BIOMASS GROWTH AND k_La ESTIMATION USING ONLINE AND OFFLINE MEASUREMENTS

Z. I. T. A. Soons[a,b], J. Shi[a], L. A. van der Pol[b], G. van Straten[a] and A. J. B. van Boxtel[a]

[a]Wageningen University, Systems and Control Group
P.O. box 17, 6700 AA Wageningen, The Netherlands, Zita.Soons@wur.nl
[b]Netherlands Vaccine Institute
Antonie van Leeuwenhoeklaan 9-11, P.O. box 457, 3720 AL Bilthoven, The Netherlands

Abstract: Measurement of the key process variables is essential during biopharmaceutical production. These measurements are often not available online. This work combines frequent online measurements with infrequent offline measurements to estimate the specific growth rate, biomass, and the oxygen mass transfer coefficient during continuous and fed-batch cultivations of *Bordetella pertussis* online using an Extended Kalman filter, parameter adaptation, and learning.

Keywords: Extended Kalman filter, online measurements, delayed offline measurements, oxygen transfer, specific growth rate, biomass, bioreactor, vaccine production, *Bordetella pertussis*.

1. INTRODUCTION

Most biopharmaceuticals are produced in a batch or fed-batch cultivation. The quality of the product is formed in this step and is the result of the metabolic state of the micro-organisms. It is therefore essential to measure the physiological state of the process. Metabolic activity is difficult to measure directly due to the lack of sensors, but respiration can be monitored by the oxygen mass balance. The oxygen uptake rate can in turn be used to estimate the specific growth rate and biomass. The specific growth rate and biomass concentration are key parameters that define the metabolic state of micro-organisms.

In this application only dissolved oxygen measurements were available. Therefore, a software sensor based on an Extended Kalman filter (EKF) was developed to estimate specific growth rate and biomass every minute using the oxygen uptake rate (*OUR*) as input. The choice for an EKF is in line with the process, in which the instruments and the measurement noise are known. The performance of the EKF was compared to the asymptotic observer of previous work (Neeleman *et al.*, 2002), and was significantly better.

In bioreactors, aerated by a high air flow entering the headspace, the difference between the inlet and exhaust oxygen fraction is small and can therefore not be measured accurately. Hence *OUR* must be calculated using the oxygen balance in the liquid phase (Neeleman, 2002, Soons *et al*., 2006, 2006a):

$$OUR \approx k_L a \cdot (C_O^{in} - C_O^L) \qquad (1)$$

Where k_La is the oxygen transfer coefficient. The dissolved oxygen concentration (C_O^L) is assumed to be at pseudo-steady state i.e. accumulation of oxygen in the bioreactor is negligible. The oxygen concentration entering (C_O^{in}) the bioreactor is assumed to be equal to the oxygen concentration in the headspace.

In most applications, the oxygen transfer rate k_La is measured in advance of the cultivation in medium and is assumed to depend on stirrer speed and volume only. However, the formation of cells, proteins, and other molecules, which absorb at gas-liquid interfaces, cause interfacial blanketing and reduce the oxygen transfer rate (Doran, 1995). Because concentrations of cells, substrates, and products change during (fed-) batch cultivation, the value of k_La also changes. An example of change in k_La due to these factors is given in Sabra *et al* (2002). Changing k_La causes errors in the *OUR* calculation and the estimation of the specific growth rate and biomass. It is therefore essential to deal with time-varying k_La.

Offline measurements are mostly considered as not suited for control and estimation purposes, because they become available with a delay and at infrequent and irregular times. These measurements however contain valuable information about the states of the system and can make the estimator more robust (Dondo and Marqués, 2003). Amongst the literature on bioprocess monitoring (e.g. Bastin and Dochain, 1990) the use of offline information for online estimation is relatively small. Myers *et al.* (1995) and Tatiraju *et al.* (1997) use offline and online measurements for state estimation, but do not estimate parameters; Lubenova *et al*. (2003); Gudi *et al.* (1995); Dondo and Marqués (2003); and Ignatova *et al.* (2003) estimate parameters in addition. In these approaches the parameters are part of the input-output equations. In this work, however, better results are obtained if the parameter k_La is considered as a part of the *OUR* calculations, which is done separately from the input-out put equations (see figure 1).

