

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 lie in 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 pPICZ α AROL, was used for heterologous expression of ROL under the control of the *AOX1* promoter.³⁴ *P. pastoris* X-33 containing the vector pPICZFLD α ROL 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 NH₄OH 30% (v/v) (batch) or 5 M KOH (fed-batch) for *PAOX1* and 5 M KOH for *PFLD1* (batch and fed-batch); dissolved oxygen 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 glycerol (3.5 L volume with an initial glycerol 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

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

Parameters and Coefficients	PAOXI (Mut ⁺)		PFLDI
	Batch	Fed-batch	Fed-batch
Substrate	Glycerol	Methanol	Sorbitol
μ_{\max} (h ⁻¹)	0.260	0.059	0.030
K_s (g·L ⁻¹)	0.20	0.22	–
$Y_{CO_2/X}$ (mol _{CO₂} g _X ⁻¹)	1.57×10^{-2}	1.02×10^{-1}	1.32×10^{-2}
$m_{CO_2/X}$ (mol _{CO₂} g _X ⁻¹ h ⁻¹)	2.6×10^{-4}	3.1×10^{-4}	6.5×10^{-4}
$Y_{O_2/X}$ (mol _{O₂} g _X ⁻¹)	2.4×10^{-2}	1.69×10^{-1}	2.80×10^{-2}
$m_{O_2/X}$ (mol _{O₂} g _X ⁻¹ h ⁻¹)	3.1×10^{-4}	4.7×10^{-4}	7.1×10^{-4}
$Y_{S/X}$ (g _S g _X ⁻¹)	1.97	4.29	0.47
$m_{S/X}$ (g _S g _X ⁻¹ h ⁻¹)	0.008	0.010	0.043
$K_L a$ (h ⁻¹)	–	360	–
$O_{2,sat}$ (mol _{O₂} L ⁻¹)	–	6.6×10^{-4}	–

source and production of the recombinant protein took place.³⁴ Cultivations using *P. pastoris* PFLDI-based systems proceeded as follows: after a batch phase on glycerol, a pulse of sorbitol and methylamine was added into the bioreactor. For the induction phase, 300 g·L⁻¹ of sorbitol, 30 g·L⁻¹ of methylamine chloride, and 7.5 mL of trace salts solution feeding medium was used in an exponential feeding rate strategy.³⁴

Glycerol, methanol, sorbitol, and methylamine were added in the transition and production phases by an automatic microburette MicroBU-2031 from Crison Instruments (Alella, Barcelona, 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 (Multor, Maihak, Hamburg, Ger-

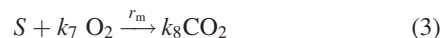
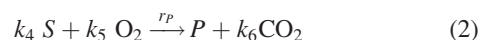
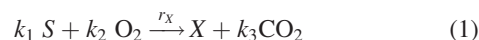
many). Carbon dioxide production rate (CPR) and oxygen uptake rate (OUR) were considered approximately equal to carbon dioxide transfer rate (CTR) and oxygen transfer rate (OTR), respectively, and, thus, obtained through mass balances of CO₂ and O₂ measured in the exhaust gas.

For the estimation procedure, the integration step and the sampling time were fixed at 0.055 h (200 s). Online gas measurements are a result of 20 averaged raw data points. So, the filtered signal was used as the input for the estimator.

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

The process model

For the oxidative assimilation of substrate (glycerol, methanol, or sorbitol), equations for growth (1), protein production (2), and maintenance (3) can be stated as



where X , S , O_2 , CO_2 , 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.

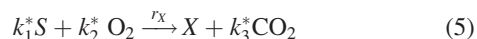
The corresponding dynamic model can be represented as follows

$$\frac{d}{dt} \begin{bmatrix} X \\ S \\ O_2 \\ CO_2 \\ P \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -k_1 & -k_4 & -1 \\ -k_2 & -k_5 & -k_7 \\ k_3 & k_6 & k_8 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ q_p \\ q_m \end{bmatrix} \cdot X - D \begin{bmatrix} X \\ S \\ O_2 \\ CO_2 \\ P \end{bmatrix} + \begin{bmatrix} 0 \\ D \cdot S_0 \\ OTR \\ -CTR \\ 0 \end{bmatrix} \quad (4)$$

where S_0 is the substrate concentration in the feed, OTR is the oxygen transfer rate from gas to liquid phase, and CTR is the carbon dioxide transfer rate from liquid to gas phase. $D = F_{in}/V$ is the dilution rate; F_{in} is the inlet volumetric flow rate, and V is the reactor volume ($dV/dt = F_{in}$).

μ , q_m , and q_p are biomass, product, and maintenance specific reaction rates, respectively.

Neglecting product formation in mass balance equations and considering a global reaction scheme for growth and maintenance, then, Eqs. 1–4 are transformed into Eqs. 5 and 6, where overall yield coefficients ($k_i^* = Y_{c/x}^*$) are used.



$$\frac{d}{dt} \begin{bmatrix} X \\ S \\ O_2 \\ CO_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -k_1^* \\ -k_2^* \\ k_3^* \end{bmatrix} \cdot \mu \cdot X - D \begin{bmatrix} X \\ S \\ O_2 \\ CO_2 \end{bmatrix} + \begin{bmatrix} 0 \\ D \cdot S_0 \\ OTR \\ -CTR \end{bmatrix} \quad (6)$$

Kinetic parameters and yield coefficients used for *P. pastoris* process under PAOXI (Mut⁺) and PFLDI for ROL production

were obtained from previous experiments in batch and fed-batch cultivations.^{34,37,38} The yield and maintenance coefficients in the kinetic models were identified as previously reported^{39,40} 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 estimators, for the determination of state variables and specific 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 estimate 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 concentration (X) and the specific growth rate (μ) through a component to biomass intrinsic yield ($Y_{c/x}$) and a maintenance

coefficient ($m_{c/x}$) described by the so-called Luedeking–Piret relationship as follows

$$\text{CKR} = Y_{c/x}\mu X + m_{c/x}X \approx Y_{c/x}^*\mu X \quad (7)$$

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 observer²⁷ represented by

$$\frac{d\hat{z}}{dt} = F(s, \hat{z})\hat{z} + G(u, s, \hat{z}) - \Lambda^{-1}(s, \hat{z})S_{\theta}^{-1}C^T(C\hat{z} - y) \quad (8)$$

