ISA Transactions 50 (2011) 303–314

Contents lists available at ScienceDirect

ISA Transactions

journal homepage: www.elsevier.com/locate/isatrans

Comparison of different estimation techniques for biomass concentration in large scale yeast fermentation

A. Hocalar [a,*], M. Türker [a], C. Karakuzu [b], U. Yüzgeç [c]

[a] Pakmaya, P.K. 149, 41001 Izmit, Kocaeli, Turkey

[b] Department of Computer Engineering, Bilecik University, Bilecik, Turkey

[c] Department of Electrical and Electronics Engineering, Bilecik University, Bilecik, Turkey

ARTICLE INFO

Article history:
Received 10 July 2009
Received in revised form
11 August 2010
Accepted 5 December 2010
Available online 22 January 2011

Keywords:
State estimation
Asymptotic observer
Artificial neural network
Differential evolution
Biocalorimetry and baker's yeast

ABSTRACT

In this study, previously developed five different state estimation methods are examined and compared for estimation of biomass concentrations at a production scale fed-batch bioprocess. These methods are i. estimation based on kinetic model of overflow metabolism; ii. estimation based on metabolic black-box model; iii. estimation based on observer; iv. estimation based on artificial neural network; v. estimation based on differential evaluation. Biomass concentrations are estimated from available measurements and compared with experimental data obtained from large scale fermentations. The advantages and disadvantages of the presented techniques are discussed with regard to accuracy, reproducibility, number of primary measurements required and adaptation to different working conditions. Among the various techniques, the metabolic black-box method seems to have advantages although the number of measurements required is more than that for the other methods. However, the required extra measurements are based on commonly employed instruments in an industrial environment. This method is used for developing a model based control of fed-batch yeast fermentations.

1. Introduction

The monitoring of dissolved oxygen (DO), pH, temperature and other easily measured variables provides information at minimum level for the monitoring and control of bioprocesses. Complex and high technology based (e.g. optical density, capacitance, ultrasonic, NIR and flow injection analysis) sensors have been used for the main state variables such as biomass, substrate or metabolite concentrations while state estimation methods based on software sensors are still popular, especially in large scale applications due to the capacity and scope of their applications. The unmeasured state variables mentioned above are generally determined by offline analysis that is costly, time-consuming and contains great information lags [1].

Over past decades, *in situ* optical density sensors, capacitance and conductance based measurements, ultrasonic measurements, NIR or MIR spectroscopy and flow injection analysis based on sample withdrawal and measurement have become commercially available. However, few sensors fulfill the requirement of interferences and are sufficiently reliable for use at high biomass concentrations and viscosity values in an industrial environment [2–4]. *In situ* biomass concentration measurements have disadvantages in industrial fermentations, especially in high biomass concentrations and complex substrate containing applications. Therefore, indirect methods using software sensors have been developed in order to overcome the drawbacks of *in situ* measurements. Many techniques have been developed over the last few decades for online estimation in biotechnological processes. Soft sensors are based on mathematical algorithms that can produce estimates of unmeasured variables from measured variables [1,5].

The application of software sensors to biotechnological processes has been discussed in many papers [1,6–9]. They can be classified according to the models used for their estimation:

- Mechanistic model based estimators.
- Black-box model based estimators.
- Hybrid model based estimators.

The mechanistic models are based on the first principle equations such as energy or mass balances. The black-box models, which are not directly related to the physics of the process, use empirical models for estimation. The empirical coefficients of the model are adjusted according to the process experimental data. The hybrid models on the other hand use a combination of mechanistic terms together with an empirical model [1].

Several mechanistic model based estimation studies have been published for example, baker's yeast [10–14] and penicillin [15,16]. Several papers have been also published on black-box

* Corresponding author. Tel.: +90 262 373 24 00, +90 262 321 44 52; fax: +90 262 321 81 17.

E-mail addresses: akifh@pakmaya.com.tr (A. Hocalar), mustafat@pakmaya.com.tr (M. Türker), cihan.karakuzu@bilecik.edu.tr (C. Karakuzu), ugur.yuzgec@bilecik.edu.tr (U. Yüzgeç).