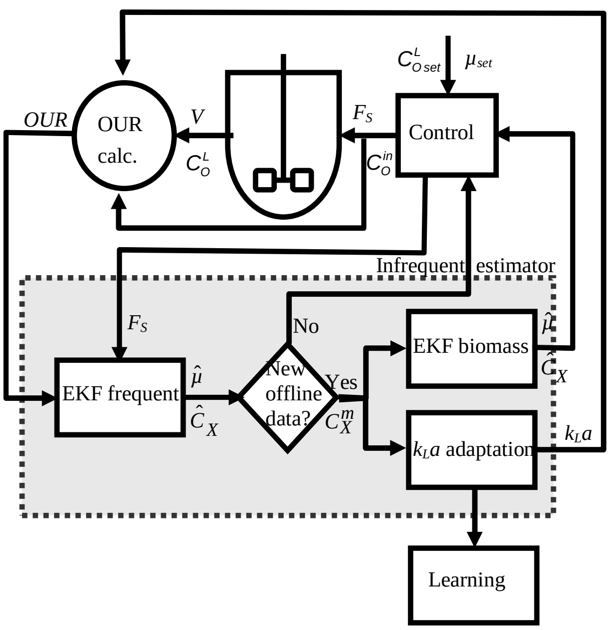

Fig. 1. System configuration (symbolic: see text or nomenclature).

This work combines frequent online measurements (oxygen uptake rate) with infrequent offline measurements (biomass) to estimate the specific growth rate and biomass accurately. Figure 1 shows an overview of the system, in which two types of estimators an learning are involved: a *frequent* estimator using the online data and an *infrequent* estimator activated by sampled offline data. The offline measurements are also used to adapt the time-varying k_La. After each run improved values for k_La are obtained and can be used in a learning process to acquire the appropriate k_La time pattern.

2. EXTENDED KALMAN FILTER

The estimator is based on a nonlinear continuous-time model:

$$\begin{aligned} \frac{dx}{dt} &= f(x,u) \\ y &= h(x) \end{aligned} \tag{2}$$

With f a nonlinear function of the states x and inputs u, and y the output.

Lewis (1986) gives a good explanation of an Extended Kalman filter. The application in biotechnological applications is amongst others discussed by Gudi *et al.* (1995), Neeleman (2002) and Keesman (2002). The structure of a discrete time EKF is shown in Fig. 2, and is based on the following equations following Lewis (1986):

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + w_k \\ y_k &= C_k x_k + v_k \end{aligned} \tag{3}$$

Where A_k and C_k and B_k at each time instant follow from discretization and linearization of Eq. 2 for a time step τ:

$$\begin{aligned} F &= \left[\frac{\partial f}{\partial x}\right]_{\hat{x},u}, \quad C_k = \left[\frac{\partial h}{\partial x}\right]_{\hat{x},u} \\ A_k &= I + F\tau \quad , B_k = \left[\frac{\partial f}{\partial u}\right]_{\hat{x},u} \tau \end{aligned} \tag{4}$$

and u_k is the input vector. The initial states x_0 are stochastic variables with average $\bar{x}_0$ and variance P_0: $x_0 \sim (\bar{x}_0, P_0)$; $w_k \sim (0, Q_k)$ is system noise and consists of model errors an unknown inputs; and $v_k \sim (0, R_k)$ is measurement noise. The algorithm has two steps. The time update and the measurement update.

2.1 *Time update*

When a sample comes available at time k, first the time update $k+1$ is calculated using the original nonlinear model.

$$\bar{x}_{k+1} = f(\hat{x}_k, u_k)$$
$$\bar{y}_{k+1} = h(\bar{x}_{k+1}) \tag{5}$$

giving the predicted states $\bar{x}_{k+1}$ and the output $\bar{y}_{k+1}$. The prediction of the variance of the states $\bar{P}_{k+1}$ is based on the system A_k and system noise (Q_k).