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 by the Luedeking–Piret type model mentioned earlier. For instance, considering CPR as the CKR, Eqs. 9 and 10 are taken into account.

$$\text{CPR}(t) = Y_{\text{CO}_2/X}\mu(t)X(t) + m_{\text{CO}_2/X}X(t) \quad (9)$$

$$z_1 = \text{CPR} \quad (10)$$

Thus, the structure of the estimator is developed as follows⁴¹

$$\frac{d\hat{z}_1}{dt} = \hat{z}_2\hat{z}_1 - D\hat{z}_1 - 3\delta(\hat{z}_1 - z_1) \quad (11)$$

$$\frac{d\hat{z}_2}{dt} = \hat{z}_3 - 3\frac{\delta^2}{z_1}(\hat{z}_1 - z_1) \quad (12)$$

$$\frac{d\hat{z}_3}{dt} = -\frac{\delta^3}{z_1}(\hat{z}_1 - z_1) \quad (13)$$

Therefore, the μ estimates can be obtained from Eq. 14

$$\frac{d\hat{\mu}}{dt} = (\hat{z}_2 - \hat{\mu})\left(\hat{\mu} + \frac{m_{\text{CO}_2/X}}{Y_{\text{CO}_2/X}}\right) \quad (14)$$

In this approach, *a priori*, only a single tuning parameter δ 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

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

$$\frac{d\hat{\xi}}{dt} = KH(\hat{\xi})\rho(\hat{\xi}) - D\hat{\xi} + F - Q(\hat{\xi}) \quad (15)$$

where $\hat{\xi}$ is the state vector (the set of n component concentrations), K is a $(n \times m)$ matrix of known yield coefficients, D is the dilution rate, F is the feed rate vector with $\dim(F) = n$ and Q is the gaseous outflow rate vector which $\dim(Q) = n$. D , F , and Q are measured online. The

reaction rates are defined as $\varphi(\hat{\xi}) = H(\hat{\xi})\rho(\hat{\xi})$ to take advantage of any possible knowledge of the kinetic model, $H(\hat{\xi})$ being a $(m \times r)$ matrix of known functions of the state and $\rho(\hat{\xi})$ a vector of r unknown functions of state. The reaction rates can be estimated using the following the *AO* general structure²⁸

$$\frac{d\hat{\xi}}{dt} = KH(\hat{\xi})\hat{\rho} - D\hat{\xi} + F - Q - \Omega(\hat{\xi} - \hat{\xi}) \quad (16)$$

$$\frac{d\hat{\rho}}{dt} = [KH(\hat{\xi})]^T\Gamma(\hat{\xi} - \hat{\xi}) \quad (17)$$

As previously reported,²⁸ the stability considerations for the related linear time varying perturbing system are developed for *AO* with SODE. The first step is the selection of the r equations from the full state space model to be used in the estimator. For the specific growth rate estimation from the global reaction defined in Eq. 5, it is necessary to introduce a “ c ” one-dimensional subspace related with the measured variable (S , O_2 , or CO_2), its feeding rate or component transfer rate and its respective yield coefficient k_c^* .

Furthermore, the SODE algorithm usually uses a transformed state variable, defined as

$$\Psi = K_C^{-1} \cdot \xi_C = \frac{1}{\pm k_C^*} C \quad (18)$$

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

$$\frac{d\hat{\psi}}{dt} = H\hat{\rho} - D\hat{\psi} + K_C^{-1}(F_C - Q_C) - \Omega_C(\hat{\psi} - \hat{\psi}) \quad (19)$$

$$\frac{d\hat{\rho}}{dt} = H^T\Gamma_C(\hat{\psi} - \hat{\psi}) \quad (20)$$

where the gain matrices (Ω_C and Γ_C) and their parameters are defined as follows

$$\Gamma_C = \text{diag}\{-\gamma_i\}, \quad \Omega_C = \text{diag}\{-\omega_i\} \quad (21)$$

$$\gamma = \frac{2 \cdot \zeta}{\tau} - \frac{1}{\hat{X}} \frac{d\hat{X}}{dt}, \quad \omega = \frac{1}{\hat{X} \cdot \tau^2} \quad (22)$$

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)

$$\frac{d}{dt} \begin{bmatrix} X \\ C \end{bmatrix} = \begin{bmatrix} 1 \\ \pm k_C^* \end{bmatrix} \cdot \mu \cdot X - D \begin{bmatrix} X \\ C \end{bmatrix} \pm \begin{bmatrix} 0 \\ F_C - Q_C \end{bmatrix} \quad (23)$$

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

$$\hat{Z} = \hat{X} + \frac{1}{\pm k_C^*} C \quad (24)$$

$$\frac{d\hat{Z}}{dt} = -D\hat{Z} + \frac{1}{\pm k_C^*} (F_C - Q_C) \quad (25)$$

$$\hat{X} = \hat{Z} - \frac{1}{\pm k_C^*} C \quad (26)$$

As previously described,^{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,^{19,31} to use filtered data with setting bounds on μ ($0.05\mu_{\max} \leq \mu \leq \mu_{\max}$) to avoid unrealistic estimated values for the specific growth rate and $Y_{c/x}^*$ 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.^{42,43} In this way, some approaches have been developed to carry out the estimation from measurement of oxygen³¹ or carbon dioxide in the exhaust gas of the fermentor.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.

$$\frac{dCPR(t)}{dt} = \frac{d(\mu(t))}{dt} X(t) Y_{CO_2/X} + \frac{dX(t)}{dt} Y_{CO_2/X} \mu(t) + m_{CO_2/X} \frac{d(X(t))}{dt} \quad (27)$$

Then, considering the mass balance for biomass

$$\frac{dCPR(t)}{dt} = \frac{d(\mu(t))}{dt} X(t) Y_{CO_2/X} + (\mu(t) - D(t)) CPR(t) \quad (28)$$

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

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

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

$$CPR(t_{k+1}) = \hat{\theta}(t_k) CPR(t_k) + CPR(t_{k-1}) \quad (30)$$

From the previous estimate, the specific growth rate is obtained

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

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

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

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

$$K(t_k) = Q(t_k) \Psi(t_k) \quad (34)$$

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

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

where $P(t)$ is the covariance matrix, $Q(t)$ is an auxiliary matrix, $y(t)$ is the CPR(t), and Ψ 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.³² The calculation of the VFF is based on model error and both the data vector and covariance matrix.

