State and Specific Growth Estimation in Heterologous Protein Production by *Pichia pastoris*

José M. Barrigón, and Ramon Ramon  
Dept. of Chemical Engineering, Escola d’Enginyeria, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain  
Isabel Rocha  
IBB—Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, 4710-057 Braga, Portugal  
Francisco Valero  
Dept. of Chemical Engineering, Escola d’Enginyeria, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain  
Eugénio C. Ferreira  
IBB—Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, 4710-057 Braga, Portugal  
José L. Montesinos  
Dept. of Chemical Engineering, Escola d’Enginyeria, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain  
DOI 10.1002/aic.12810  
Published online November 22, 2011 in Wiley Online Library (wileyonlinelibrary.com).

Estimation of biomass, substrate, and specific growth rate (μ) by two nonlinear observers (nonlinear observer-based estimator—NLOBE, asymptotic observer with second-order dynamics tuning—AO-SODE) and a linear estimator (recursive least squares with variable forgetting factor—RLS-VFF) is presented. Heterologous protein production in *Pichia pastoris* PAOX1 (Mut+) and PFLD1-based systems is closely related to μ and has been addressed due to its high relevance in modern biotechnology and bioprocess engineering. μ was estimated by online gas analyses or substrate measurements, biomass, and substrate considering yield coefficients and mass balances. In simulation studies, NLOBE showed high sensitivity to tuning and initialization variables. Validation experiments demonstrated AO-SODE performs better than the RLS-VFF for moderate to rapid changes of μ and model parameters being known. If low changes on μ are presented, for instance, in substrate regulation, RLS-VFF comes up as the best option, because of its reduced requirements. © 2011 American Institute of Chemical Engineers AIChE J, 58: 2966–2979, 2012  
Keywords: state estimation, observers, recursive least squares, heterologous protein production, *Pichia pastoris*

Introduction

To achieve high productivity, constant product quality, and also to allow optimization and control of biotechnological processes, real-time monitoring of the key fermentation variables (biomass, substrate, and products) is of major academic and industrial relevance. Some reviews and specific papers have addressed this topic, covering the wide range of techniques that have been applied. 1–4 Biomass is usually the central variable of the mathematical models used to describe microbial growth, where it is included as a state variable. Several analytical methods have been adapted to monitor cell density evolution during bioprocesses. However, a standard method for the online determination of biomass is not currently available, as each available technique has its own advantages and disadvantages. 5,6

The methylotrophic yeast *Pichia pastoris* has been widely reported as a suitable expression system for both basic research and industrial application. 7 In recent years, more than 500 proteins have been expressed using this system. 8 The strong and tightly regulated alcohol oxidase 1 promoter (PAOX1) is the most widely used *P. pastoris* promoter for recombinant expression, being induced by methanol. 9,10 To avoid methanol use, FLD1 promoter has been recently considered. 11 The gene encodes a formaldehyde dehydrogenase (FLD), a key enzyme required for the catabolism of methanol and certain primary amines, such as methylamine used as nitrogen source in methylotrophic yeast. 12 With this promoter, methanol is replaced by sorbitol as carbon source and methylamine is used as sole nitrogen source and inducer of protein production.

Correspondence concerning this article should be addressed to J. L. Montesinos at joseluis.montesinos@uab.es.
The specific growth rate is a critical parameter in the optimization of the heterologous protein production by *P. pastoris*. There is an optimum growth rate for optimal product formation, which is protein specific. Nevertheless, besides being difficult to estimate online, this parameter may vary during the fermentation process, since this is usually operated in batch and fed-batch modes. Moreover, there are variations in substrates and operational conditions applied within the different phases of the production process.\(^{13}\)

Online estimation of the specific growth rate(s) is usually performed together with the determination of state variables such as biomass, as the direct estimation of the specific growth rate is not possible. The determination of these variables can be achieved by the development of estimation algorithms or software sensors that require the existence of accurate mathematical models of the process.\(^{14-19}\) Nonetheless, the development of a suitable mathematical description and its identification is difficult due to the complex interactions exhibited by the microorganism, as well as to the operating conditions and the state of the system.

To eliminate errors in estimation owing to inaccurate models, it is possible to design adaptive algorithms that estimate simultaneously the state and some parameters of the process, considered as time-varying parameters.

This type of estimation can be achieved by an adaptive extended Kalman filter,\(^{20,21}\) a recursive prediction error method,\(^{22}\) an adaptive extended Luenberger observer,\(^{23}\) and also by nonlinear observers,\(^{24}\) which are based on the accurate knowledge of both model structure and parameters. However, these observers present a problem of convergence over a wide range of operational conditions. In addition, guessing suitable initial values for different state variables and parameters is rather critical to obtain precise estimates.

Adaptive observers somewhat easier to implement have been designed.\(^{25}\) They estimate the state and the kinetics of the process, considered as time-varying parameters, through partial measurements of the state.

Observer-based estimators (OBE), whose fundamentals are based on the nonlinear systems theory, have also been proposed. In this approach, estimation of reaction rates is carried out from the measurement of state variables.\(^{26,27}\) The observer gain depends on measured state variables and also constraints, which are added for the calibration of the gain when the measurements used are relatively noisy. Furthermore, specific nonlinear observer-based estimators (NLOBEs) have been developed.\(^{27}\) The main characteristic of those estimators is the ease with which they may be implemented and specifically calibrated. Their tuning is reduced to the calibration of a simple tuning parameter.

Because of the complexity to tune OBE’s, some authors studied their stability, dynamics of convergence, and suitable values for tuning parameters.\(^{28}\) Recently, an asymptotic observer (AO) for the estimation of state variables has been designed and its performance compared with that of a classical observer (extended Kalman observer).\(^{29}\) The specific growth rates were obtained using an estimator based on the reformulation proposed by Bastin and Dochain.\(^{25,30}\) The AO allows reconstructing the missing state variables, even if the process is not exponentially observable and the kinetics is unknown. Another advantage of this kind of observers is that there are no tuning parameters. However, it is assumed that the yield coefficients are known.

In addition, estimation of parameters by the recursive least squares (RLS) method has been proposed, especially for the specific growth rate and for the process state, thus, achieving unbiased convergence and making possible the estimation of changing process parameters.\(^{31,32}\) The main advantage of the RLS methods is that, it does not require an accurate knowledge of the system because a linear model that only depends on the online measures is considered.

The modulating-function method has also been used for the online identification of a microbial growth model.\(^{33}\) Nevertheless, due to large number of estimated coefficients and parameters, the method is of difficult application to experimental conditions.

In this article, the estimation of biomass, substrate, and specific growth rate are presented and applied to the heterologous production of *Rhizopus oryzae* lipase (ROL) by *P. pastoris* in batch and fed-batch operational modes. Different algorithms and procedures are studied and discussed, always considering their applicability in terms of overall performance, taking into account aspects, which are not often considered from an experimental point of view.