$$\bar{P}_{k+1} = A_k \hat{P}_k A_k^T + Q_k \tag{6}$$

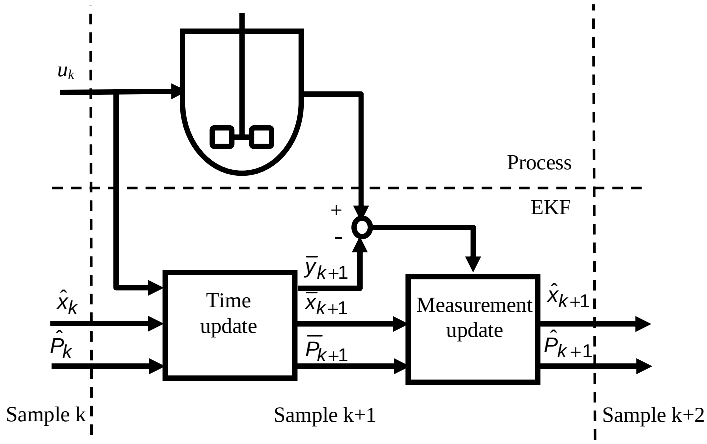

Fig. 2. Calculation scheme for the Extended Kalman filter.

2.2 *Measurement update*

Next the measurement update calculates new estimates using the predicted model states $\bar{x}_{k+1}$ and outputs $\bar{y}_{k+1}$, the actual measurements y_{k+1}, and the predicted variance $\bar{P}_{k+1}$:

$$\hat{P}_{k+1} = \bar{P}_{k+1} - \bar{P}_{k+1} C_{k+1}^T R_{k+1} C_{k+1} \bar{P}_{k+1}$$
$$K_k = \hat{P}_{k+1} C_{k+1}^T R_{k+1}^{-1} \tag{7}$$
$$\hat{x}_{k+1} = \bar{x}_{k+1} + K_k \left(y_{k+1} - \bar{y}_{k+1} \right)$$

Where the Kalman gain K_k is the magnitude of the correction and aims to minimize the covariance $\hat{P}_k$.

3. FREQUENT ESTIMATOR

3.1 *Model*

The frequent estimator is based on a generic model (Eqs. 8 and 9) with μ as an additional state. The model contains only two parameters: the yield (Y_O) and maintenance (m_O) coefficient on oxygen (which were calculated using separate experiments):

$$\mu_{k+1} = \mu_k$$
$$C_{X,k+1} = \left(1 + \tau(\mu_k - \frac{F_{S,k}}{V_k}) \right) C_{X,k} \tag{8}$$

$$OUR_k = \left(\frac{\mu_k}{Y_O} + m_O \right) C_{X,k} \tag{9}$$

The states x are specific growth rate and biomass (μ and C_X), the output y is the oxygen uptake rate (*OUR*). The volume V is assumed to be exactly known and doesn't need to be estimated. F_S is the feed rate. The prediction of the states follows every instant (one minute) by calculation of Eq. 8, the measurement update from calculation of Eq. 7. The observability matrix is full rank, which renders the linearized system observable if the oxygen uptake rate is not negligible.

3.2 *Tuning*

The parameterization of the EKF is based on simulation. Synthetic output noise is derived from real process data. System noise (Q)

$$Q = \begin{bmatrix} Q_\mu & 0 \\ 0 & Q_{C_X} \end{bmatrix} \tag{10}$$

was chosen in such a way that the deviation (E) between estimation (EKF) and generated data by a detailed Monod model (Soons *et al*., 2006) is minimal for a worst case scenario with a high level of noise.

$$E = \sqrt{\frac{\sum_{k=1}^{N} \left[\hat{\mu}(k) - \mu_{\text{mod}}(k)\right]^2}{\bar{\mu}_{\text{mod}} \cdot (N-1)}} + \sqrt{\frac{\sum_{k=1}^{N} \left[\hat{C}_X(k) - C_{X\,\text{mod}}(k)\right]^2}{\bar{C}_{X\,\text{mod}} \cdot (N-1)}} \tag{11}$$

With Q_μ the intensity of the system noise on the specific growth rate and Q_{Cx} on the biomass. Both terms are divided by their average to make them equally important.