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

\sum_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 μ estimation to be always positive and bounded ($\leq \mu_{\max}$).

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 μ , X , and S estimations. RMNS, ITAE, and RT were obtained for μ estimates.

The SSE can be defined as follows

$$SSE = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (38)$$

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 PAOX1 (Mut⁺) for ROL Production

Process Variables	Units	Batch	Fed-batch
Substrate (S)	[g·L ⁻¹]	Glycerol	Methanol
Initial concentration (S ₍₀₎)	[g·L ⁻¹]	40	0
Inlet concentration (S ₀)	[g·L ⁻¹]	–	790
Process time (t)	[h]	14.7	60.3
Initial volume (V ₍₀₎)	[L]	3.5	3.5
Initial Biomass (X ₍₀₎)	[g·L ⁻¹]	0.50	20
Specific growth rate set-point (μ _{set})	[h ⁻¹]	–	0.02
Initial specific growth rate (μ ₍₀₎)	[h ⁻¹]	0.00	0.00
Sampling time (T)	[h]	0.055	0.055

The RMSE is defined by

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (39)$$

that is, the average difference between estimated and actual target variables (specific growth rate, biomass, and substrate), where \hat{y}_i and y_i are the estimated and true value for the i th data point; n is the total number of data points. Alternatively,

$$\text{RMOPI} = \sqrt{w_\mu \text{RMSE}_\mu^2 + w_x \text{RMSE}_x^2 + w_s \text{RMSE}_s^2 + w_{ns} \text{RMNS}^2} \quad (43)$$

w_k is the weighting factors of the single metrics, selected to bring the terms within the same order of magnitude but with lower contribution for errors on S . $w_\mu = 0.40 \times 10^5$, $w_x = 0.19$, $w_s = 0.01$, $w_{ns} = 0.40 \times 10^5$.

$$\text{CMRE} = w_{c\mu} \text{MRE}_\mu + w_{cx} \text{MRE}_x \quad (44)$$

w_k is the weighting factors of the MRE for μ and X , $w_{c\mu} = 0.50$, $w_{cx} = 0.50$.

ITAE weight errors that exist after long time are much more heavy than those at the start of the estimation. Its evaluation is of interest in batch and fed-batch processes, where the need to have rapidly reliable estimates is a crucial issue,¹⁵ because it can be taken as an indicator for speed of convergence.

$$\text{ITAE} = \int t e \, dt \quad (45)$$

with e is the absolute error between μ estimations and actual values $|(\hat{\mu} - \mu)|$, t is the time of the process. ITAE is computed over the simulated period using the trapezium rule to approximate the integral.

Moreover, RT is defined at the time taken for the estimator response to first reach 95% of the change on specific growth rate. The lower RT is the faster response the estimation system has.

Another point that has to be noted is the different effect, which tuning parameters and estimator initialization values can be produced on the estimator's global performance. So, sensitivity of RMOPI and CMRE functions were examined through variations of their nominal values.

calculation of the MRE is also carried out to examine the “goodness-of-estimation.”

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (40)$$

The sum of squared noise sensitivity accounts for the continuous variation of μ estimations and it is defined as follows

$$\text{SSNS} = \sum_{i=1}^{n-1} (\hat{\mu}_{i+1} - \hat{\mu}_i)^2 \quad (41)$$

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.

$$\text{RMNS} = \sqrt{\frac{\sum_{i=1}^{n-1} (\hat{\mu}_{i+1} - \hat{\mu}_i)^2}{n-1}} \quad (42)$$

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.

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 $\text{OTR} = K_L a (\text{O}_{2,\text{sat}} - \text{O}_2)$, where $K_L a$ is the global mass-transfer coefficient and $\text{O}_{2,\text{sat}}$ is the saturation concentration of dissolved oxygen.⁴⁴

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

$$F_{\text{in}}(t) = \frac{\mu_{\text{set}} [X_{(0)} V_{(0)}]}{Y_{X/S}^* S_0} \exp[\mu_{\text{set}} [t - t_{(0)}]] \quad (46)$$

Thus, the substrate feeding rate depends on the specific growth rate set-point, $\mu_{\text{set}} = 0.02 \text{ h}^{-1}$; the initial volume, $V_{(0)}$; the

Table 3. Initial Guesses, Nominal Values, and Acceptable Ranges for Tuning Parameters and Initialization Estimator Variables

Algorithm	Parameters and Estimator Variables	Units	Initial Guesses	Nominal Value	Acceptable Range
<i>NLOBE</i>	δ	[h ⁻¹]	20	51	±1
	\hat{z}_{10}	[mol·L ⁻¹ ·h ⁻¹]	0.5	4.0	±0.3
	\hat{z}_{20}	[h ⁻¹]	5.0	2.0	±0.1
	\hat{z}_{30}	[h ⁻²]	50	100	±3
<i>AO-SODE</i> (Methanol)	ζ	[-]	1.0	0.70	±0.40
	τ	[h ⁻¹]	0.05	0.20	±0.11
<i>AO-SODE</i>	ζ	[-]	1.0	0.70	±0.40
	τ	[h ⁻¹]	0.05	0.20	±0.11
<i>RLS-VFF</i>	Σ_0	[h ⁻²]	5.0×10^{-8}	1.0×10^{-8}	$\pm 0.9 \times 10^{-8}$
	θ_0	[-]	2.0×10^{-3}	1.5×10^{-2}	$\pm 0.5 \times 10^{-2}$

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

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: $(X_{(0)}, \hat{\mu}_{(0)}, S_{(0)}) = (X_{(0)}, \mu_{(0)}, S_{(0)})$.

In general, tuning parameters and some initialization variables were determined empirically considering the convergence speed and noise sensitivity,¹⁹ 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 RMOPI indicator described before was used as cost criterion to be minimized. Depending on the sensitivity of this RMOPI 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 δ , \hat{z}_{10} , \hat{z}_{20} , and \hat{z}_{30} were selected according to the structure of the estimator: $\delta \approx 1/T$; $\hat{z}_{10} \approx \text{CPR}_{\max}$; $\hat{z}_{20} < ((\mu_{\max}/T)/\mu)$ for the maximal value of $d\mu/dt$; $\hat{z}_{30} = 10 \cdot \hat{z}_{20}$. Initial guesses and nominal values, obtained by RMOPI 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: ζ is the damping coefficient and τ is the time constant of a second-order dynamic response of estimated to the true values. The damping coefficient, ζ , was preliminary fixed at 1 according to a common engineering rule of thumb.²⁵ Tuning rules for discrete time implementation²⁸ were applied to determine preliminary value of τ according to $0 < T < 2\zeta\tau$, thus, setting τ to 0.05. Nominal values were finally obtained

by minimization of the cost criterion function RMOPI and they are shown in Table 3 for the two *AO-SODE* methods tested.

