Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

State Estimation in Biotechnological Processes Using a Software-Sensor Combining Full-Horizon Observer and Neural Networks

Joachim Hörrmann * Michael Kräling Dorothee Barth Helmut Röck

* Christian-Albrechts-University Kiel, Kaiserstrasse 2, 24143 Kiel, Germany (Tel: 0049-431-8806285; e-mail: jh@tf.uni-kiel.de)

Abstract:
This paper presents an innovative method for the online determination of biomass in fermentation, using a combination of model-based full-horizon observer and Neural Network. The performance of the Neural Network depends highly on correct initial conditions of the unknown process values. Unfortunately, in biological processes it is impossible to guarantee reproducible initial conditions. On the contrary, the variations in the innoculated amount of bacteria oscillate between 30% in lab scale and more than 100% in industrial applications.
To reduce the effect of these variations to the Neural Network, we herein propose the use of an optimization based estimator to determine the initial conditions of the process values online in early process stages in order to improve the estimation results of the Neural Network.

Keywords: Parameter and State Estimation, Modelling and Identification, Neural Network, Full-Horizon Observer, Biotechnology, *Streptococcus thermophilus*

1. INTRODUCTION

The bacterium *Streptococcus thermophilus*is an important agent in dairy industry (Gronau (2005); Hols et al. (2005)). It is used in many different production processes, e.g. as a starter culture in the manufacture of yoghurt or cheese, where its main purpose is acidifying the milk by metabolizing the disaccharide lactose into the acid lactate.
Up to now the fermentation processes do not have the same high level of automation as found in the pharma industry, where modern process control is already regarded essential. Instead, the production of *Streptococcus thermophilus* is to a large extend still carried out in traditional ways, where experience and experimental data are of high importance for process development.
Since the demands on quality and quantity of this bacteria increase, the starter culture producers have to improve the fermentation processes in order to fullfill these higher requirements. Concepts of modern process control must be applied to achieve better performance of the process, to obtain constant output rates and to achieve constant high quality. A bottleneck for process control is the limited possibility of measuring important process values, like biomass, the substrate lactose and the product lactate, in real-time. Since these values are fundamental for the evaluation of the actual process, further concepts have to be developed in order to gain online information about these interesting process states. A suitable method of achieving this goal is the use of software-sensors in order to estimate the desired process states in real time.
This contribution introduces a combination of a model-based observer, the full-horizon observer, with the state estimation using a Neural Network. These two methods are brought together in one software-sensor in order to improve the state estimation of unknown process variables in starter culture fermentation.

2. REAL-TIME-MEASUREMENT IN THE FERMENTATION OF *STREPTOCOCCUS THERMOPHILUS*

Since the performance of a high-level process control depends on the knowledge of process determining values, several different process states have to be determined online during the fermentation in order to influence the process in the desired way. Several sensors are available in order to measure various process values in real-time.
In the considered process the pH-value of the medium as well as its conductivity and the amount of base added to the medium in order to control the pH-value are used as online measurement values.
Unfortunately, none of the interesting process states, the substrate lactose, the product lactate nor the biomass, can be measured directly. The concentration of the medium components are determined offline by analyzing manually taken samples using an HPLC, while the biomass is determined by cell count two days after the sample has been taken.
However, it is possible to correlate the amount of produced lactate via a second degree polynomial with the pH-value and the conductivity, respectively.
Since the pH-value is, for process specific reasons, kept at an almost constant level in later process times, which is done by adding base to the medium, the correlation

978-1-1234-7890-2/08/$20.00 © 2008 IFAC

10.3182/20080706-5-KR-1001.0365

between pH-value and lactate concentration is only valid in early process stages. Due to the increasing amount of base, the conductivity of the medium increases, which motivates the use of the correlation between the product lactate and the conductivity of the medium in later process times. The variation in time of these two key variables during the fermentation process is presented in figure 1, while the already mentioned correlation is shown in figure 2.