Two of the estimators studied belong to the OBE class (NLOBE and AO) and the other is based on the RLS method, all three selected owing to their main advantages described earlier. The aim of this work has been to compare the performance of different algorithms in *P. pastoris* bioprocesses. Online exhaust gas analyses or substrate concentration measurements were used to estimate the specific growth rate. With the aim to simplify and to reduce instabilities and the number of tuning parameters, biomass and substrate were straightforwardly obtained by solving their corresponding mass balances.

This article is structured as follows: first, process description and estimation algorithms are presented. Second, simulations for different alternatives are performed with the *P. pastoris* PAOX1-based system, Mut\(^{+}\) phenotype, discussing the most suitable ones using single and global performance metrics. Finally, experimental validation for both *P. pastoris* AOX Mut\(^{+}\) and FLD promoter is shown.

### Process Description and Modeling

#### Strains

The wild type *P. pastoris* X-33 strain (Mut\(^{+}\)), containing the vector pPICZzAROL, was used for heterologous expression of ROL under the control of the AOX1 promoter.\(^{34}\) *P. pastoris* X-33 containing the vector pPICZ2FLDzROL was chosen for the expression of ROL controlled by the FLD1 promoter.\(^{34}\)

#### Cultivation setup and operational conditions

Batch and fed-batch processes for the heterologous production by *P. pastoris* were studied. Cells were cultured in a 5-L Braun Biostat ED fermenter (Braun Biotech, Melsungen, Germany). Fermentation conditions were: stirring rate 800 rpm; temperature 30°C; pH controlled at 5.5 adding NH\(_4\)OH 30% (v/v) (batch) or 5 M KOH (fed-batch); KOH concentration was controlled above 30% with an air flow rate of 1.5–20 L min\(^{-1}\).

For the *P. pastoris* PAOX1-based systems, the cultivations were carried out in two phases: a first batch growth phase on glucose (3.5 L volume with an initial glucose concentration of 40 g L\(^{-1}\)) followed by a second phase (fed-batch), where first a mixture of glycerol plus methanol was fed to the culture and after that, only methanol was used as the sole carbon source.
The induction phase, 300 g
exceeded as follows: after a batch phase on glycerol, a pulse of
burette MicroBU-2031 from Crison Instruments (Alella, Bar-

2968 DOI 10.1002/aic Published on behalf of the AIChE October 2012 Vol. 58, No. 10 AIChE Journal

P. pastoris PFLD1

source and production of the recombinant protein took place.34
Cultivations using P. pastoris PFLD1-based systems pro-

ceeded as follows: after a batch phase on glycerol, a pulse of
sorbitol and methanol was added into the bioreactor. For

the induction phase, 300 g L−1 of sorbitol, 30 g L−1 of methyl-
amine chloride, and 7.5 mL of trace salts solution feeding
medium was used in an exponential feeding rate strategy.34

Glycerol, methanol, sorbitol, and methanol were added in the
transition and production phases by an automatic micro-
burette MicroBU-2031 from Crison Instruments (Alella, Bar-

celona, Spain). Online substrate measurements are possible by

the implementation of automated manifolds as previously

reported.35,36 Off-gas analyses were performed by infrared and

paramagnetic detectors (M ultor, Maihak, Hamburg, Ger-

many). Carbon dioxide production rate (CPR) and oxygen
uptake rate (OUR) were considered approximately equal to

carbon dioxide transfer rate (CTR) and oxygen transfer rate

(OTR), respectively, and, thus, obtained through mass bal-

ances of CO2 and O2 measured in the exhaust gas.

For the estimation procedure, the integration step and the
sampling time were fixed at 0.055 h (200 s). Online gas
measurments are a result of 20 averaged raw data points. So,

the filtered signal was used as the input for the estima-

tor.

A more detailed description of materials and methods can
be found elsewhere.34

The process model

For the oxidative assimilation of substrate (glycerol, meth-

anol, or sorbitol), equations for growth (1), protein produc-

tion (2), and maintenance (3) can be stated as

\[ k_1 S + k_2 O_2 \xrightarrow{r_x} X + k_3 CO_2 \]  

(1)

\[ k_4 S + k_5 O_2 \xrightarrow{r_p} P + k_6 CO_2 \]  

(2)

\[ S + k_7 O_2 \xrightarrow{r_m} k_8 CO_2 \]  

(3)

where X, S, O2, CO2, and P represent biomass, substrate,
dissolved oxygen, dissolved carbon dioxide, and product,
respectively, (in the sequel, the same symbols are used to
represent component concentrations); \( r_x, r_p, \) and \( r_m \) are
the reaction rates; \( k_i \) are the yield (stoichiometric) coefficients.

The corresponding dynamic model can be represented as

\[
\frac{d}{dt}
\begin{bmatrix}
X \\
S \\
O_2 \\
CO_2 \\
P
\end{bmatrix}
= 
\begin{bmatrix}
1 & -k_1 & -k_2 & -k_3 & -k_4 & -k_5 & -k_6 & -k_7 \\
0 & k_3 & k_6 & k_8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\cdot 
\begin{bmatrix}
\mu \\
q_p \\
q_m
\end{bmatrix}
\cdot 
\begin{bmatrix}
X - D \\
0 \\
0
\end{bmatrix}
+ 
\begin{bmatrix}
D \\
0 \\
0
\end{bmatrix}
\]  

(4)

were obtained from previous experiments in batch and fed-
batch cultivations.34,37,38 The yield and maintenance coeffi-
cients in the kinetic models were identified as previously
reported39,40 but adapted to fed-batch processes. Process
model parameters and coefficients are listed in Table 1.

Estimation Algorithms

The main objective of this work is to implement different
dicators, for the determination of state variables and spec-
cific growth rate using, if it is possible, single and easily
available measurements from a bioprocess system. Equally,
estimation algorithms can use either gas transfer rates or
substrate concentration as online measurements to first esti-
mate the specific growth rate without requiring a kinetic
model for growth, after that, state variables are calculated.

The estimators considered are two nonlinear observers
(NLOBE, AO) and another one based on the RLS method.

They are appropriate to systems, where component kinetic
rates (CKRs) can be considered related to biomass concen-
tration (X) and the specific growth rate (\( \mu \)) through a com-
ponent to biomass intrinsic yield (\( Y_{c/x} \)) and a maintenance

Table 1. Parameters and Coefficients Used for P. pastoris Process Under PAOXI (Mut+) and PFLD1 for ROL Production