The tuning of the *RLS-VFF* estimator starts with the assignation of first trials to the tuning parameter Σ_0 and the initial value θ_0 .

Σ_0 has to be set to manage slow or moderate variations of μ by the VFF. So, Σ_0 initial settings were fixed to be the same order of magnitude that square deviations in Eq. 37. For the initial value θ_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 RMOPI minimization, for Σ_0 and θ_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 μ values. However, all μ and X estimations were conducted properly with deviations lower than 1% (MRE) and with similar values of the overall performance indexes (RMOPI 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 μ .

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 bioprocess approaches the end.

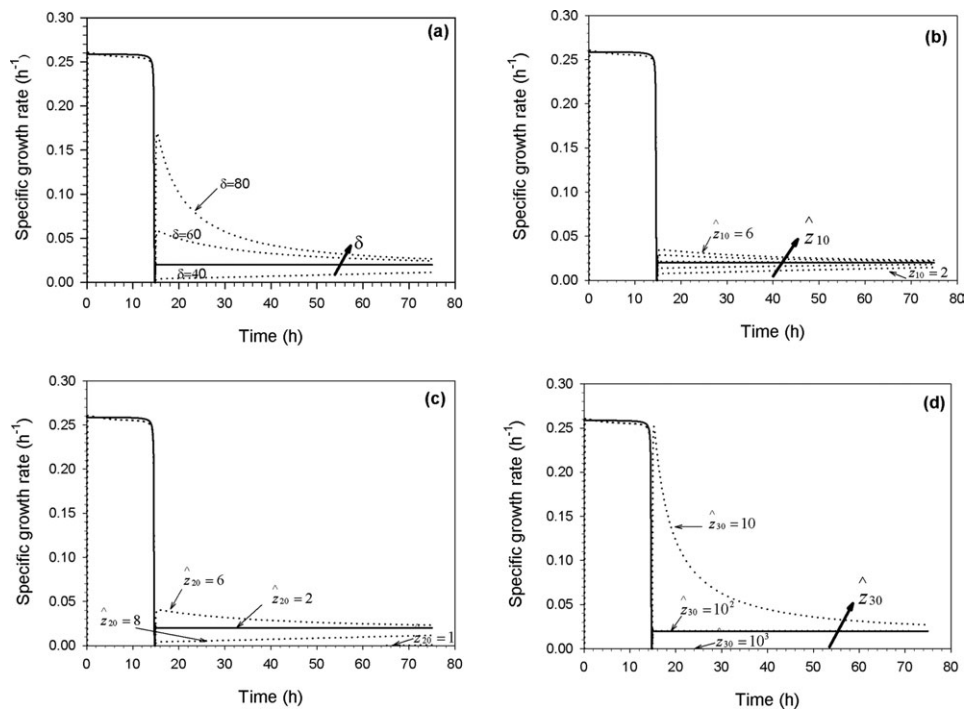


Figure 1. *NLOBE* estimator.

Effect of tuning and initial parameters on μ estimation: (a) δ , (b) \hat{z}_{10} , (c) \hat{z}_{20} , and (d) \hat{z}_{30} . Solid line stands for model data and dotted lines stands for estimation from CPR simulated data.

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 it 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 δ and the initial value of some variables on the estimation results. The effect of δ , \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 δ 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 δ variations as displayed in Figures 1 and 4. In spite of having only one tuning parameter, the extremely sensitivity of the RMOPI to δ 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 δ . 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 .⁴¹

Second, the effect of *AO-SODE* tuning parameters ζ and τ 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 τ , the response becomes faster and decreasing ζ 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 ζ and τ . These facts allow searching and selecting the tuning parameters with intuitive basis, simplifying the search for optimum.²⁸

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 τ values rather far from nominal values. This initial X offset remained quite stable through the bioprocess, whenever μ estimation was well-fitted to the simulated system response. Finally, to calculate the sensitivity of the tuning parameters ζ and τ , the RMOPI combined metrics was explored. In contrast to Figure 2, it can be observed in Figure 4 that the effect of ζ is not different than those of τ . This fact can be explained due to the tuning parameter τ in Figure 2 was tested in a wider range than ζ , 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 ζ and τ were very similar to those illustrated for the methanol case. No significant differences on X and μ 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 μ 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

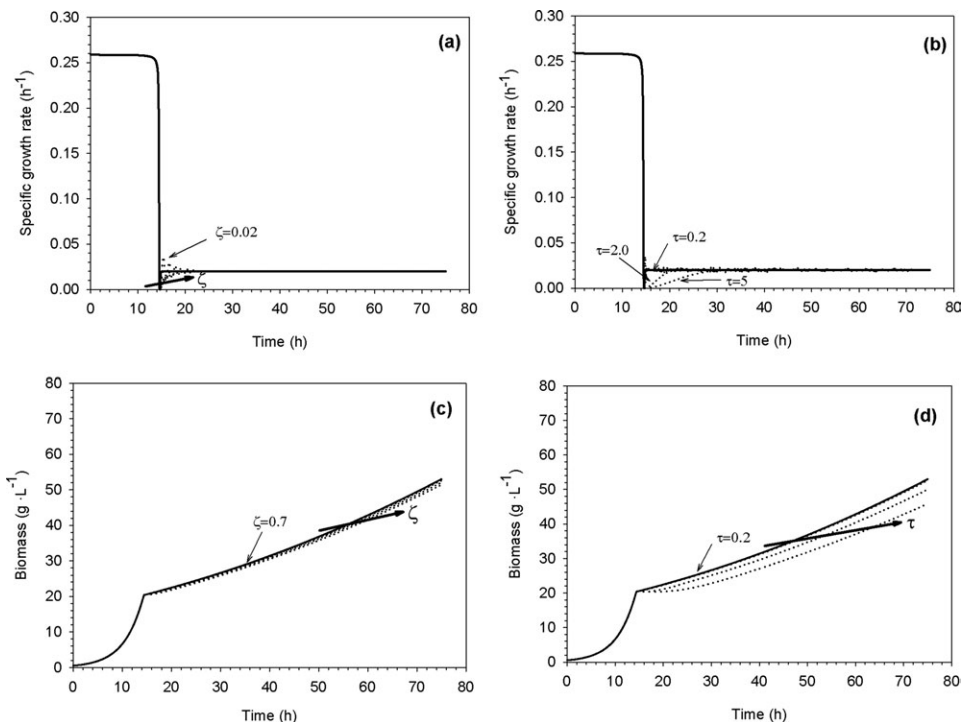


Figure 2. AO-SODE estimator.