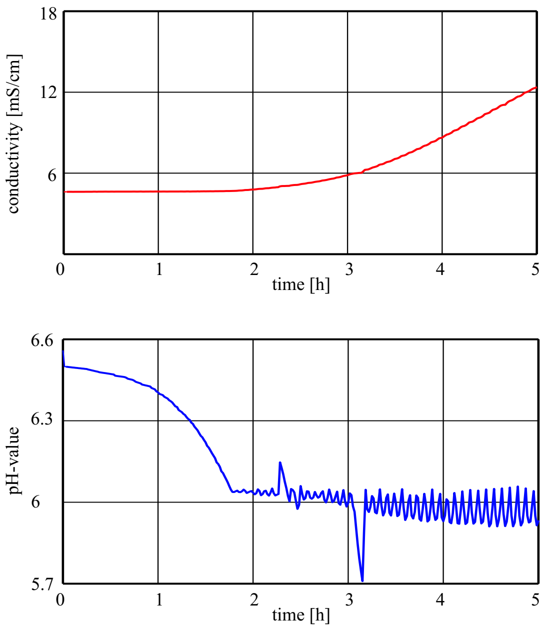

Fig. 1. The variation in time of the conductivity (upper picture) and the pH-value (lower picture), respectively during a fermentation process of *Streptococcus thermophilus*.

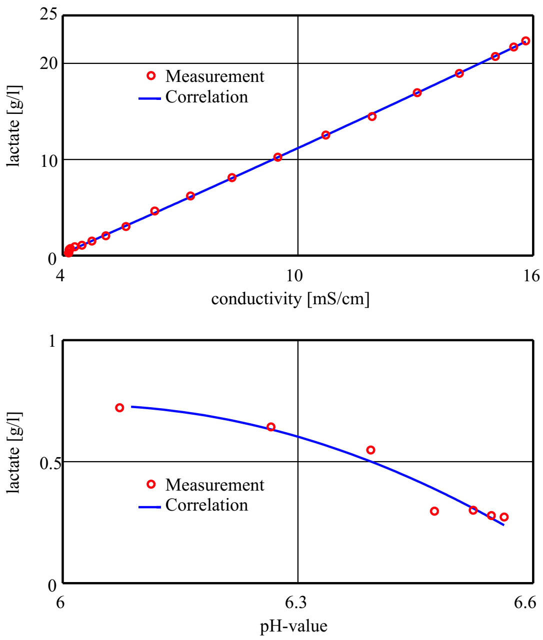

Fig. 2. Corelation between the concentration of lactate in the medium and the conductivity value (upper picture) and the pH-value (lower picture), respectively.

Unfortunately, the other interesting process values, lactose and biomass, cannot be correlated to online available measurement data. To overcome this deficiency, the application of a software-sensor is proposed in several publications in order to gain information about the formation of biomass and the consumption of substrate (Buchen et al. (2005); Hörrmann et al. (2007)). A schematic representation of this method is shown in figure 3.

