 4 as solid black lines named ‘PLS input’. Dynamic changes as for example for the glucose depletion around 15h are detected and could be confirmed by reference analytics afterwards. However, the predicted glucose values after 16h are not in accordance with manual measurements and biological knowledge, highlighted in fig. 6.

Moreover, the PLS model predicts a decreasing level of optical density from 0.5h to 4h. This development contradicts the at-line measurements of OD. From a biological point of view, the biomass should grow exponentially as long as no substrate has reached a limiting concentration.

4.2 Prediction by EKF and at-line OD – Scenario B

In fig. 5 the results of the EKF using the at-line measurement of the OD are presented. Additionally, the predicted uncertainty is plotted together with the state. At first sight the EKF predicts the concentrations satisfactorily. As the continuous PLS-generated data were used in the parameter estimation, the dynamic phenomena can be adequately reflected. Since most manual measurements are within the calculated uncertainties, the described EKF design worked well. One can critically remark that the moment of glucose limitation at \( t = 16 \)h as predicted by the EKF is not captured correctly. The exact moment was determined separately using off-gas analysis and showed the prediction in scenario A to be correct, see fig. 6.

4.3 Combination of EKF and PLS – Scenario C

If the on-line PLS-data is filtered by the EKF, all inconsistencies of scenarios A and B are eliminated, see fig. 4 and 6. The moment of glucose depletion that could not be correctly predicted in scenario B is correctly discovered in scenario C based on the on-line measurements. Furthermore, the falsely predicted higher glucose levels of the PLS model after depletion are corrected by the EKF to values which are confirmed by the off-line measurements. The nonphysically decreasing OD-values, predicted by the PLS model, are adjusted by the EKF to provide physically reasonable values.

Comparing the RMSEV values of scenario A, B and C in tab. 3, calculated with the off-line measurements, one could conclude that the NIR estimation is the worst and does not improve the estimation compared with an EKF.
Table 3. Validation error of different scenarios

| RMSEV     | Scenario A | Scenario B | Scenario C |
|-----------|------------|------------|------------|
| Glucose / g/L | 3          | 0.19       | 0.16       |
| Ethanol / g/L | 2.6        | 1.7        | 1.6        |
| Phosphate / g/L | 0.19       | 0.12       | 0.18       |
| Ammonium / g/L | 0.19       | 0.18       | 0.16       |
| Optical Density | 3.9        | 5.3        | 4.4        |

applying at-line OD measurement. But the comparison of fig. 4 and fig. 5 shows the superiority of the PLS model to identify dynamic phenomena as for example the glucose depletion. Furthermore the at-line measurements require manual sample drawing and are therefore costly in labor and influenced by dead time.

The predicted uncertainty of the EKF may be an under-estimation as some off-line measurements are outside the predicted 2-σ interval. Possibly, the calculated approximations of the parameter uncertainties are too small even if 2 percentage points were added. The parameter covariance matrix can be higher than the inverse Fisher matrix, as it is a lower bound. Furthermore, the symmetrical uncertainty description cannot describe the nonlinear behavior correctly. In fig. 6, the lower 2-σ interval surpasses 0 g/L and thus violates physical constraints. Nonlinear state estimation techniques as, e.g., the sigma-point Kalman filter presented by van der Merwe et al. (2004) and applied to biological processes by Chai et al. (2007) could circumvent these problems and will be addressed in the future.

5. CONCLUSION

The combination of two well-known and widely applied techniques, namely PLS and EKF, leads to improved online state estimation even if the two methods show large uncertainty when applied separately. A better PLS model could be obtained through an iterative cycle of parameter estimation and PLS fitting using simulation results, as demonstrated by Solle et al. (2003). This, however, requires the integration of the multivariate modeling in the nonlinear parameter estimation. The methodology presented here enables the user to use commercial PLS tools and still get a more reasonable estimate even with few data. Based on reliable state estimation, feedback control may now be used for further cultivations in order to improve model quality and enzyme production.