Effect of tuning parameters on μ estimation from substrate measurements: (a) ζ , (b) τ ; and on reconstructed biomass: (c) ζ , (d) τ . Effect of tuning parameters on μ estimation from pO_2 and OTR: (e) ζ , (f) τ ; on reconstructed biomass: (g) ζ , (h) τ ; and on reconstructed substrate: (i) ζ , (j) τ . Solid line stands for model data and dotted lines stands for estimation from simulated data.

biomass concentration than in their ζ and τ 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 θ_0 and the tuning parameter Σ_0 . Their effect on μ , 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 θ_0 values than for Σ_0 . In Figures 3c,d, the tuning parameter Σ_0 for biomass calculations seems to have a stronger effect than the initial value of θ_0 due to the wide range explored for Σ_0 . From Figure 3, if a low Σ_0 value is introduced in the estimation of μ , 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 θ_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 varia-

bles, 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 ζ , τ , and Σ_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 (ζ , τ) and the far low sensitivity of CMRE to ζ , τ , and Σ_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 θ_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 δ , \hat{z}_{10} , \hat{z}_{20} , and \hat{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,

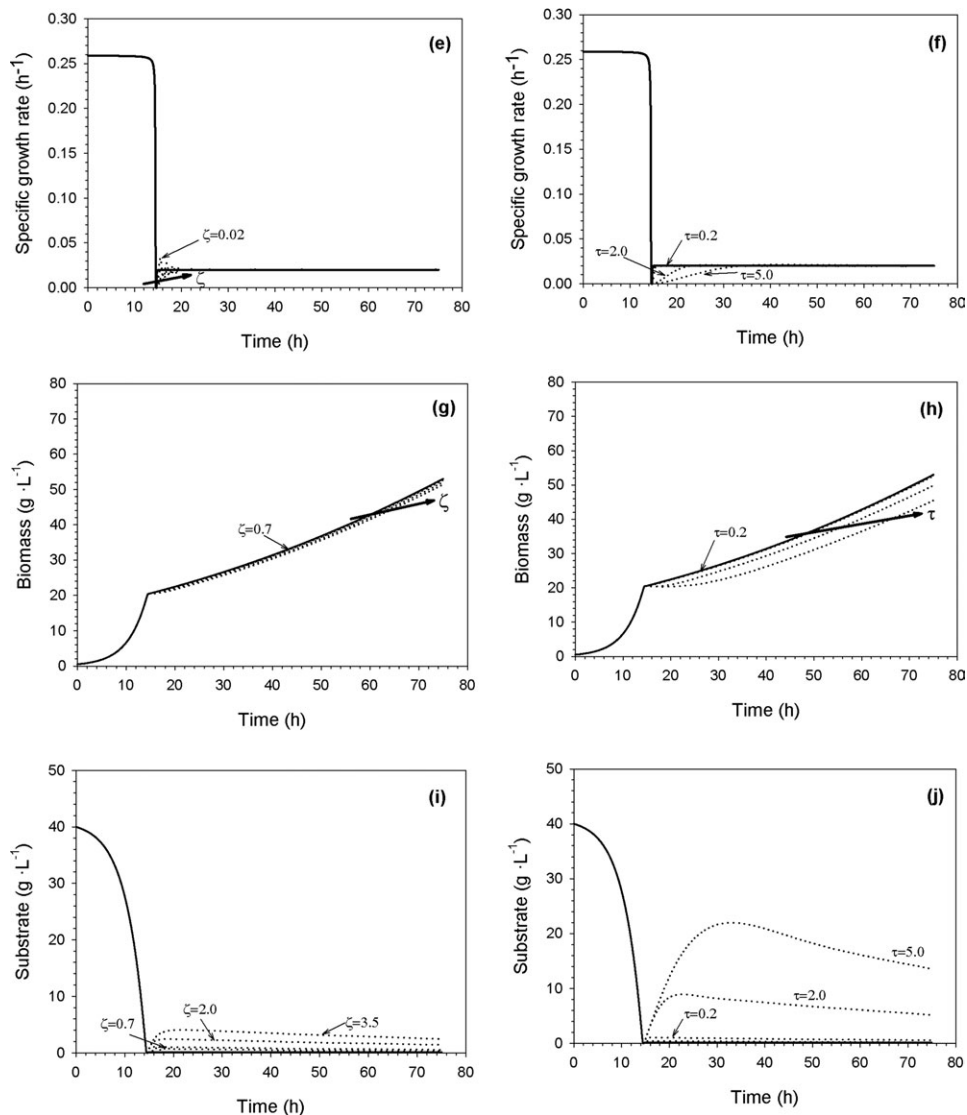


Figure 2. Continued

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 PAOXI*- and *PFLDI*-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 *PAOXI* 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, μ estimation and computed biomass estimation were carried out properly to the system