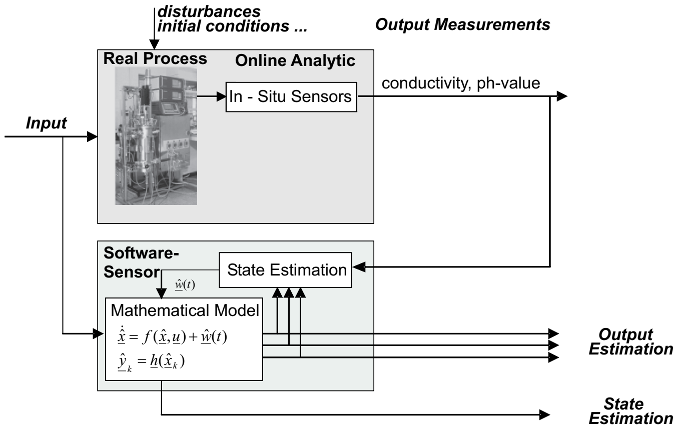

Fig. 3. Schematic representation of a software-sensor

The output measurements are used to synchronize the software-sensor with the real process, so that possible disturbances and uncertainties in the process can be included in the estimate.
Many different methods for the realization of a software-sensor are proposed (Jenzsch et al. (2006); Bastin and Dochain (1990); Haag (2004); Buchen (2007)). A very common method is the implementation of a previously developed mathematical model of the process, e.g. in an Extended Kalman Filter (Gelb et al. (1992)). In the herein considered application, however, a Neural Network is operating in parallel to the process, in early process stages supported by the use of a model-based full-horizon observer. The online available measurement data, the concentration of the pH-value, the conductivity of the medium and the amount of base added to the medium, are used for the synchronization between the real process and the process model. Using this data, the desired process values for biomass, lactose and lactate are estimated.

3. ARTIFICIAL NEURAL NETWORK AS SOFTWARE-SENSOR

The relationship between different process variables in biological processes is mostly nonlinear. One possibility to express these nonlinear correlations is the use of artificial neural networks (ANN), whose areas of application are increasing during the last decades (Baughman and Liu (1995)).

Herein the ANN is used as a software-sensor in order to estimate the amount of biomass and lactose online during the fermentation process. Therefore a direct mapping of online-measurable values onto biomass and lactose concentration is suggested, as illustrated in figure 4.

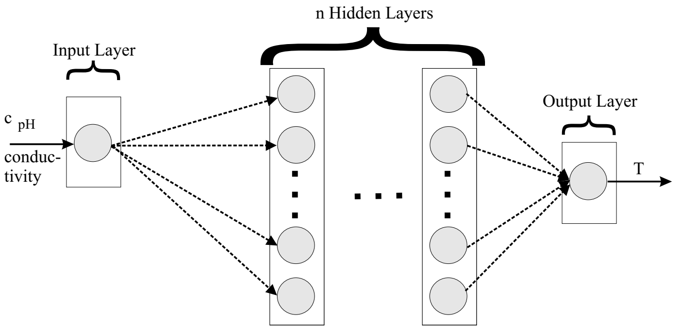

Fig. 4. Strucutre of an ANN

For process specific reasons two different ANNs are used for each output value. As long as the pH-value decreases ($t = 0$ until $t \approx 1.25[h]$), the pH-measurement serves as input for the first ANN. When the pre-specified value of pH = 6.05 is reached, base is fed to the medium in order to keep the pH-value constant. This action leads to an increase of the medium conductivity, which serves as input for the second ANN, which then calculates the state estimate.

In order to get reliable results for the biomass and the substrate, the output values have to be normalized

$$T_1(0) = 1 \tag{1}$$

$$\hat{X}(t) = X_0 \cdot T_1(t), \quad t < t_k \tag{2}$$

$$T_2(t_k) = 1 \tag{3}$$

$$\hat{X}(t) = X_0 \cdot T_1(t_k) \cdot T_2(t), \quad t \geq t_k, \tag{4}$$

where X_0 is the initial estimate of the biomass at $t = 0$, while T_1 and T_2 are the normalized estimated values of the different ANNs using pH-value and conductivity as input, respectively. The value t_k denotes the point in time, when the second ANN, using the conductivity as input value, is launched into operation with the process. The normalization of the values obtained by the ANNs for the estimation of the substrate lactose is identical to the one mentioned above. Due to the suggested normalization, the amount of hidden layers and neurons of the net and therefore its training epochs decrease.

Since the ANN represents the correlation between input and output values, the training has to be performed applying the mentioned normalization. Therefore the batch training using data sets of prior fermentations is disposed, using classical training algorithms like the backpropagation method.

4. THE OPTIMIZATION-BASED STATE ESTIMATOR

The results of the estimation process are very sensitive to the initial conditions. While the concentration of the substrate lactose and the product lactate can be determined a priori by analyzing the medium before the bacteria is inoculated, the amount of bacteria at the beginning of the process can hardly be considered as being constant. The variations reach from 30% in labscale up to more than 100% in industrial applications. To overcome these uncertainties, during the first 50 minutes of the fermentation process a full-horizon observer is used to determine the most likely initial amount of bacteria in the fermenter.