4. INFREQUENT ESTIMATOR

The infrequent estimator updates the biomass and k_La estimation using infrequent offline and frequent online measurements (Fig. 1). Since inaccurate values of k_La may cause biased estimations for the frequent estimator, k_La is also updated when offline samples are available.

Using a second Extended Kalman filter to estimate C_X and k_La for offline samples would be the logical step to incorporate the effect of product and biomass formation on k_La. The infrequent estimator is extended with the ten delayed biomass concentrations (analyzing biomass samples takes approximately ten sample instants, Eq. 12), k_La, and dissolved oxygen. Although this estimator is observable, the algorithm is complex and tuning is difficult.

As an alternative, the following three-step algorithm is designed. First the frequent estimator estimates μ and C_X using frequent online available data (section 3). Next, if samples are available, the biomass estimator estimates biomass and finally the adaptive estimator estimates k_La. The main advantage of this approach is that convergence of k_La-adaptation is determined by one tuning parameter which can be enforced *a priori*.

4.1 EKF Biomass

The biomass estimator contains the following discrete model, in which the sample and analysis delay is assumed to be constant and is incorporated by addition of fictitious delayed states Gudi *et al.* (1995):

$$
\begin{aligned}
&C_{X_1,k+1} = \left(1+\tau(\mu_k - \frac{F_{S,k}}{V_k})\right)C_{X_1,k}\\
&C_{X_2,k+1} = C_{X_1,k}\\
&.\\
&.\\
&C_{X_{11},k+1} = C_{X_{10},k}
\end{aligned}
\tag{12}
$$

where τ, as before, is the frequent sampling instant. The states of the biomass estimator are the (delayed) biomass concentrations ($C_{X_1}, C_{X_2},...C_{X_{11}}$); the output is biomass delayed with ten minutes ($C_{X_{11}}$). The prediction of the states follows every infrequent instant (if an offline sample is available, Fig. 1) from calculation of Eq. 12, the measurement update from calculation of Eq. 7.

The observability matrix has full rank. Tuning is performed similar to the frequent estimator by optimization of the covariances of Q.

4.2 Adaptive k_La Estimator

During cultivation the microorganisms and their products may affect the physical properties of the liquid and as a result k_La may change. Deviations between the measured and actual biomass are assumed to be caused by errors in k_La. The adaptive estimator adapts k_La when samples are available. The relative correction c is calculated every infrequent instant and assumed constant until the next sample becomes available:

$$
\begin{aligned}
&\textit{if } C^m_{X,k} \textit{ exists}\\
&\quad k_La^{Stir}_k = c\cdot f(V,R^{Stir})\\
&\quad k_L\hat{a}_k = k_La^{Stir}_k + \gamma(C^m_{X,k} - \hat{C}_{X,k})\cdot T_m\\
&\quad c = \frac{k_L\hat{a}_k}{k_La^{Stir}_k}\\
&\textit{else}\\
&\quad k_La^{Stir} = c\cdot f(V,R^{Stir})\\
&\quad k_L\hat{a}_k = k_La^{Stir}_k\\
&\quad c = c
\end{aligned}
\tag{13}
$$

Depending on the oxygen demand during the cultivation, agitation speed is adjusted. The oxygen transfer coefficient k_La is measured in advance of the cultivation in biomass-free medium and depends on the stirrer speed and on the volume as specified in the function $f(V,R^{Stir})$ in Eq. 13. This function is used to adjust k_La *a priori* for known changes stirrer speed.

The tuning parameter γ determines the convergence speed of the observer. T_m is the infrequent sampling instant, which can be irregular and infrequent.

4.3. Learning

In some situations (e.g. production) it is not desirable to take many offline samples, because the process may be disturbed or it is not possible to take samples. The estimators discussed so far are applied online. Learning can be applied offline after each run.

In order to improve the estimation when samples are not available and c in Eq. 13 cannot be calculated, the effect of product formation on k_La (based on previous runs) can used instead and reads:

$$c(C_X) = 1 + f(C_X) \tag{14}$$

where $f(C_X)$ determines the decrease of k_La as a function of biomass. Batch-to-batch improvement can be obtained by estimating $f(C_X)$ by regression after each run using data of all runs.