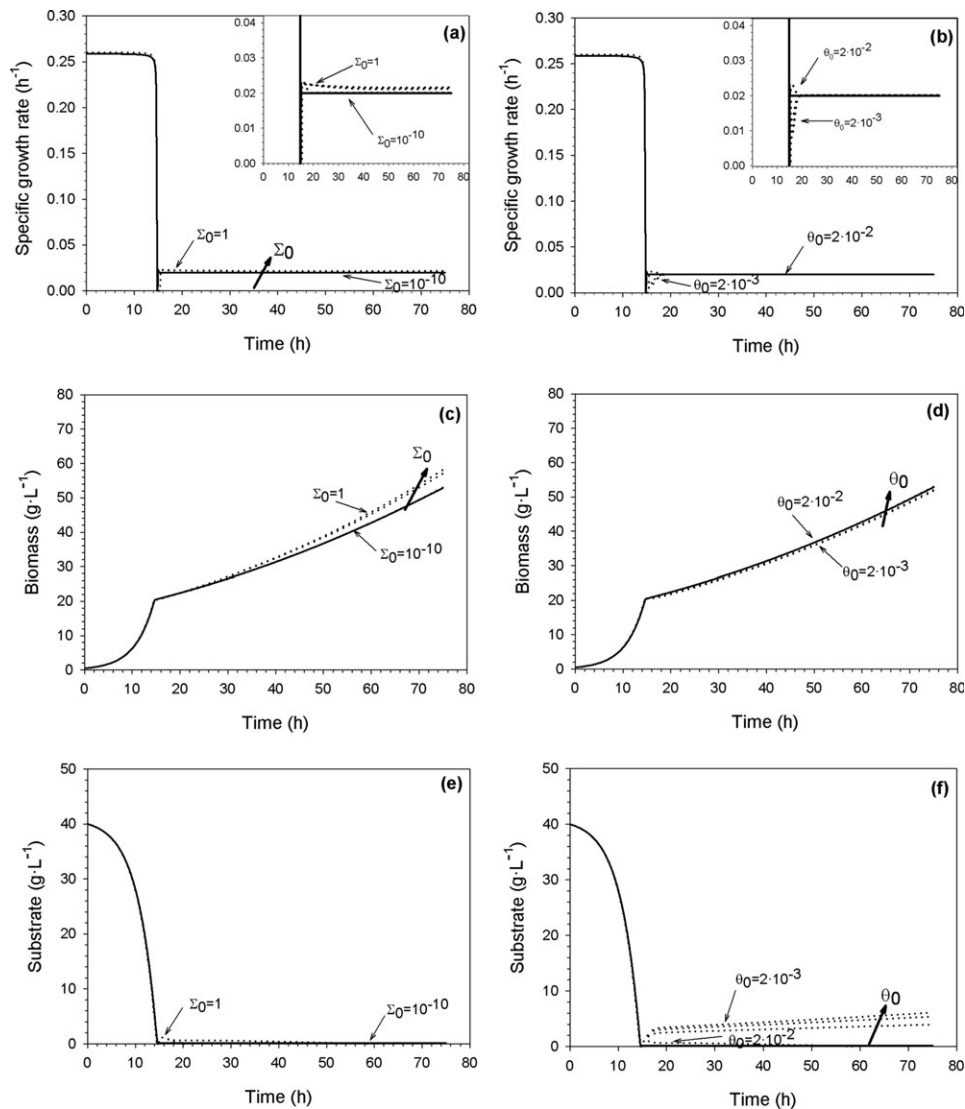


Figure 3. RLS-VFF estimator.

Effect of tuning parameters on μ estimation: (a) Σ_0 , (b) θ_0 ; on reconstructed biomass: (c) Σ_0 , (d) θ_0 ; and on reconstructed substrate: (e) Σ_0 , (f) θ_0 . Solid line stands for model data and dotted lines stands for estimation from CPR simulated data.

response and a rapid convergence was obtained for the μ estimation during the batch phase. When the batch phase was completed, as found at 20 h, unbiased μ 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 PFLDI*-based system was processed. In *PFLDI* 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

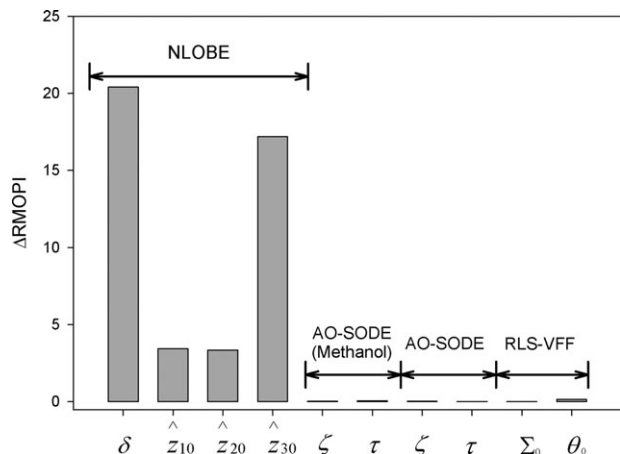


Figure 4. Absolute variation in the RMOPI of the estimation procedure on variation in the estimator parameters by 20%.

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 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 O_2 and OUR data or CO_2 and CPR measurements, similar results were observed. However, *AO-SODE* (O_2 , 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
 μ [h^{-1}]; X [$g \cdot L^{-1}$]; S [$g \cdot L^{-1}$]

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

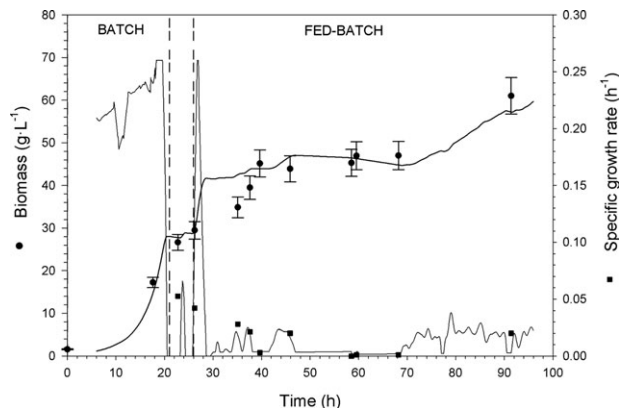


Figure 5. Specific growth rate and biomass estimation using CPR experimental data for the *P. pastoris PAOX1*-based system (Mut^+) obtained with the *RLS-VFF* method.

Symbols correspond to real values and lines stand for estimation.

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

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

Algorithm	Estimated Variable	RMSE	MRE
RLS-VFF (CPR)	Biomass [g·L ⁻¹]	1.2	0.036
	Substrate [g·L ⁻¹]	1.3	0.16
AO-SODE (CO ₂ , CPR)	Biomass [g·L ⁻¹]	1.0	0.026
	Substrate [g·L ⁻¹]	1.5	0.19
AO-SODE (O ₂ , OUR)	Biomass [g·L ⁻¹]	0.9	0.025
	Substrate [g·L ⁻¹]	1.4	0.18
AO-SODE (Sorbitol)	Biomass [g·L ⁻¹]	2.0	0.051

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.