0019-0578/$ – see front matter © 2010 ISA. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.isatra.2010.12.003

Nomenclature

A	Area (m^2)
C_i	Concentration of i (kg/m^3)
D	Dilution rate (1/h)
DO	Dissolved oxygen concentration
E_m	Measurable part of elemental composition matrices
E_c	Calculable part of elemental composition matrices
F_i	Flow rate of i (m^3/h)
K	Yield coefficient matrices
M_i	Molar weight of i (kg)
q_i	Specific conversion rates of i (kg/kg h, C mol/C mol h)
r_i	Rate of consumption or production of i (C-mol/m^3 h)
r_m	Measurable part of conversion rate vector
r_c	Calculable part of conversion rate vector
S_o	Substrate concentration (kg/m^3)
X	Biomass (kg/m^3)
V	Volume (m^3)
$Y_{i/j}$	Yield of i over j
T	Temperature (°C)

Subscripts

ox	Oxidative
red	Reductive
eth	Ethanol
m	Maintenance
pr	Production
ae	Aerobic
in	Inlet
out	Outlet
T	Transpose
n	Nitrogen
o	Oxygen
e	Ethanol
s	Substrate
p	Product
x	Biomass
w	Water
exch	Exchanger

Greek Letters

μ	Specific growth rate, (h^{-1})
ξ	State variable
α	Weight factor
β	Weight factor

model estimators [17–19] and hybrid model based estimators [20,21]. The black-box estimation techniques use appropriate mathematical techniques to extract process behaviour from input–output data. Neural network, linear ARMA, nonlinear NARMA and Volterra modelling are commonly used in models based on black-box estimation. Neural estimators use online primary measurements to estimate process states. After online data are collected, offline data must be found so that the NN can be trained. Many works can be found in the literature for the purpose of estimating biomass concentration [15,17,22,23]. State observers are interesting tools for the estimation of process states in bioprocess monitoring and control. Indirect measurement of process states that eliminating the unknown kinetic parameters can be possible by means of observers. Classical observers like Luenberger and Kalman observers are based on exact knowledge of the model structure. Asymptotic observers are based on the idea that the uncertainty in the process models lies in the process kinetic models. The design of these asymptotic observers is based on the mass or energy balances without requiring knowledge of process kinetics [24–26].

In recent years, the stochastic search techniques such as genetic algorithm, immune genetic algorithm and differential evolution have gained popularity due to their easy searching abilities, probabilistic natures and robustness. The Differential Evolution (DE) developed by [27] is an improved version of the genetic algorithm technique. In general, a simple genetic algorithm uses binary coding for representing problem parameters, whereas DE uses real coding of floating point numbers. DE's algorithms have been implemented in many areas such as digital filter design [28], neural network learning [29], parameter estimation [30] and optimization of chemical process [31,32] due to its advantages (speed, robustness, ease of use and simple structure).

In this study, the results of five different biomass estimation methods are compared on the basis of the estimation of biomass concentration in industrial fed-batch yeast fermentation process. These methods are used in different model based state estimation and control studies individually in our group at different times [17,33–36]. The proposed methods are: (1) estimation based on kinetic model of overflow metabolism; (2) metabolic black-box model based estimation; (3) asymptotic observer based estimation; (4) artificial neural network based estimation and (5) differential evolution based estimation.

The same data sets are used to compare performances of estimation methods. The obtained estimation results and their convergence to offline measurements are presented. The advantages and disadvantages of estimation methods are discussed based on the number of measurements required, ease of implementation and required knowledge about process.

2. Basic metabolism of *Saccharomyces cerevisiae*

Yeast *S. cerevisiae* are used for the production of baker's yeast, ethanol and as a host organism for producing recombinant proteins (e.g. for production of insulin for diabetics, vaccines, etc.). Metabolic behaviour of the baker's yeast fermentation process can be defined by three different pathways: oxidative growth on carbon source, fermentative growth on carbon source and oxidative growth on ethanol [37,38]. The overall metabolism of these three pathways can be shown by the following equation.

$$r_s CH_2ON_{0.0125} + r_n NH_3 + r_o O_2 \rightarrow r_x CH_{1.8}O_{0.5}N_{0.2} + r_p CH_3O_{0.5} + r_c CO_2 + r_w H_2O + r_q. \quad (1)$$

Yeast cells consume the carbon substrate as a source of energy and biosynthetic reactions. Depending on the concentration of substrate in the medium, yeast cells can follow two different pathways: microbial growth or product formation. Product formation can only be induced in excess glucose concentrations or in the absence of sufficient oxygen.