5. EXPERIMENTAL RESULTS

Fed-batch and continuous-flow stirred-tank (CSTR) cultivations with the dual substrate consuming bacterium *Bordetella pertussis* were performed with glutamate and L-lactate as the main carbon sources. Bioreactor conditions, medium, analysis, and software were applied as reported by Soons *et al.* (2006). First, batch cultivation was performed until the limiting substrates were depleted. Next, the fed-batch phase of the cultivation automatically started when the specific growth rate dropped below the set-point; the CSTR phase started when the biomass reached an optical density of one. The dynamic method was used to estimate k_La in medium before inoculation for a range of stirrer speeds and volumes. The CSTR experiments were performed to test the frequent EKF only, the fed-batch experiments to test the infrequent estimator (frequent EKF, infrequent EKF, and k_La adaptation).

Maintenance and yield coefficients were calculated on the basis of a series of four CSTR experiments with low biomass. Blanketing effects are therefore small so that k_La can be assumed independent of biomass and product formation, which allows calculation of proper maintenance and yield coefficients.

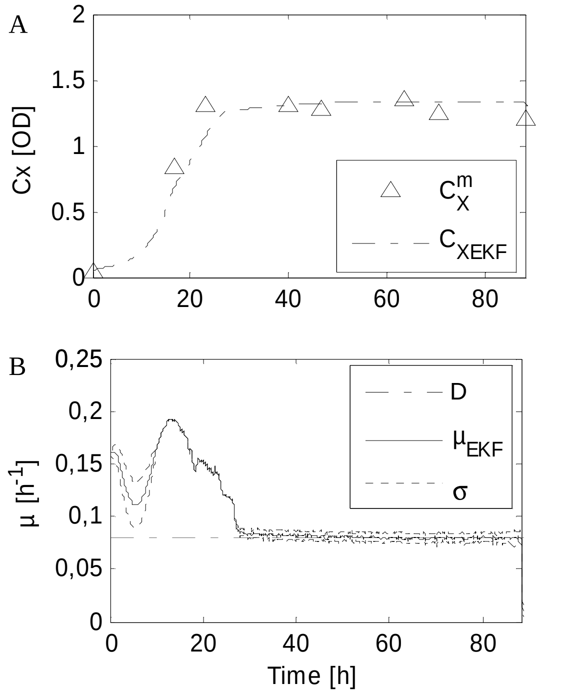

Fig. 3. Frequent estimator for CSTR cultivation A. Biomass. B. Estimated specific growth rate and its standard deviation (σ).

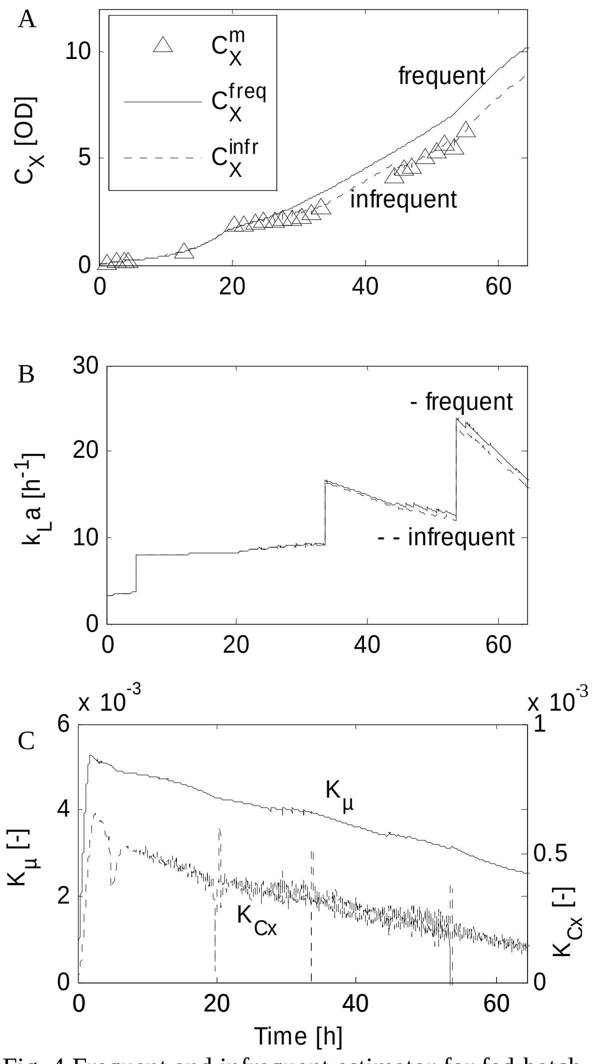

Fig. 4 Frequent and infrequent estimator for fed-batch cultivation. A. Biomass. B. Oxygen transfer coefficient. C. Kalman gains