*AO*s 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

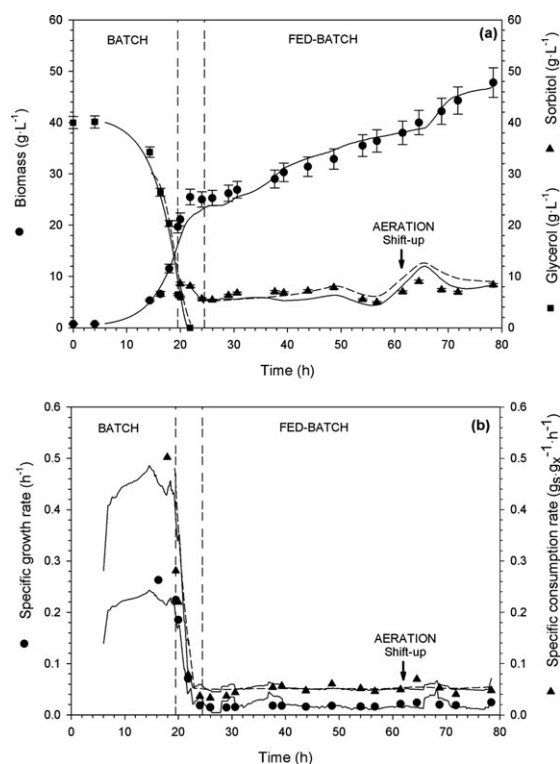


Figure 6. *RLS-VFF* validation from CPR experimental data for the *P. pastoris* PFLD1-based system.

(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.

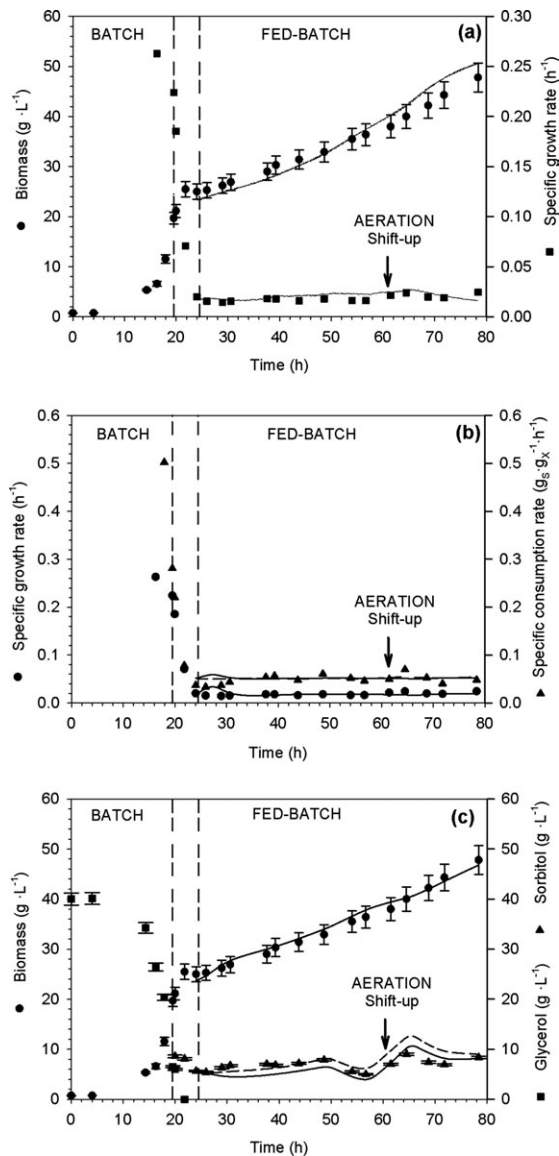


Figure 7. *AO-SODE* validation from experimental data for the *P. pastoris* PFLD1-based system.

(a) Specific growth rate and biomass estimation using *S* measurements; (b) specific growth rate estimation and specific consumption rate using *O*₂ and *OTR*; (c) biomass estimation and reconstructed substrate using *O*₂ 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.

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* (*O*₂, *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*.

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 Processes Technologies (CTQ2007-60347; CTQ2010-15131), HP2007-0045 from Science and Innovation Ministry, 2009-SGR-00281 and the Reference Network in Biotechnology (XRB) from the DURSI (Generalitat de Catalunya). Support from “Luso-Spanish Integrated Actions” (Action 100/08) are also acknowledged. JMB would like to thank the Spanish Education Ministry for his predoctoral fellowship.