<table>
<thead>
<tr>
<th>Parameters and Coefficients</th>
<th>PAOXI (Mut+)</th>
<th>PFLD1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch</td>
<td>Fed-batch</td>
<td>Fed-batch</td>
</tr>
<tr>
<td>Substrate</td>
<td>Glycerol</td>
<td>Methanol</td>
</tr>
<tr>
<td>( \mu_{\text{max}} ) (h−1)</td>
<td>0.260</td>
<td>0.050</td>
</tr>
<tr>
<td>( K_s ) (g L−1)</td>
<td>0.20</td>
<td>2.22</td>
</tr>
<tr>
<td>( Y_{\text{CO}_2,X} ) (mol \text{CO}_2 g X−1)</td>
<td>1.57 × 10−2</td>
<td>1.02 × 10−1</td>
</tr>
<tr>
<td>( m_{\text{CO}_2,X} ) (mol \text{CO}_2 g X−1 h−1)</td>
<td>2.6 × 10−4</td>
<td>3.1 × 10−4</td>
</tr>
<tr>
<td>( Y_{\text{O}_2,X} ) (mol \text{O}_2 g X−1)</td>
<td>2.4 × 10−2</td>
<td>1.69 × 10−1</td>
</tr>
<tr>
<td>( m_{\text{O}_2,X} ) (mol \text{O}_2 g X−1 h−1)</td>
<td>3.1 × 10−4</td>
<td>4.7 × 10−4</td>
</tr>
<tr>
<td>( Y_{\text{S},X} ) (mol \text{S} g X−1)</td>
<td>1.97</td>
<td>4.29</td>
</tr>
<tr>
<td>( m_{\text{S},X} ) (mol \text{S} g X−1 h−1)</td>
<td>0.008</td>
<td>0.010</td>
</tr>
<tr>
<td>( K_d ) (h−1)</td>
<td>–</td>
<td>360</td>
</tr>
<tr>
<td>( O_{\text{2},a} ) (mol \text{O}_2 L−1)</td>
<td>–</td>
<td>6.6 × 10−4</td>
</tr>
</tbody>
</table>
coefficient \((m_{c/x})\) described by the so-called Luedeking–Piret relationship as follows

\[
\text{CKR} = \frac{Y_{c/x} \mu X + m_{c/x} X}{Y_{c/x} \mu X} \approx \frac{Y_{c/x} \mu X}{Y_{c/x} \mu X}
\]  

(7)

So, the proposed methodology can be implemented, where either the CPR, the OUR, or the substrate uptake rate (SUR) can be determined online.

**Nonlinear observer-based estimator**

The first estimator tested is based on a general design for a nonlinear observer\(^{27}\) represented by

\[
\frac{d\hat{z}}{dt} = F(s, \hat{z})\hat{z} + G(u, s, \hat{z}) - \Theta^{-1}(s, \hat{z})S_0^{-1}c^T(C\hat{z} - y)
\]  

(8)

where \(\hat{z}\) is the estimated state and \((u, s)\) and \(y\) are, respectively, the input and the output of the system.

The used estimator considers that only \(z_1\) is available as online measurement. It is correlated to biomass and the specific growth rate \(\mu\) component.

Asymptotic observer with second-order dynamics tuning

The second estimator tested was an AO, which is used for the estimation of state variables. The specific growth rate is obtained using an estimator based on the reformulation proposed by Bastin and Dochain.\(^{25,30}\)

A general dynamic model for stirred tank bioreactors can be described following mass balance equations written in a matrix form as

\[
\frac{d\hat{\mu}}{dt} = (\hat{z}_2 - \hat{\mu})\left(\hat{\mu} + m_{\text{CO}_2/X}/Y_{\text{CO}_2/X}\right)
\]  

(14)

In this approach, a priori, only a single tuning parameter \(\delta\) is needed to calibrate the estimation procedure. However, suitable initial values for \(\hat{z}_{10}, \hat{z}_{20}\), and \(\hat{z}_{30}\) have to be provided to obtain satisfactory estimates.

**Asymptotic observer with second-order dynamics tuning**

Combining Eqs. 16 and 17 with Eq. 18, the estimator is obtained

\[
\frac{d\hat{\mu}}{dt} = H\hat{\mu} - D\hat{\mu} + F - Q - \Omega(\hat{\mu} - \hat{\mu})
\]  

(19)

where the gain matrices \((\Omega_C, \Gamma_C)\) and their parameters are defined as follows

\[
\Gamma_C = \text{diag}\{-\gamma\}, \quad \Omega_C = \text{diag}\{-\omega_i\}
\]  

(20)

The estimation of biomass \((X)\) is obtained from one measurement \(C\) (substrate, oxygen, or carbon dioxide concentration and their respective feeding or gaseous outflow rates), according to the choice of a subspace \(c\). Thus, the system model from Eq. 6 is simplified, according with the expression (23)

\[
\frac{d}{dt} \begin{bmatrix} X \\ C \end{bmatrix} = \begin{bmatrix} 1 & \pm k_C^c \\ \pm k_C^c & D \end{bmatrix} \cdot \begin{bmatrix} X \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ F_C - Q_C \end{bmatrix}
\]  

(23)

Finally, biomass estimation is calculated by an auxiliary variable \((Z)\) that depends on \(C\) concentration, CKR measurements, or the feeding rate, the overall yield coefficient and biomass.\(^{25}\)

\[
\hat{Z} = \hat{\mu} + \frac{1}{\pm k_C^c} C
\]  

(24)
\[
\frac{dZ}{dt} = -DZ + \frac{1}{4k_c} (F_c - Q_c)
\]  
(25)

\[
\hat{X} = Z - \frac{1}{4k_c} C
\]  
(26)

As previously described,\textsuperscript{28,29} when \(D\), the dilution rate, is kept close to zero for long period of time, the performance of the \(AO\) is expected to be not satisfactory, as the rate of convergence of the estimation fully depends on the values of that variable. Therefore, the \(AO\) estimation procedure was not applied in the batch phase and will be tested and validated only in fed-batch operation. Furthermore, although the kinetics of the process may be considered unknown in these observers, the estimation of the state variables requires an accurate knowledge about the reaction scheme and stoichiometric coefficients. Consequently, uncertainties on these model parameters, besides noisy online measurements can generate a large bias in the estimation procedure. For that reason, it is advisable, as implemented by some authors,\textsuperscript{19,31} to use filtered data with setting bounds on \(\mu\) (0.05\(\mu_{\text{max}}\) \(\leq \mu \leq \mu_{\text{max}}\)) to avoid unrealistic estimated values for the specific growth rate and \(Y_{s/C}\) in Eq. 7.

**Recursive least squares with variable forgetting factor**

Estimators based on the RLS method consider a linear model for the system within the time interval, where the identification procedure is performed. Some authors have shown the capacity of linear estimators to adequately estimate the fermentation process without increasing the model structure complexity.\textsuperscript{42,52} In this way, some approaches have been developed to carry out the estimation from measurement of oxygen\textsuperscript{31} or carbon dioxide in the exhaust gas of the fermentor.\textsuperscript{32} It has been applied to a type of processes where a sole substrate for growth and induction was added in a noncontinuous mode. From Eq. 7, considering the CPR and its time-derivative, Eq. 27 can be obtained.

\[
\frac{d\text{CPR}(t)}{dt} = \frac{d(\mu(t))}{dt}X(t)Y_{\text{CO}_2/X} + \frac{dX(t)}{dt} Y_{\text{CO}_2/X}\mu(t)
\]  
(27)

Then, considering the mass balance for biomass

\[
\frac{d\text{CPR}(t)}{dt} = \frac{d(\mu(t))}{dt}X(t)Y_{\text{CO}_2/X} + (\mu(t) - D(t))\text{CPR}(t)
\]  
(28)

where \(\mu(t)\) is an unknown time-varying parameter, CPR(t) is an indirect online measurement, and biomass \(X(t)\) cannot be measured online.