Figure 3 shows that the estimated biomass coincides well with the offline biomass measurements during CSTR. The estimated specific growth rate converged exactly to the preset dilution rate ($D = F_S.V^{-1}$). Uncertainty on μ was small throughout the cultivation.

Blanketing effects however do occur during fed-batch cultivation with higher biomass concentrations, resulting in overestimation of the biomass if only the frequent estimator is used (Fig. 4). Figure 4 compares the frequent estimator with the infrequent estimator for fed-batch cultivation. An exponentially increasing feed was added to the bioreactor to keep the specific growth rate constant, resulting in an increase in volume and a decrease in k_La. sampling causes a small decrease in volume and therefore slight increase in k_La. The abrupt changes of k_La were caused by changes of stirrer speed (to meet the increasing oxygen demands). Biomass was overestimated using only the frequent estimator (Eqs. 8 and 9). Biomass was estimated more accurately when the infrequent estimator (Eqs. 8, 9, 12, and 13) corrected the k_La and C_X estimations for biomass and product formation.

The availability of more sparse biomass measurements will give larger sampling intervals T_m and larger k_La adaptation per offline measurement (Eq. 13). The effect of more sparse measurements on observer performance will therefore be small, however, the observer becomes more noise sensitive when only few noisy measurements are available. Obviously, some measurements are required to find the proper k_La pattern.

Figure 5 shows the values of the relative k_La ($c = k_L\hat{a} / k_L a^{Stir}$) for two fed-batch cultivations at different biomass concentrations. The relative k_La decreased by up to 8% for a cultivation grown to 8 *OD* due to biomass and product formation. Using a linear relation between biomass and relative k_La (Eqs. 14 and 15) improves the biomass estimation (d=0.01):

$$c(C_X) = 1 + d \cdot C_X \tag{15}$$

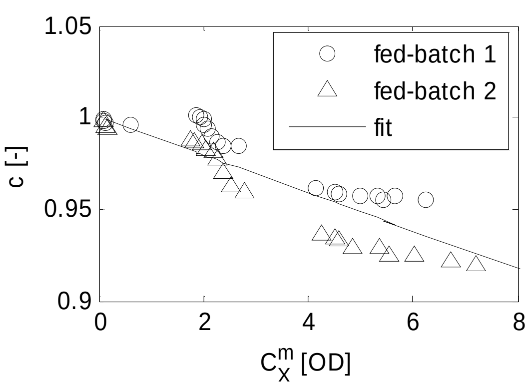

Fig. 5 Relative k_La (k_La during fed-batch compared to k_La in pure medium, $c = k_L\hat{a} / k_L a^{Stir}$) as function of the biomass concentration for two fed-batch experiments.

For high cell density cultivation (e.g. *E. coli*), in which cells can grow up to approximately 150 *OD*, k_La will drastically decrease. It illustrates the potential of incorporating this effect in the estimation.

6. CONCLUSIONS

Application of the presented observer design improves bioprocess monitoring in biopharmaceutical production. The estimator consists of three parts operating on three time-scales: a *frequent* estimator that estimates specific growth rate and biomass online; an *infrequent* estimator, activated by offline measurements, that estimates the biomass and k_La online; and *learning* that improves the relation between biomass and the oxygen transfer coefficient from batch-to-batch. The estimator showed convergence and fitted well to the data.

Biomass and products influence mass transfer in a bioreactor. Estimation of these k_La changes is essential for proper estimation of the states during fed-batch and other high cell density cultivations that use the oxygen balance in the liquid phase.