ACKNOWLEDGEMENTS

Support with the BioPAT spectro PMD500 by Sartorius Stedim Biotech GmbH, Germany is gratefully acknowledged as the supply of the organism by Organobalance GmbH.

REFERENCES

Bastin, G. and Dochain, D. (1990). *On-line estimation and adaptive control of bioreactors*, volume 1 of *Process measurement and control*. Elsevier, Amsterdam and New York.

Biegler, L.T. (2010). *Nonlinear programming: Concepts, algorithms, and applications to chemical processes*. SIAM, Philadelphia.

Chai, Q., Furenes, B., and Lie, B. (2007). Comparison of state estimation techniques, applied to a biological wastewater treatment process. In *10th Int IFAC Symp on Comput Appl Biotech*, 353–358.

De la Hoz Siegler, H., McCaffrey, W.C., Burrell, R.E., and Ben-Zvi, A. (2012). Optimization of microalgal productivity using an adaptive, non-linear model based strategy. *Bioresource Technol*, 104, 537–546.

Gelb, A. (1974). *Applied optimal estimation*. M.I.T. Press, Cambridge and Mass.

Gill, P.E., Murray, W., and Saunders, M.A. (2008). User’s Guide for SNOPT Version 7: Software for Large-Scale Nonlinear Programming.

Glauche, F., Genth, R., Kiesewetter, G., Cuda, F., Goelling, D., Raab, A., Lang, C., and Neubauer, P. (2016). Rapid process development for recombinant protein production in Saccharomyces cerevisiae. *submitted*.

Heine, T., Kawohl, M., and King, R. (2003). Parameterschätzung für nichtlineare dynamische Modelle auf Basis stark verrauschter, statistisch unvollständig bestimmter Messdaten. *Chem-Ing-Tech*, 75(5), 543–550.

Hutchison, D., Kanade, T., Kitterl, J., Kleinenberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Saunders, C., Grobelnik, M., Gunn, S., and Shawe-Taylor, J. (2006). *Subspace, Latent Structure and Feature Selection*, volume 3940. Springer Berlin Heidelberg, Berlin and Heidelberg.

Jong, S.d. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemomert Intell Lab*, 18(3), 251–263.

Kawohl, M., Heine, T., and King, R. (2007). Model based estimation and optimal control of fed-batch fermentation processes for the production of antibiotics. *Chem Eng Process*, 46(11), 1223–1241.

Landgrebe, D., Hanke, C., Höpfnner, T., Beutel, S., Hitzmann, B., Scheper, T., Rhiel, M., and Reardon, K.F. (2010). On-line infrared spectroscopy for bioprocess monitoring. *Appl Microbiol Biotechnol*, 88(1), 11–22.

Rhine, E.D., Mulvaney, R.L., Pratt, E.J., and Sims, G.K. (1998). Improving the Berthelot Reaction for Determining Ammonium in Soil Extracts and Water. *Soil Sci Soc Am J*, 62(2), 473.

Schneider, R. and Georgakis, C. (2013). How To NOT Make the Extended Kalman Filter Fail. *Ind & Eng Chem Res*, 52(9), 3354–3362.

Solle, D., Geißler, D., Stark, E., Scheper, T., and Hitzmann, B. (2003). Chemometric Modelling based on 2D-Fluorescence Spectra without a Calibration Measurement. *Bioinformatics*, 19(2), 173–177.

Valappil, J. and Georgakis, C. (2000). Systematic estimation of state noise statistics for extended Kalman filters. *AICHeJ*, 46(2), 292–308.

van der Merwe, R., A. Wan, E., and I. Julier, S. (2004). Sigma-point Kalman filters for nonlinear estimation and sensor-fusion: Applications to integrated navigation. *P AIAA GN&EC*.

Walter, E. and Pronzato, L. (1997). Identification of parametric models from experimental data. Communications and control engineering. Springer and Masson, Berlin and New York and Paris.