Considering that \(\mu\) varies slowly within the sampling interval, the term corresponding to the \(\mu\) first-time derivative in Eq. 28 can be neglected. Discretization of the resulting equation is conducted using a first-order central Euler approximation

\[
\frac{\text{CPR}(t_{k+1}) - \text{CPR}(t_k)}{2\Delta t} = (\mu(t_k) - D(t_k))\text{CPR}(t_k)
\]  
(29)

Thereupon, a time varying parameter \(\hat{\theta}(t_k)\) can be defined, which will be recursively estimated with the RLS method.

\[
\text{CPR}(t_{k+1}) = \hat{\theta}(t_k)\text{CPR}(t_k) + \text{CPR}(t_{k-1})
\]  
(30)

From the previous estimate, the specific growth rate is obtained

\[
\mu(t_k) = \frac{\hat{\theta}(t_k)}{2\Delta t} + D(t_k)
\]  
(31)

The proposed estimator uses the RLS method to estimate \(\hat{\theta}(t_k)\) according to the set of equations presented below

\[
\hat{\theta}(t_k) = \hat{\theta}(t_{k-1}) + K(t_k)(y(t_k) - \hat{y}(t_k))
\]  
(32)

\[
\hat{y}(t_k) = \Psi^T(t_k)\hat{\theta}(t_{k-1})
\]  
(33)

\[
K(t_k) = Q(t_k)^{1/2}P(t_{k-1})
\]  
(34)

\[
Q(t_k) = \frac{P(t_{k-1})}{\lambda + \Psi(t_k)^T P(t_{k-1}) \Psi(t_k)}
\]  
(35)

\[
P(t_k) = \frac{1}{\lambda} \left\{ P(t_{k-1}) - \frac{\Psi(t_k)^T P(t_{k-1}) \Psi(t_k)}{\lambda + \Psi(t_k)^T P(t_{k-1}) \Psi(t_k)} \right\}
\]  
(36)

where \(P(t)\) is the covariance matrix, \(Q(t)\) is an auxiliary matrix, \(y(t)\) is the CPR(t), and \(\Psi\) the data vector.

In the heterologous protein production by \(P.\) pastoris, sometimes there are important changes in the system characteristics, such as substrates, concentrations, and operating conditions, which may lead to less satisfactory estimation results. The nonlinear dynamics of the specific growth rate can be included in the estimator, considering a variable forgetting factor (VFF) \(\lambda(t)\) to conform the proposed estimator. The time-varying forgetting factor prevents the constant reduction in the value of the covariance matrix during the dynamic process.\textsuperscript{32} The calculation of the VFF is based on model error and both the data vector and covariance matrix.

\[
\lambda(t_k) = 1 - \frac{(y(t_k) - \hat{y}(t_k))^2}{\sum_0^n \left( 1 - \frac{\Psi(t_k)^T P(t_{k-1}) \Psi(t_k)}{\lambda(t_{k-1}) + \Psi(t_k)^T P(t_{k-1}) \Psi(t_k)} \right)}
\]  
(37)

\(\Sigma_0\) is the only tuning parameter and it needs to be determined empirically, because it depends on process dynamics, sampling time, and measurement noise.

Finally, it is necessary to impose biological restrictions in the \(\mu\) estimation to be always positive and bounded (\(\mu_{\text{max}}\)).

**Performance indicators**

Single metrics sum of squared error (SSE), root mean squared error (RMSE), mean relative error (MRE), root mean squared noise sensitivity (RMNS), integral time-weighted absolute error (ITAE), and rise time (RT) were used for the evaluation of the “goodness-of-estimation” in the simulation phase. SSE, RMSE, and MRE were calculated for \(\mu\), \(X\), and \(S\) estimations. RMNS, ITAE, and RT were obtained for \(\mu\) estimates.

The SSE can be defined as follows

\[
\text{SSE} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2
\]  
(38)

where \(n\) is the number of data points, \(\hat{y}_i\) is the \(i\)th estimated value, \(y_i\) is the corresponding \(i\)th actual value from process model.
Table 2. Simulation Conditions in Batch and Fed-batch *P. pastoris* Process Under FAAOX (Mut⁺) for ROL Production

<table>
<thead>
<tr>
<th>Process Variables</th>
<th>Units</th>
<th>Batch</th>
<th>Fed-batch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate (S)</td>
<td>g L⁻¹</td>
<td>Glycerol</td>
<td>Methanol</td>
</tr>
<tr>
<td>Initial concentration (S₀)</td>
<td>g L⁻¹</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>Inlet concentration (Sᵢ)</td>
<td>g L⁻¹</td>
<td>–</td>
<td>790</td>
</tr>
<tr>
<td>Process time (t)</td>
<td>h</td>
<td>14.7</td>
<td>60.3</td>
</tr>
<tr>
<td>Initial volume (V₀)</td>
<td>L</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Initial Biomass (X₀)</td>
<td>g L⁻¹</td>
<td>0.50</td>
<td>20</td>
</tr>
<tr>
<td>Specific growth rate</td>
<td>h⁻¹</td>
<td>–</td>
<td>0.02</td>
</tr>
<tr>
<td>set-point (μₛₑ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial specific growth rate (μ₀)</td>
<td>h⁻¹</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Sampling time (T)</td>
<td>h</td>
<td>0.055</td>
<td>0.055</td>
</tr>
</tbody>
</table>

The RMSE is defined by

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}
\]

that is, the average difference between estimated and actual target variables (specific growth rate, biomass, and substrate), where \(y_i\) and \(\hat{y}_i\) are the estimated and true value for the \(i\)th data point; \(n\) is the total number of data points. Alternatively, calculation of the MRE is also carried out to examine the “goodness-of-estimation.”

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i}
\]

The sum of squared noise sensitivity accounts for the continuous variation of \(\mu\) estimations and it is defined as follows

\[
SSNS = \sum_{i=1}^{n-1} (\hat{\mu}_{i+1} - \hat{\mu}_i)^2
\]

Root mean squared noise sensitivity (RMNS) is then calculated as the average differences between consecutive estimated specific growth rate values \(\hat{\mu}_{i+1}\) and \(\hat{\mu}_i\) from \(i = 1\) to \(n−1\) data points.

\[
RMNS = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} (\hat{\mu}_{i+1} - \hat{\mu}_i)^2}
\]

In addition, overall performance indicators: root mean overall performance index (RMOPI) and combined mean relative error (CMRE), which consider some former single metrics, were used.

\[
RMOPI = \sqrt{w_\mu RMSE^2_\mu + w_x RMSE^2_x + w_s RMSE^2_s + w_w RMSE^2_w + w_m RMSE^2_m}
\]

\[
CMRE = w_x MRE_\mu + w_m MRE_s
\]

\[
ITAE = \int te
dt
\]

\[
F_{in}(t) = \frac{\mu_{set}X(0) V(0)}{V_{clear} S_b} \exp[\mu_{set} \frac{t - t(0)}{1}]
\]

Simulation Results

The simulated system accounts for the cultivation of the Mut⁺ phenotype of *P. pastoris* under pAOX for the ROL production. To study the performance of the different estimators, a process model was developed considering mass balances (X, S, O₂, and CO₂) and CKR expression for carbon dioxide, oxygen, and substrate (CPR, OUR, and SUR) as stated in Eqs. 6 and 7.