4.1 The Full-Horizon Observer

A general formulation of this estimation method is given in (Robertson et al. (1996)).
Using a given initial value $\underline{x}_0$ and available measurement data $\underline{y}_1, \cdots, \underline{y}_k$, the error in the initial condition $\underline{w}_0$ and the error sequence $\{\underline{w}_1, \cdots, \underline{w}_{k-1}\}$ can be determined by minimizing a cost function J at each time step k,

$$\min \quad J(\underline{w}_0, \cdots, \underline{w}_{k-1}) = \underline{w}_0^T \mathbf{P}^{-1} \underline{w}_0 + \sum_{l=1}^{k} \underline{v}_l^T \mathbf{R}^{-1} \underline{v}_l + \sum_{l=1}^{k-1} \underline{w}_l^T \mathbf{Q}^{-1} \underline{w}_l \quad , \tag{5}$$

with

$$\hat{\underline{x}}_0 = \underline{x}_0 + \underline{w}_0 \tag{6}$$

$$\underline{x}_l = F(\underline{x}_{l-1}) + \underline{w}_{l-1} \tag{7}$$

$$\underline{v}_l = \underline{y}_l - h(\underline{x}_l) \quad . \tag{8}$$

The first term in equation (5) pays attention to the deviation in the initial conditions, the second and third term to modelling and measurement errors, respectively. The matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$, describe the confidence into the corresponding terms and are chosen with respect to the a priori knowledge about the different errors. The nonlinear function F in equation (7) represents the mathematical modell of the process while the vector $\underline{v}_l$ in equation (8) is the difference between calculated process value and the corresponding measurement.
The solution of this optimization can be realized by a combination of different methods of numerical integration and nonlinear optimization.

4.2 The Growth Model of the Bacteria Streptococcus thermophilus

In order to determine the unknown process values via an optimization based observer, a sufficiently exact model of the fermentation process has to be developed. A brief description of the unstructured model is introduced in the following. Hereby the intra-cellular metabolism of the bacteria is considered as a black box and merely the dynamics of the concentration of the substrate lactose S, the product lactate P and the amount of biomass X are taken into account.
The popular kinetic approach of Monod (Monod (1942)) is used to develop a mathematical expression for the growth dynamic, paying attention to the phenomena of product inhibition and substrate limitation. One then obtains a set of coupled differential equations,

$$\dot{X} = \mu \cdot X \tag{9}$$

$$\dot{S} = Y_S \cdot \mu \cdot X \tag{10}$$

$$\dot{P} = Y_P \cdot \mu \cdot X \tag{11}$$

where Y_S and Y_P are the yield coefficients for the substrate and the product, respectively, while μ is the specific growthrate, calculated by

$$\mu = \mu_{max} \cdot \frac{S}{K_S + S} \cdot \frac{K_P}{K_P + P} \tag{12}$$

with μ_{max} representing the maximum growth rate and K_S and K_P being the limitation and inhibition constants, respectively. Figure 5 compares simulated values which are obtained by the mathematical growth model mentioned above and experimental results of three different fermentations with each other.