3. Materials and method

The experimental data have been obtained from two different technical scale bubble column fermenters, one with 25 m^3 volumes (type D) and the other with 100 m^3 volumes (type A, B, C). The reactors are operated according to the predetermined substrate (molasses, ammonia and airflow) profiles, temperature and pH profiles (Figs. 1 and 2). The flows entering the fermenters are substrate (molasses), ammonia, air, acid and anti-foam agent. The carbon dioxide, oxygen (Servomex, 1400 B4 SPX) and ethanol (Vogelbuch GS 2/3) concentrations are measured in gas phase in the exhaust line. Molasses and ammonia flow rates are measured by electromagnetic flow meters (Krohne, IFM 090) and airflow is measured by a vortex type (EMCO, V-Bar 700) flow meter. The temperature and pH of the broth are controlled at 30 °C, pH

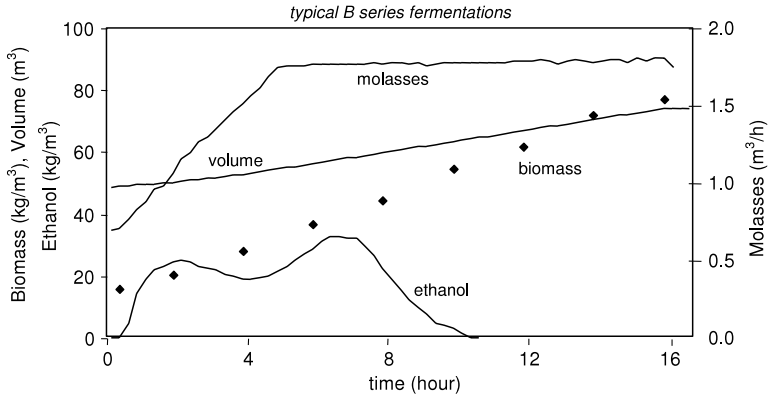

Fig. 1. Progress of typical fermentation obtained in the 100 m^3 technical scale reactor.

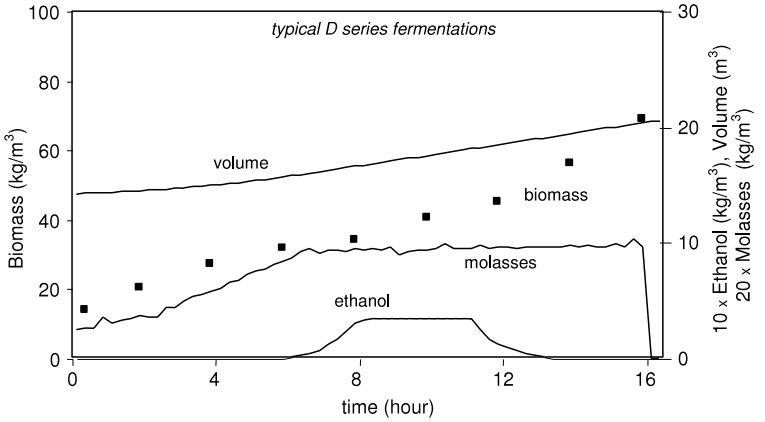

Fig. 2. Progress of a typical fermentation obtained in the 25 m^3 pilot reactor.

4–7 optional according to the predetermined profiles by a closed loop controller in PLC (Modicon, Micro CPU 612 03). In order to remove heat generated during fermentation, the 100 m^3 reactor is equipped with an external plate heat exchanger while the other 25 m^3 reactor has cooling coils. The gravimetric method is used for the reference of biomass concentration measurements, in which the biomass concentration is quantified by its dry weight.

Four different types of fermentations were selected in order to evaluate the performances of the estimation techniques. The differences in the initial conditions and operating conditions arise from the restrictive marketing conditions for the final products. The following fermentations are carried out for testing the performances of different estimation techniques:

- Type A fermentations are designed and operated to produce minimum ethanol (≈%0) during fermentation at 100 m^3 reactor.
- Type B fermentations are designed and operated to produce and then consume a small amount of ethanol (≈%0.2) during the first hours of fermentation at 100 m^3 reactor.
- Type C fermentations are designed and operated to produce some ethanol (≈%1) during the exponential phase of fermentation at 100 m^3 reactor.
- Type D fermentations are carried out at 25 m^3 reactor and operated according to produce and then consume a small amount of ethanol (≈% 0.2) during the first hours of fermentation.

The measurements of the substrate feeding, biomass and ethanol concentration are given in Figs. 1 and 2 respectively.

The substrate flows are controlled via the PID controllers. The PID loops are managed by PLC (Modicon, Micro CPU 612 03) and process measurements are collected via the SCADA system (Nematron, Paragon V5.4). All the primary measurements are collected in a one-minute sampling period. Ten minutes periods are used in the calculation of conversion rates and the determination of biomass concentration. The primary measurements and calculated conversion rates are given in Table 1.