**NLOBE**, two sorts of AO-SODE and RLS-VFF estimation algorithms were studied. First, **NLOBE** and **RLS-VFF** estimation were carried out from CPR measures. Second, the estimation was made using substrate concentration (methanol) for the first of the **AO-SODE**’s. Third, the **AO-SODE** estimation performance, for the second **AO-SODE** estimator, was obtained from oxygen balance, in which the main variables are the dissolved oxygen concentration and the OTR. Thus, it is necessary to include growth kinetics (\(\mu(S)\)) and also the OTR model for the **AO-SODE** simulations. The specific growth rate is approximated by the “Monod” equation. The OTR is computed as OTR = \(K_L\alpha (O_{2 sat} - O_2)\), where \(K_L\alpha\) is the global mass-transfer coefficient and \(O_{2 sat}\) is the saturation concentration of dissolved oxygen.

For the simulation, and after a batch phase on glycerol, an induction phase in fed-batch mode was started by an exponential addition of methanol as substrate.
Table 3. Initial Guesses, Nominal Values, and Acceptable Ranges for Tuning Parameters and Initialization Estimator Variables

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameters and Estimator Variables</th>
<th>Units</th>
<th>Initial Guesses</th>
<th>Nominal Value</th>
<th>Acceptable Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLOBE</td>
<td>$\delta$</td>
<td>[h$^{-1}$]</td>
<td>20</td>
<td>51</td>
<td>$\pm 1$</td>
</tr>
<tr>
<td></td>
<td>$z_{10}$</td>
<td>[mol L$^{-1}$ h$^{-1}$]</td>
<td>0.5</td>
<td>4.0</td>
<td>$\pm 0.3$</td>
</tr>
<tr>
<td></td>
<td>$z_{20}$</td>
<td>[h$^{-1}$]</td>
<td>5.0</td>
<td>2.0</td>
<td>$\pm 0.1$</td>
</tr>
<tr>
<td></td>
<td>$z_{90}$</td>
<td>[h$^{-1}$]</td>
<td>50</td>
<td>100</td>
<td>$\pm 3$</td>
</tr>
<tr>
<td>AO-SODE (Methanol)</td>
<td>$\zeta$</td>
<td>[h$^{-1}$]</td>
<td>1.0</td>
<td>0.70</td>
<td>$\pm 0.40$</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>[h$^{-1}$]</td>
<td>0.05</td>
<td>0.20</td>
<td>$\pm 0.11$</td>
</tr>
<tr>
<td>AO-SODE</td>
<td>$\zeta$</td>
<td>[h$^{-1}$]</td>
<td>1.0</td>
<td>0.70</td>
<td>$\pm 0.40$</td>
</tr>
<tr>
<td></td>
<td>$\tau$</td>
<td>[h$^{-1}$]</td>
<td>0.05</td>
<td>0.20</td>
<td>$\pm 0.11$</td>
</tr>
<tr>
<td>RLS-VFF</td>
<td>$\Sigma_0$</td>
<td>[h$^{-2}$]</td>
<td>$5.0 \times 10^{-8}$</td>
<td>$1.0 \times 10^{-8}$</td>
<td>$\pm 0.9 \times 10^{-8}$</td>
</tr>
<tr>
<td></td>
<td>$\theta_0$</td>
<td>[-]</td>
<td>$2.0 \times 10^{-3}$</td>
<td>$1.5 \times 10^{-2}$</td>
<td>$\pm 0.5 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

The initial biomass concentration, $X(0)$; the substrate concentration in the feed, $S_0$; and the biomass to substrate overall yield, $Y_{X/S}$ = $Y_{X/S}^{-1}$.

The CTR (CPR) and OTR (OUR) simulated data were corrupted by an additive Gaussian noise of mean zero and relative standard deviation of 2%, considered as a typical value after data filtering. Noise applied to substrate measurements was 6%.

Initial conditions for the simulations are given in Table 2. Initial estimates are taken as same as the system model and are representative of the real process: $(X(0), \mu(0), S(0)) = (X(0), \mu(0), S(0))$.

In general, tuning parameters and some initialization variables were determined empirically considering the convergence speed and noise sensitivity, which are tightly related to the deviation between estimations and real data. Weighted average of mean square deviations can be used as part of the objective function.

In this sense, the RMOPl indicator described before was used as cost criterion to be minimized. Depending on the sensitivity of this RMOPl on parameters and initialization values, nominal values were rounded to one or two significant digits. In some cases, restrictions related to stability and convergence properties were applied to get preliminary values. Conversely, sometimes these values can be obtained straightforwardly attending to the particular structure of the estimator. Details related to each estimator are specified separately.

Tuning parameters and initialization variables

In the NLOBE estimator, initial guesses for $\delta$, $z_{10}$, $z_{20}$, and $z_{30}$ were selected according to the structure of the estimator: $\delta \approx 1/T$; $z_{10} \approx CPR_{max}$; $z_{20} < (\mu_{max}/T)/\mu$ for the maximal value of $d\mu/dt$; $z_{30} = 10\times z_{20}$. Initial guesses and nominal values, obtained by RMOPl minimization, are listed in Table 3. Reinitialization of the estimator parameters is needed when the operational mode is switched from batch to fed-batch, when protein production is induced. Because of different system dynamics during batch and fed-batch culture, the estimator is not able to overcome the change on feeding rate ($D > 0$), if a similar estimation quality is required.

Two parameters must be tuned in the AO-SODE algorithm: $\zeta$ is the damping coefficient and $\tau$ is the time constant of a second-order dynamic response of estimated to the true values. The damping coefficient, $\zeta$, was preliminary fixed at 1 according to a common engineering rule of thumb. Tuning rules for discrete time implementation were applied to determine preliminary value of $\tau$ according to $0 < T < 2\zeta\tau$, thus, setting $\tau$ to 0.05. Nominal values were finally obtained by minimization of the cost criterion function RMOPl and they are shown in Table 3 for the two AO-SODE methods tested.

The tuning of the RLS-VFF estimator starts with the assignment of first trials to the tuning parameter $\Sigma_0$ and the initial value $\theta_0$.

$\Sigma_0$ has to be set to manage slow or moderate variations of $\mu$ by the VFF. So, $\Sigma_0$ initial settings were fixed to be the same order of magnitude that square deviations in Eq. 37.

For the initial value $\theta_0$, according to the inner structure of the RLS estimator, a preliminary value $\theta_0 \approx 2T\mu$ can be assigned.