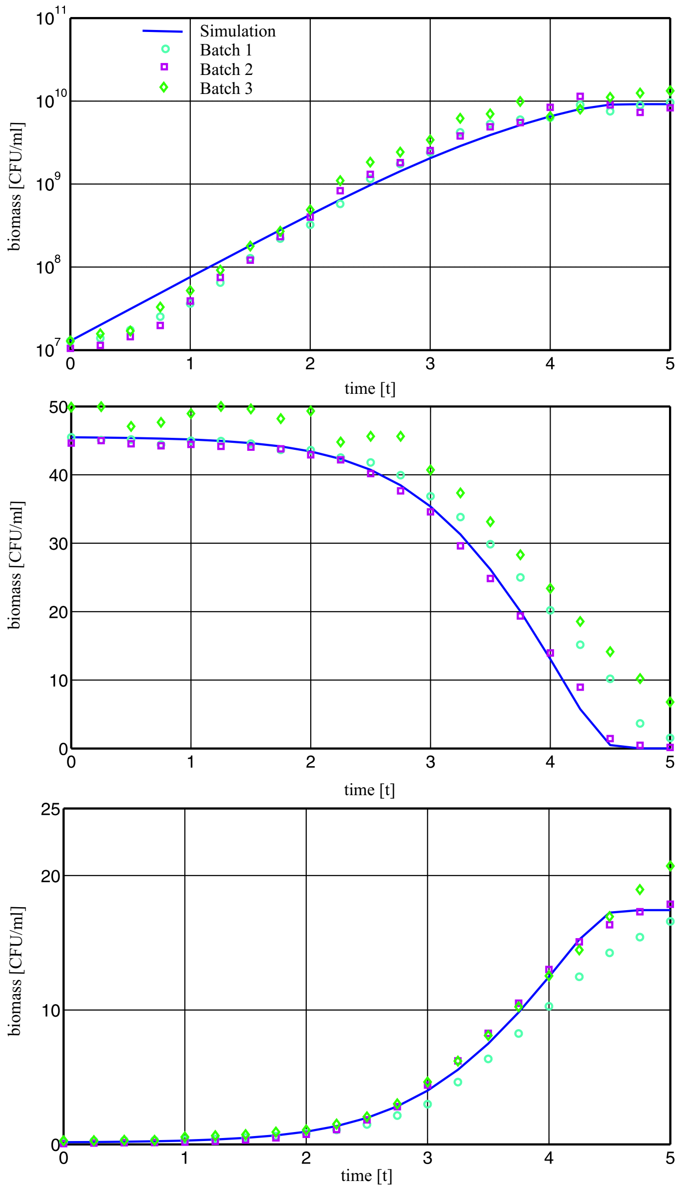

Fig. 5. Experminental results of three different fermentations compared with simulated values based on a mathematical growth model.

The five unknown parameters Y_S, Y_P, K_S, K_P and μ_{max} which determine the characteristics of the growth model, are obtained by least squares estimation and presented in the following table.

Parameters for the growth model of the bacteria *Streptococcus thermophilus*	
$Y_S[g/CFU]$	$4.947 \cdot 10^{-9}$
$Y_P[g/CFU]$	$1.887 \cdot 10^{-9}$
$K_S[g/l]$	1.7411
$K_P[g/l]$	14.989
$\mu_{max}[1/h]$	1.8584

Obviously, this very simple mathematical model gives already a sufficiently good representation of the fermentation process and therefore can be used for the estimation of process variables using a model-based observer.

5. RESULTS AND CONCLUSIONS

In order to test the performance of the proposed software-sensor it was applied to a common fermentation process, where the biomass, lactose and lactate were to be determined online.

5.1 Conditions of the Fermentation

The fermentations are performed in a 2 liter batchreactor, using a complex medium, with lactose serving as major substrate. The pH-value of the medium is adjusted to 6.5 prior to inoculation. Due to the growth of the bacteria, the acid lactate is produced, which leads to a decrease of the pH-value. When the marginal value of 6.05 is reached, a controller steadies the value by adding base to the medium. These process specifications are defined by industrial producers, who rely on long fermentation experience.
The values of medium components and biomass are determined offline. Every 15 minutes a sample is taken from the process and analyzed manually. An HPLC serves to determine lactose and lactate, while the results of the biomass are obtained via cell count.

5.2 Estimation Results

At the beginning of the fermentation, the initial conditions of the 3 process variables have to be determined. The amount of lactose and lactate in the medium are analyzed a priori to the inoculation using an HPLC, hence these values are known. The initial condition for the biomass is assumed to be $X_0 = 9 \cdot 10^6$, a value which is determined in former experiments. Once again, it shall be pointed out, that there cannot be an exact prediction of the initial value, due to unreproducable growth of the bacteria in preculture and errors in the inocculation amount. Although a sample is taken immediately after the pre-culture is inoculated into the fermenter, it takes at least two days until reliable results of the biomass at the beginning of the process, $X_{0,real}$, are available. For this reason the full-horizon observer is used to estimate an improved value of the initial biomass concentration X_0.
Every ten minutes a new prediction is calculated and so a new initial value of biomass is obtained. The development of the improved values is shown in figure 6.