Literature Cited

- Junker BH, Wang HY. Bioprocess monitoring and computer control: key roots of the current PAT initiative. *Biotechnol Bioeng.* 2006;95:226–261.
- Surribas A, Geissler D, Gierse A, Scheper Th, Hitzmann B, Montesinos JL, Valero F. State variables monitoring by in situ multi-wavelength fluorescence spectroscopy in heterologous protein production by *Pichia pastoris*. *J Biotechnol.* 2006;124:412–419.
- Becker Th, Hitzmann B, Muffler K, Poertner R, Reardon KF, Stahl F, Ulber R. Future aspects of bioprocess monitoring. *Adv Biochem Eng Biotechnol.* 2007;105:249–293.
- Amigo JM, Surribas A, Coello J, Montesinos JL, MasPOCH S, Valero F. On-line parallel factor analysis. A step forward in the monitoring of bioprocess in real time. *Chemom Intell Lab Syst.* 2008;92:44–52.
- Kiviharju K, Salonen K, Moilanen U, Eerokäinen T. Biomass measurement online: the performance of in situ measurements and software sensors. *J Ind Microbiol Biotechnol.* 2008;35:657–665.
- Madrid RE, Felice CJ. Microbial biomass estimation. *Crit Rev Biotechnol.* 2005;25:97–112.
- Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast.* 2005;22:249–270.
- Keck Graduate Institute. Heterologous proteins expressed in *Pichia pastoris*. Available at: <http://www.kgi.edu/faculty-and-research/profiles/james-m-cregg.html>, Accessed September 7, 2011.
- Lin Cereghino GP, Lin Cereghino J, Sunga AJ, Johnson MA, Lim M, Gleeson MAG, Cregg JM. New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. *Gene.* 2001;263:159–169.
- Cregg JM, Shen S, Johnson M, Waterham HR. *Classical genetic manipulation*. In: Higgins DR, Cregg JM, editors. *Methods in Molecular Biology*. Pichia protocols. Totowa, NJ: Humana Press, 1998:17–26.
- Shen S, Sulter G, Jeffries T, Cregg JM. A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. *Gene.* 1998;216:93–102.
- Harder W, Veenhuis M. *Metabolism of one-carbon compounds*. In: Rose AH, Harrison JS, editors. *The Yeasts*. London: Academic Press, 1989:289–316; vol. 3.
- Cos O, Ramon R, Montesinos JL, Valero F. Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microbial Cell Factories.* 2006;5:17.
- James SC, Legge RL, Budman, H. On-line estimation in bioreactors: A review. *Rev Chem Eng.* 2000;16:311–340.
- Dochain D. State and parameter estimation in chemical and biochemical processes: a tutorial. *J Proc Control.* 2003;13:801–818.
- Sunström H, Enfors SO. Software sensors for fermentation processes. *Bioprocess Biosyst Eng.* 2008;31:145–152.
- Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. *Comput Chem Eng.* 2009;33:795–814.
- Jensch M, Simutis R, Eisbrenner G, Stückrath I, Lübbert A. Estimation of biomass concentrations in fermentation processes for recombinant protein production. *Bioprocess Biosyst Eng.* 2006;29:19–27.
- Van Impe JF, Claes JE. On-line estimation of the specific growth rate based on viable biomass measurements: experimental validation. *Bioprocess Eng.* 1999;21:389–395.
- Valero F, Lafuente J, Poch M, Solà C. Biomass estimation using on-line glucose monitoring by FIA: application to *Candida rugosa* batch growth. *Appl Biochem Biotechnol.* 1990;24/25:591–602.
- Arndt M, Kleist S, Miksch G, Friehs K, Flaschel E, Trierweiler J, Hitzmann BA. Feedforward–feedback substrate controller based on a Kalman filter for a fed-batch cultivation of *Escherichia coli* producing phytase. *Comput Chem Eng.* 2005;29:1113–1120.
- Montesinos JL, Lafuente J, Gordillo MA, Valero F, Solà C, Charbonnier S, Cheruy A. Structured modeling and state estimation in a fermentation process: lipase production by *Candida rugosa*. *Biotechnol Bioeng.* 1995;48:573–584.
- Loeblein C, Perkins JD. Structural design for on-line process optimization. I. Dynamic economics of MPC. *AIChE J.* 1999;45:1018–1029.
- Nadri M, Trezzani I, Hammouri H, Dhurjati P, Longin P, Lieto J. Modeling and observer design for recombinant *Escherichia coli* strain. *Bioprocess Biosyst Eng.* 2006;28:217–225.
- Bastin G, Dochain C. *On-line estimation and adaptive control of bioreactors*. Amsterdam: Elsevier science publishers, B.V., 1990.
- Lubanova V, Rocha I, Ferreira EC. Estimation of multiple biomass growth rates and biomass concentration in a class of bioprocesses. *Bioprocess Biosyst Eng.* 2003;25:395–406.
- Farza M, Hammouri H, Jallut C, Lieto, J. State observation of a nonlinear system: application to (bio)chemical processes. *AIChE J.* 1999;45:93–106.
- Oliveira R, Ferreira EC, Feyer de Azevedo S. Stability, dynamics of convergence and tuning of observer-based kinetics estimators. *J Process Control.* 2002;12:311–323.
- Veloso ACA, Rocha I, Ferreira EC. Monitoring of fed-batch *E. coli* fermentations with software sensors. *Bioprocess Biosyst Eng.* 2009; 32:381–388.
- Pomerleau Y, Perrier M. Estimation of multiple specific growth rates: experimental validation. *AIChE J.* 1992;38:1751–1760.
- Estler MU. Recursive on-line estimation of the specific growth rate from off-gas analysis for the adaptive control of fed-batch processes. *Bioprocess Eng.* 1995;12:205–207.
- Golobic I, Gjerkes H, Malensek J. On-line estimation of the specific growth rate in the bacitracin fermentation process. *AIChE J.* 1999;45:2550–2556.
- Ungarala S, Co TB. Model parameter tracking in microbial growth processes. *AIChE J.* 1998;44:2129–2134.
- Cos O, Resina D, Ferrer P, Montesinos JL, Valero F. Heterologous production of *Rhizopus oryzae* lipase in *Pichia pastoris* using the alcohol oxidase and formaldehyde dehydrogenase promoters in batch and fed-batch cultures. *Biochem Eng J.* 2005;26:86–94.
- Ramon R, Feliu JX, Cos O, Montesinos JL, Berthet FX, Valero F. Improving the monitoring of methanol concentration during high cell density fermentation of *Pichia pastoris*. *Biotechnol Lett.* 2004;26:1447–1452.
- Horstkotte B, Arnau C, Valero F, Elsholz O, Cerdà V. Monitoring of sorbitol in *Pichia pastoris* cultivation applying sequential injection analysis. *Biochem Eng J.* 2008;42:77–83.
- Cos O, Serrano A, Montesinos JL, Ferrer P, Cregg JM, Valero F. Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures. *J Biotechnol.* 2005;116:321–335.
- Cos O. Monitorització i control del procés de producció de proteïnes heteròlogues en el llevat metilotròfic *Pichia pastoris*, PhD Thesis, Departament d'Enginyeria Química, ETSE, Universitat Autònoma de Barcelona, 2005.
- Petkov SB, Davis RA. On-line biomass estimation using a modified oxygen utilization rate. *Bioprocess Eng.* 1996;15:43–45.
- Surribas A, Montesinos JL, Valero F. Biomass estimation using fluorescence measurements in *Pichia pastoris* bioprocess. *J Chem Technol Biotechnol.* 2006;81:23–28.
- Farza M, Nadri M, Hammouri H. Nonlinear observation of specific growth rate in aerobic fermentation processes. *Bioprocess Eng.* 2000;23:359–366.
- Roux G, Dahhou B, Queinnec I. Modelling and estimation aspects of adaptive predictive control in a fermentation process. *Control Eng Practice.* 1996;4:55–66.
- Carrier JF, Stephanopoulos G. Wavelet-based modulation in control-relevant process identification. *AIChE J.* 1998;44:341–360.
- Pérez J, Montesinos JL, Gòdia F. Gas–liquid mass transfer in an upflow cocurrent packed-bed biofilm reactor. *Biochem Eng J.* 2006;31:188–196.

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