Initial and nominal values, obtained by RMOPl minimization, for $\Sigma_0$ and $\theta_0$ can be found in Table 3. In contrast with NLOBE estimator, retuning of the estimator parameters was not necessary, when the operational mode was changed. The RLS-VFF method was able to cope up with the different dynamics of the operational modes, batch and fed-batch process.

Initial value for the covariance matrix $P_0$ was set at 1, because simulations with high and low $P_0$ showed that it converges quickly, and weakly influence, to the estimation.

Estimators performance

The estimation results for NLOBE, AO-SODE, and RLS-VFF methods are illustrated in Figure 1, Figure 2, and Figure 3, respectively. The performance of the estimators at their nominal values for the tuning parameters and initialization variables are presented in Table 4. The goodness of estimation for NLOBE and RLS-VFF methods is nearly as satisfactory as for AO-SODE, which shows the higher values for the single metrics (SSE, RMSE, and MRE) for $\mu$ values. However, all $\mu$ and $X$ estimations were conducted properly with deviations lower than 1% (MRE) and with similar values of the overall performance indexes (RMOPl and CMRE). At this point, it is remarkable to emphasize that substrate estimations have a large MRE > 1, due to substrate low values, which is confirmed by its RMSE. Terms in substrate balance corresponding to inlet and uptake rate are quite similar and so, computed substrate concentration is very sensitive to the estimation of $\mu$.

The speed of convergence, indirectly measured by ITAE and RT, is also satisfactory for all estimators. Fast enough response to the change on the substrate feeding strategy is considered for RT appeared to be shorter than 1 h, being the longest time for both AO-SODE estimators. Nonetheless, ITAE values are slightly better for AO-SODE estimators. In addition, all estimators showed that dynamics of convergence are time varying and becoming faster, when the bio-process approaches the end.
Finally, the noise sensitivity of the estimators was quantified by the RMNS index and displayed in Table 4. The NLOBE estimator is almost as sensitive as others but provides the higher time-varying dynamics. However, NLOBE estimator gives the lowest estimation error RMSE. In contrast, AO-SODE methods present the highest RMSE values with low RMNS. These facts are a consequence of the tuning procedure applied to obtain nominal values, which are a compromise between estimation error and noise sensitivity evaluated through the RMOPI.

**Sensitivity analysis**

To make a thorough sensitivity analysis, first, NLOBE’s simulations were carried out to study the effect of the tuning parameter $d$ and the initial value of some variables on the estimation results. The effect of $d$, $\hat{z}_{10}$, $\hat{z}_{20}$, and $\hat{z}_{30}$ on the estimation of the specific growth rate from CPR measurements is visualized in Figure 1. Regarding the effects of the tuning parameter, high values of $d$ provoke a typical overshoot at the start-up of the fed-batch phase. Moreover, a high variation on overall performance is observed for low to moderate $d$ variations as displayed in Figures 1 and 4. In spite of having only one tuning parameter, the extremely sensitivity of the RMOPI to $d$ is the main drawback of the NLOBE. Equally, the results obtained for the initialization variables were presented in Figures 1 and 4. Results presented in Figure 4 shows more clearly the largest sensitive response. RMOPI is less sensitive to $\hat{z}_{10}$ and $\hat{z}_{20}$ than $\hat{z}_{30}$, but the last one is no harder sensitive than $d$. These results were somewhat expected because $z_3$ was defined as the $z_2$’s time derivative, which is supposed to be unknown and bounded. The dynamics of $z_3$ introduces some type of integral action, which eliminates any static error when estimating $z_2$.

Second, the effect of AO-SODE tuning parameters $\zeta$ and $\tau$ on the dynamics of convergence can be assessed from the plots in Figure 2. Figures 2a,b illustrate that specific growth rate values are in agreement with typical second-order dynamic responses. It is shown that decreasing $\tau$, the response becomes faster and decreasing $\zeta$ produces more oscillatory responses. With regard to stability conditions, the range allowed for the integration step $T$, set equal to the sampling time, is bounded and conditioned by the choice of $\zeta$ and $\tau$. These facts allow searching and selecting the tuning parameters with intuitive basis, simplifying the search for optimum.

As it is shown in Figure 2, X estimations are quite sensitive to deviations of the specific growth rate. When the estimation started, long time was taken to reach the specific growth rate for $s$ values rather far from nominal values. This initial $X$ offset remained quite stable through the bioprocess, whenever $\mu$ estimation was well-fitted to the simulated system response. Finally, to calculate the sensitivity of the tuning parameters $\zeta$ and $\tau$, the RMOPI combined metrics was explored. In contrast to Figure 2, it can be observed in Figure 4 that the effect of $\zeta$ is not different than those of $\tau$. This fact can be explained due to the tuning parameter $\tau$ in Figure 2 was tested in a wider range than $\zeta$, with the aim to easily appreciate the overall estimation behavior.

Third, in the second case of AO-SODE, Figures 2e–j, based on dissolved O$_2$ and OUR data, the dynamics of convergence for tuning parameters $\zeta$ and $\tau$ were very similar to those illustrated for the methanol case. No significant differences on $X$ and $\mu$ estimations were shown for all AO-SODE estimators tested (Methanol, O$_2$ and OUR, CO$_2$ and CPR). Nevertheless, AO-SODE estimators using gas measures also allowed computing substrate concentration. This estimation was performed by applying the substrate balance, using $X$ and $\mu$ estimation, as indicated in Eqs. 18–26. Thereupon, it can be observed in Figures 2i,j, substrate estimations are more sensitive to deviations of the specific growth rate and
biomass concentration than in their \( \zeta \) and \( \tau \) nominal values, as commented in the previous section. In addition, for the cases where the specific growth rate is estimated using gas measures (O\(_2\) or CO\(_2\) and OTR or CTR), the effect of the dissolved concentration in the AO-SODE algorithm is negligible compared with the gaseous transfer rates.

Finally, the study of RLS-VFF was focused on the effect of the initial value \( \theta_0 \) and the tuning parameter \( \Sigma_0 \). Their effect on \( \mu, X, \) and \( S \) estimation, using CPR measurements, is presented in Figures 3 and 4. As it can be observed, the performance of the estimation algorithm is more sensitive to \( \theta_0 \) values than for \( \Sigma_0 \). In Figures 3c,d, the tuning parameter \( \Sigma_0 \) for biomass calculations seems to have a stronger effect than the initial value of \( \theta_0 \), due to the wide range explored for \( \Sigma_0 \). From Figure 3, if a low \( \Sigma_0 \) value is introduced in the estimation of \( \mu \), it gives robustness to the overall procedure, providing a very satisfactory state estimation. These results clearly highlight the practical implementation of the proposed algorithm. With suboptimal \( \theta_0 \) values, a good convergence speed of the estimator is not attained, as can be seen in Figure 3 at the beginning of the fed-batch process, and, consequently, large deviations of substrate estimations are produced.

**Selection of the estimator**

Estimations showed similar good agreement with their actual values for the specific growth rate, biomass, and substrate concentrations for all estimators tested. This goodness of estimation has been evaluated by overall performance indexes RMOPI and CMRE presented in Table 4. However, significant differences were observed concerning the range of application for tuning parameters and initialization variables, speed of convergence, number of tuning parameters needed and requirements for system model.