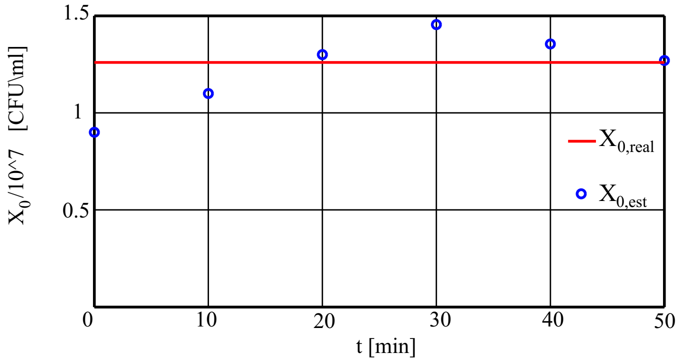

Fig. 6. Development of the estimated initial condition of biomass $X_{0,est}$ over process time t.

The estimated initial value $X_{0,est}$ evolves over process time and finally approaches the exact initial value $X_{0,real}$, which is determined manually by offline analysis.
These estimated results are immediately used for the ANN based estimation of biomass. The results of this state estimation is shown in figure 7.

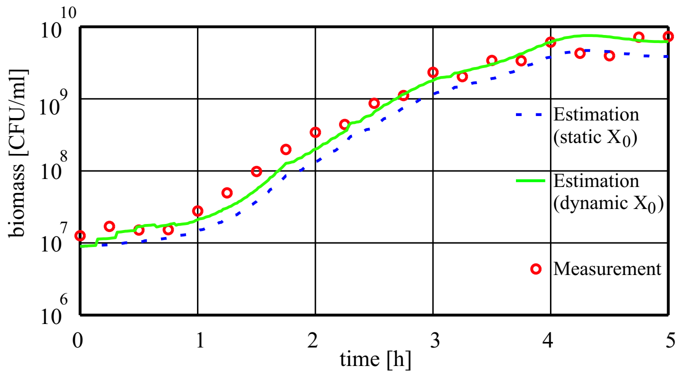

Fig. 7. Results of biomass estimation with static and dynamic initial condition for X_0.

The sensitivity of the estimation result to the initial condition X_0 is evident. In the case where the initial guess for X_0 is used over the whole process time, the estimated values do not agree with the measured ones. In fact the qualitative behavior of the process is well represented, merely the whole course of the estimation lies below the measurements.
In contrast, the estimation results of the ANN with dynamic initial conditions are very good. During the first hour of the process one clearly sees the steps in the estimation curve due to updated initial condition guesses and consequently the influence of the updated initial condition $X_{0,est}$ on the output value of the software-sensor. The continuously attained results of the ANN using dynamic initial conditions correspond well to the real process values and are superior to the results obtained with the static $X_{0,est}$.

In addition to the estimation of biomass, also the concentrations of lactose and lactate have been determined. The results are shown in figure 8.