NOMENCLATURE

A, B, C	System matrices	
c	Relative k_La	
C_X	Biomass concentration	OD
C_O^{in}	Oxygen concentration into reactor	$mmol.l^{-1}$
C_O^{L}	Oxygen concentration in medium	$mmol.l^{-1}$
d	Constant	
E	Performance criterion	
F_S	Substrate feed rate	$l.h^{-1}$
k_La	Oxygen transport coefficient	h^{-1}
m_O	Maintenance on oxygen	$mmol.OD^{-1}.h^{-1}$
OTR	Oxygen transfer rate	$mol.l^{-1}.h^{-1}$
OUR	Oxygen uptake rate	$mol.l^{-1}.h^{-1}$
P	Variance of the states	
Q, R	System noise, output noise	
R^{Stir}	Stirrer speed	rpm
t	Cultivation time	h
T_m	Infrequent sampling interval	h
u, x, y	Inputs, states, outputs	
V	Liquid volume	l
v, w	Measurement noise, system noise	
Y_O	Biomass yield on oxygen	$OD.mmol^{-1}$
γ	Tuning parameter	
μ	Specific growth rate	h^{-1}
τ	Frequent sampling interval	h
$\hat{\ }$, $_{EKF}$	Estimated values	
$\bar{\ }$	Predicted values	

ACKNOWLEDGMENTS

Wageningen University and the Netherlands Vaccine Institute (NVI) work together in a project with Applikon Biotechnology BV and Siemens NV in order to improve the production process by release of biopharmaceuticals on the basis of new online techniques. Senter supports the project under project name TSGE3067. Laboratory experiments were performed at NVI, The Netherlands.

REFERENCES

Bastin, G. and D. Dochain (1990), O*n-line estimation and adaptive control of bioreactors*, Elsevier, Amsterdam.

Dondo, R., D. Marqués (2003), Simulation results for on-line optimization of a batch bioreactor using nonlinear filtering and optimal control, lSA Transactions **42**, 289-303.

Doran, P. M. (1995), *Bioprocess engineering principles*, Academic Press, San Diego.

Gudi, R. D., S. L. Shah, M. R. Gray, P. K. Y. Aran (1997), Adaptive multirate estimation and control of nutrient levels in a fed-batch fermentation using off-line and on-line measurements, *The Canadian journal of chemical engineering* **75**, 562-573 .

Keesman, K. J. (2002), State and parameter estimation in biotechnical batch reactors, *Control Engineering Practice* **10**, 219-225

Lewis, F.J. (1986), *Optimal estimation,* John Wiley and Sons, New York.

Ignatova, M. N. (2003), Adaptive biomass observer on the basis of measurements of oxygen, *the 11th Mediterranean conference on control and automations*, Rhodes, Greece.

Lubenova, V., I. Rocha, E. C. Ferreira (2003), Estimation of multiple biomass growth rates and biomass concentration in a class of bioprocesses, *Bioprocess Biosyst. Eng.* **25**, 395-406.

Myers, M.A., S. Kang, R. H. Luecke (1995), State estimation and control for systems with delayed off-line measurements, *Computers Chem Engineering* **20**, 585-588.

Neeleman, R. (2002), *Biomass performance – Monitoring and control in bio-pharmaceutical production*, Thesis, Wageningen University, The Netherlands.

Sabra W., E. Kim, A. Zeng (2002), Physiological responses of *Pseudomonas aeruginosa* PAO1 to oxidative stress in controlled microaerobic and aerobic cultures, *Microbiology* **148**, 3195-3202.

Soons, Z. I. T. A., J. A. Voogt, G. Van Straten, A. J. B. Van Boxtel (2006), Constant specific growth rate in fed-batch cultivation of *Bordetella pertussis* using adaptive control, *J. Biotechnology* **125**, 252-268.

Soons, Z. I. T. A., G., Van Straten, A. J. B. Boxtel (2006a), Automatic tuning and adaptation for specific growth rate control of fed-batch cultivation, *Chemical Process Control 7*, Lake Louise, Canada.

Tatiraju, S., M. Soroush, R. Mutharasan (1999), Multi-rate nonlinear state and parameters estimation in a bioreactor, *Biotechnology and Bioengineering* **63**, 22-32.