The acceptable ranges of tuning parameters were shown in Table 3. These ranges were calculated according to the sensitivity analysis on CMRE indicator to obtain variations always lower than 5%, except in the cases of \( \zeta \), \( \tau \), and \( \Sigma_0 \) with 1.5% maximal CMRE variation. This is justified due to stability criteria used for AO-SODE, which imposes setting bounds on tuning parameters (\( \zeta \), \( \tau \)) and the far low sensitivity of CMRE to \( \zeta \), \( \tau \), and \( \Sigma_0 \).

From results given in Table 3 and Figure 4, it can be concluded that tuning parameters and initial values for the NLOBE have stronger influence on the overall estimator performance than the AO and RLS methods. Furthermore, there is no much difference on influence between AO and RLS methods, although the RLS showed more dependence on the \( \theta_0 \) initial value.

Speed of convergence indicated in Table 4 by the RT was similar to all three estimators studied (<1 h), but presenting a rather sluggish initial response for the AO-SODE. It is not clearly reflected in the ITAE value because it weights initial errors less heavily than those which persist on time. The number of critical initialization variables is greater in the NLOBE estimator, with some of them having high influence on the RMOPI (Figure 4). It has to be pointed out the effect of \( \delta, \bar{z}_{10}, \bar{z}_{20}, \) and \( \bar{z}_{30} \), which a variation of 2, 8, 5, and 3%, respectively, provides an 5% increase on CMRE as can be straightforwardly obtained from Table 3.

Observer-based estimators (NLOBE, AO) are not recommended in a preliminary choice with poorly known kinetic parameters because of their fundamentals relies on an accurate knowledge of yields and maintenance equations. However,
their stability, dynamics of convergence, and parameter tuning can be improved using a variable gain-structure-like AO-SODE estimators. With this kind of variable gain observers, moderate to rapid changes in the specific growth rate can be anticipated properly. They allow higher identification frequency and reveal good adaptive behavior to process changes with high speed of convergence. As its main disadvantage, its application in the batch mode is not recommended due to expected poor performance and weak stability.

Making use of RLS-VFF methods allows decreasing the knowledge requirements about the system, not only in quantitative terms but also in a qualitative or schematic system description. Among other advantages, the use of linear equations instead of differential equations can be pointed out. Finally, a reduced number of tuning parameters gives an added value to this kind of estimators. The adaptive computation of the forgetting factor in the RLS method makes possible that moderate changes in the system, such as in substrates, state variable concentrations, and operational conditions, can be processed to conduct a satisfactory tracking of the specific growth rate. This approach allows that, when the operational mode is switched from batch to fed-batch, reinitialization of estimator variables is not necessary.

On the whole, taking into account, simulation studies RLS-VFF methods and AO-SODE estimators are the most promising ones. The main advantage for both is their tuning simplicity compared to the NLOBE tested. In addition, RLS-VFF does not require kinetic coefficients and AO-SODE presents good stability within a wide range of tuning parameter values. Thereupon, in the next section, experimental validation is carried out for both estimators.

**Experimental Validation**

The RLS-VFF method was applied to experimental data from both *P. pastoris* PAOX1- and PFLD1-based systems. Experimental data for the specific growth rate were obtained from off-line biomass measurements using suitable smoothing spline functions.37

First, the cultivation of the Mut+ phenotype of *P. pastoris* under PAOX1 for the ROL production was studied. After batch cultivation on glycerol and transition phase, an induction phase in fed-batch mode was started by a preprogrammed exponential feeding of methanol.

As it can be seen in Figure 5, $\mu$ estimation and computed biomass estimation were carried out properly to the system.
response and a rapid convergence was obtained for the \( \mu \) estimation during the batch phase. When the batch phase was completed, as found at 20 h, unbiased \( \mu \) and \( X \) estimations to substrate changes were also shown. This fast adjustment is again confirmed, as in previous section, during the fed-batch mode. First, growth slowdown is observed, the growth was almost practically stopped mainly due to the regulation of the microorganism to the new conditions. After that, an increase in the growth rate is obtained. The estimator performance allows adapting to growth variations and, then, estimating the computed biomass successfully.

With the aim to validate the estimation algorithms in other systems with different promoters and process dynamics, experimental data from \( P. \) *pastoris PFLD1*-based system was processed. In PFLD1 system, a comparable experimental strategy than for the PAOXI system was used, but now using sorbitol and methylamine instead of methanol and ammonium hydroxide as carbon and nitrogen sources, respectively. Variable substrate feeding was performed, with a nonautomatic substrate control procedure, attempting to keep the substrate concentration at 8 g L\(^{-1}\).

With the goal to make a comparison among estimation methods, single metrics: RMSE and MRE were used to evaluate the goodness-of-estimation for the state variables, such as biomass and substrate, when they can be measured directly but they are not used as online measurement. A summary is presented in Table 5.

Results for the RLS-VFF algorithm are shown in Figure 6 and Table 5. Once the specific growth rate was properly estimated, biomass and substrate can be calculated through the use of their corresponding mass balances. Biomass and substrate were estimated successfully with an estimation error \( \leq 5\% \), although slight deviations were detected for substrate. From Figure 6, it was observed that the estimated substrate concentration and specific consumption rates are in good agreement with those obtained through substrate balancing with off-line specific growth rate data.

The estimation procedure showed suitable capacity of adjustment during the bioprocess, especially when switching from batch phase to fed-batch mode with different substrate types and changing concentration. Interestingly, the quality of the estimation was nonpermanently affected by
operational problems caused by aeration and foaming shift-up at 68 h.

Furthermore, the performance of the AO-SODE observer was also tested in the P. pastoris PFLD1-based system for validation and, finally, compared with the RLS-VFF estimator.

Results from specific growth rate and biomass estimation using substrate or oxygen measurements are displayed in Figures 7a–c. Computed biomass estimation was achieved with an estimation error ≤5%, although performance was better when gaseous measurements were used instead of substrate measurements as presented in Table 5. The low accuracy of the values of the overall yield coefficients used could explain the differences in biomass estimations.29

As previously described, apart from X and μ estimations using gaseous measurements, substrate estimation was performed from the substrate balance and SUR estimation. Similarly, the substrate concentration was reconstructed using off-line X and μ data with the aim of comparing the accuracy of computed estimates. Although slight deviations in the computed substrate were detected, reconstructed substrate has a similar precision than the estimated one.

Notably, when the AO-SODE method was applied to O2 and OUR data or CO2 and CPR measurements, similar results were observed. However, AO-SODE (O2, OUR) was evaluated achieving the least deviation error (RMSE and MRE) of state variables, as detailed in Table 5.

Finally, all estimation methods could follow substrate evolution with less deviation error than that obtained in simulation results, as presented in Tables 4 and 5.