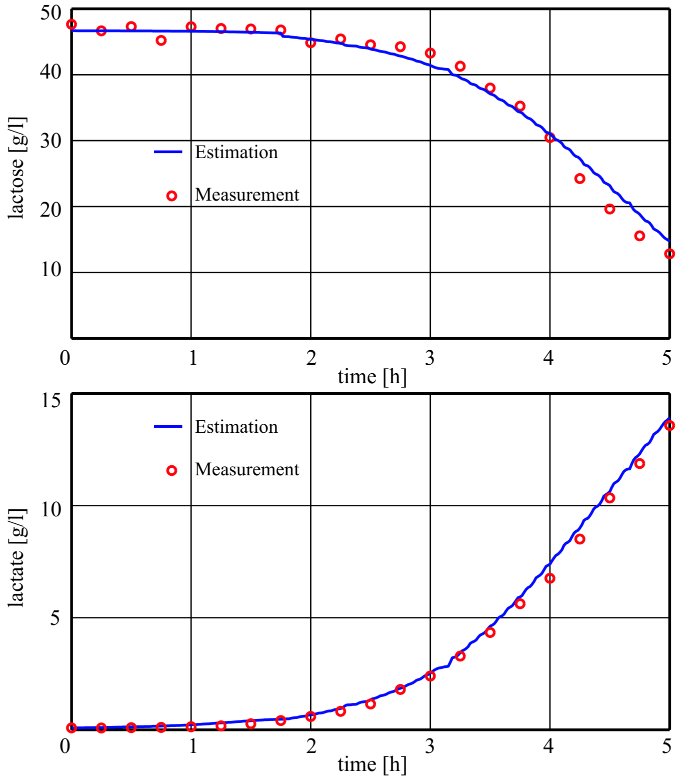

Fig. 8. Development of the estimated lactose and lactate concetrations over process time.

Lactose concentration was also estimated with an ANN. The initial condition has not been optimized dynamically, due to the a priori determination. The estimated values also correspond well to the measured ones.
The lactate values were calculated using the correlations between the pH-value and the conductivity of the medium, respectively, as already shown in section 2. The obtained results also agree well with the reality.

5.3 Conclusion

In this contribution we proposed a software-sensor for the estimation of biomass using ANNs. We showed the sensitivity of the estimation to the initial condition, which cannot be predicted reliably. To overcome this obstacle we combined the optimization based full-horizon observer with the ANN. With this modell-based method the initial condition of biomass is estimated during early process stages and so leads to a more credible value than a guess based on experience. Using this new, better value, the estimation results show a very good correspondence to reality. This method is therefore suitable for the online-measurement of biomass in biotechnological fermentation processes.
Based on the herein achieved results, the fermentation can now be optimized by applying modern control theory to the process, which will be based upon the online estimation of process values, in order to increase the output and quality of the product.

REFERENCES

G. Bastin and D. Dochain. *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier, Amsterdam, 1990.

D. R. Baughman and Y. A. Liu. *Neural Networks in Bioprocessing and Chemical Engineering*. Academic Press Inc., San Diego, 1995.

A. Buchen. *Modellierung des Wachstums und der Einschlusskörperbildung des Bakteriums Photorhabdus luminescens*. PhD thesis, Christian-Albrechts-Universität, Kiel, 2007.

A. Buchen, K. Kolahi, and H. Röck. Software-sensor based on moving horizon state estimation applied in biotechnology. *Third Int. Conf. on Systems, Signals and Devices*, 2005. Sousse, Tunisia.

A. Gelb, J. F. Kasper, and A. A. Sutherland. *Applied Optimal Estimation*. The MIT Press, 1992.

O. Gronau. Produktion von *Streptococcus thermophilus* im Diafiltrationsprozess . Master's thesis, University of Applied Sciences, Hamburg, Germany, 2005.

J. Haag. *Dynamic Modelling and State Estimation of Complex Bioprocesses*. PhD thesis, Faculte Polytechnique de Mons, 2004.

P. Hols, M. Kleerebezem, and B. Grossiord. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiology Reviews*, 29:435–463, 2005.

J. Hörrmann, D. Barth, M. Kräling, and H. Röck. State estimation in bioprocesses - extended kalman filter vs. neural networks. 9^{th} *IASTED International Conference Control and Application*, 42(8):238–242, 2007.

M. Jenzsch, R. Simutis, and A. Lübbert. Estimation of biomass concentrations in fermentation processes for recombinant protein production. *Journal of Bioprocess and Biosystems Engineering*, 29:19–27, 2006.

J. Monod. *Recherche sur la Croissance des Cultures Bacteriennes*. Herman, Paris, 1942.

D. G. Robertson, J. H. Lee, and J. B. Rawlings. A moving horizon-based approach for least-squares estimation. *AiChE Journal*, 42(8):2209–2224, 1996.