Table 4. Single and Combined Performance Metrics for NLOBE, AO-SODE, and RLS-VFF Estimators

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Variable</th>
<th>SSE</th>
<th>RMSE</th>
<th>MRE</th>
<th>RMNS</th>
<th>RT</th>
<th>ITAE</th>
<th>RMOP</th>
<th>CMRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLOBE</td>
<td>μ</td>
<td>1.2×10^-4</td>
<td>3.4×10^-4</td>
<td>5.0×10^-3</td>
<td>6.5×10^-4</td>
<td>0.27</td>
<td>0.11</td>
<td>0.17</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>24</td>
<td>0.15</td>
<td>3.1×10^-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>370</td>
<td>0.59</td>
<td>&gt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AO-SODE (Methanol)</td>
<td>μ</td>
<td>1.3×10^-3</td>
<td>1.1×10^-3</td>
<td>7.2×10^-3</td>
<td>2.2×10^-4</td>
<td>0.50</td>
<td>0.06</td>
<td>0.24</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>46</td>
<td>0.20</td>
<td>5.4×10^-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>–</td>
<td></td>
<td>&gt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AO-SODE</td>
<td>μ</td>
<td>1.8×10^-3</td>
<td>1.3×10^-3</td>
<td>1.0×10^-2</td>
<td>1.8×10^-4</td>
<td>0.72</td>
<td>0.12</td>
<td>0.27</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>1.6</td>
<td>3.8×10^-2</td>
<td>1.2×10^-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>510</td>
<td>0.68</td>
<td>&gt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS-VFF</td>
<td>μ</td>
<td>2.3×10^-4</td>
<td>4.6×10^-4</td>
<td>9.1×10^-3</td>
<td>3.5×10^-5</td>
<td>0.10</td>
<td>0.18</td>
<td>0.11</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>7.3</td>
<td>0.08</td>
<td>&gt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>170</td>
<td>0.40</td>
<td>&gt;1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

Simulation results obtained for the nonlinear OBE’s showed adequate global performance, but an optimal tuning cannot be easily derived. Despite of their stability, dynamics of convergence and parameter tuning could be improved using a variable gain-structure, it is necessary to know with high accuracy the yield and maintenance coefficients of the bioprocess model. However, these values can be different, depending on the carbon source and environmental conditions. So, when complex nutrients are consumed in a significant quantity, these parameters may also vary during the fermentation process. Thus, errors on model coefficients may produce inaccurate results in estimation of μ and, consequently, the error is propagated and amplified on X and S estimation. Hence, these observers are not recommended for processes with poorly known parameters in a preliminary selection. In addition, NLOBE was established as the most tuning sensitive, being rather dependent on their tuning parameters and initial values.

The use of RLS methods to identify the specific growth rate allows diminishing requirements about the knowledge of the system, besides providing other important advantages. Among them are: low mathematical complexity because it is only necessary to solve linear equations instead of differential equations like in OBE’s, increase of identification frequency, high adaptation capacity to process changes, short
response times, and reduced number of both tuning and initialization variables. The implementation of the RLS-VFF method allows that the whole estimation system is fairly dependent on the tuning parameters applied. Furthermore, the VFF allows that a minimal excitation of the estimator is maintained through the bioprocess preventing a constant reduction of the covariance matrix. In the cases studied, this is a great advantage, since there is a continuous addition of substrate in a relatively long process with a low specific growth rate. Moreover, it includes phases where the carbon source has to be replaced and also operational perturbations often appear. This method was able to make a suitable identification of the specific growth rate both in batch and fed-batch processes, whenever slow variation of the specific growth rate is presented.

AOs as well as RLS methods can be applied to correctly estimate the specific growth rate. Concretely, AO-SODE performed better than the RLS-VFF, when moderate to rapid changes of the specific growth rate appeared because model parameters were well-known. Conversely, when slow changes on the specific growth rate were presented in the bioprocess, for instance, in a substrate control operation, RLS-VFF comes up as the best option, because of its reduced requirements. In addition, biomass and substrate were also satisfactorily predicted, solving their corresponding mass balances once the specific growth rate had been estimated.

To conclude the method, which was resulted especially efficient for the proposed system was the AO-SODE (O2, OUR) and was evaluated achieving the least deviation error. The overall methodology presented in this article can be used to improve the global performance of the process in terms of productivity, yields, and reproducibility in the heterologous protein production by *P. pastoris* and for real-time monitoring of the key fermentation variables. This will represent a significant contribution through a more efficient process of heterologous protein production by *P. pastoris*.  

### Table 5. Errors of the Estimation Algorithms for Estimated State Variables in the *P. pastoris PFLD1*-Based System During the Fed-batch Mode

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Estimated Variable</th>
<th>RMSE</th>
<th>MRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLS-VFF (CPR)</td>
<td>Biomass [g L⁻¹]</td>
<td>1.2</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>Substrate [g L⁻¹]</td>
<td>1.3</td>
<td>0.16</td>
</tr>
<tr>
<td>AO-SODE (CO₂, CPR)</td>
<td>Biomass [g L⁻¹]</td>
<td>1.0</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>Substrate [g L⁻¹]</td>
<td>1.5</td>
<td>0.19</td>
</tr>
<tr>
<td>AO-SODE (O₂, OUR)</td>
<td>Biomass [g L⁻¹]</td>
<td>0.9</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>Substrate [g L⁻¹]</td>
<td>1.4</td>
<td>0.18</td>
</tr>
<tr>
<td>AO-SODE (Sorbitol)</td>
<td>Biomass [g L⁻¹]</td>
<td>2.0</td>
<td>0.051</td>
</tr>
</tbody>
</table>

*Figure 6. RLS-VFF validation from CPR experimental data for the *P. pastoris PFLD1*-based system.*

(a) Biomass and substrate estimation; (b) specific rates estimation. Symbols correspond to process values, solid lines stand for estimation, and dashed lines obtained through substrate balance with off-line specific growth rate data.

*Figure 7. AO-SODE validation from experimental data for the *P. pastoris PFLD1*-based system.*

(a) Specific growth rate and biomass estimation using *S* measurements; (b) specific growth rate estimation and specific consumption rate using *O₂* and OTR; (c) biomass estimation and reconstructed substrate using *O₂* and OTR. Symbols correspond to real values, solid lines stand for estimation and dashed lines were obtained through substrate balance with off-line specific growth rate data.
In a future approach, the chosen estimation procedure will allow the implementation of a “true” specific growth rate controller. This would represent an improvement regarding those based either on preprogrammed exponential feeding or an indirect μ-control, keeping constant the substrate concentration.

Acknowledgments

This work was supported by a grant from the Spanish programme of Chemical Processos Technologies (CTQ2007-66347; CTQ2010-15131). HP2007-0045 from Science and Innovation Ministry, 2009-SGR-00281 and the Reference Network in Biotechnology (XRBI) from the DURSI (Generalitat de Catalunya). Support from “Lusso-Spanish Integrated Actions” (Action 100/08) are also acknowledged. JMB would like to thank the Spanish Education Ministry for his predoctoral fellowship.

Literature Cited


Manuscript received Nov. 12, 2010, and revision received Sept. 29, 2011.