4. Development of estimation methods

4.1. Estimation based on kinetic model of overflow metabolism

The kinetic model of overflow metabolism is based on the bottleneck model developed by Sonnleitner and Kappeli [39]. Depending on the exact model, the limited oxygen capacity or excessive concentrations of sugar result in ethanol formation in addition to some biomass. The sugar uptake rates (q_s) are calculated from online molasses feed rate measurements assuming the process is in a quasi-steady state. The total sugar uptake rate is calculated as below:

$$q_{s,tot} = \frac{(FS_o)/M_s}{XV_f}. \tag{2}$$

The specific reductive sugar uptake rate ($q_{s,red}$) is calculated using an online ethanol concentration measurement as follows:

$$q_{s,red} = \frac{(\Delta C_e)/M_E Y_{E/S}}{XV_f}. \tag{3}$$

Here, ΔC_e shows the difference in the ethanol concentration measurements. The specific ethanol consumption/production ($q_{e,ox}/q_{e,pr}$) rates are calculated using online ethanol measurements in liquid phase:

$$\begin{aligned} q_{e,pr} &= q_{s,red} Y_{E/S} \\ q_{e,ox} &= \Delta C_e / X. \end{aligned} \tag{4}$$

Table 1
Primary measurements and conversion rates.

Primary measurements	Units	Conversion rates	Units	Symbol
Air flow	N m^3/h	O_2 consumption rate	mol/m^3 h	r_o
Substrate flow rate	m^3/h	CO_2 production rate	mol/m^3 h	r_c
O_2 concentration	%	Ethanol prod./cons. rate	C mol/m^3 h	r_e
CO_2 concentration	%	Substrate consumption rate	C mol/m^3 h	r_s
Ethanol concentration	%	Metabolic heat prod. rate	kJ/m^3 h	r_q
Liquid volume	m^3	Water production rate	mol/m^3 h	r_w
Cooling water flow rate	m^3/h	Nitrogen consumption rate	mol/m^3 h	r_n
Cooling water inlet temp.	°C	Biomass production rate	C mol/m^3 h	r_x
Cooling water outlet temp.	°C			
Molasses temperature	°C			
Air temperature	°C			
Fermenter temperature	°C			

The oxidative sugar uptake rate, $q_{s,ox}$, is calculated from the difference between the total sugar consumption rate and substrate utilized for ethanol formation as:

$$q_{s,ox} = q_{s,tot} - q_{s,red}. \tag{5}$$

Total specific growth rate is obtained by adding individual specific substrate uptake rates multiplied by yield coefficients:

$$\mu = Y^{ox}_{X/S} q_{s,ox} + Y^{red}_{X/S} q_{s,red} + Y^{ox}_{O/E} q_{e,ox}. \tag{6}$$

The biomass concentration is estimated by the following mass balance equation:

$$\frac{dX}{dt} = \mu X - \frac{(FX)}{V}. \tag{7}$$

In this method, glucose substrate concentration is assumed to be zero and measured substrate flow rate is converted to the specific substrate uptake rates at quasi-steady state conditions. The specific ethanol production or consumption rates are calculated by online ethanol concentration measurements.

4.2. Estimation based on metabolic black-box model

In this approach, black-box process description is used and the cell biomass is considered as a black-box exchanging material with the environment [40]. All the relevant compounds and their elemental compositions must be known in order to define the conservation relations [41]. The rate vector for general stoichiometry of yeast fermentation can be written as:

$$r = (r_x, r_p, r_c, r_w, -r_s, -r_o, -r_n, r_q)^T. \tag{8}$$

The elemental composition matrix including energy balance for fermentation process can be written as in Box I. According to general conservation law multiplying the elemental composition matrices and conversion rate vector is zero, thus:

$$Er = 0. \tag{10}$$

The conversion rate vector can be divided into measurable (r_m) and non-measurable (r_c) conversion rates. Similarly, the elemental composition matrix is also divided into measurable (E_m) and non-measurable (E_c) as

$$E_m r_m + E_c r_c = 0. \tag{11}$$

If the number of measurable rates is equal to the degree of freedom (F), E_c matrix is in quadratic form and easily inverted as:

$$r_c = -E_c^{-1} E_m r_m. \tag{12}$$

In the case E_c matrix is not in quadratic form, the matrix inversion is calculated by Eq. (16) (pseudo inverse):

$$r_c = -(E_c^T E_c)^{-1} E_c^T E_m r_m. \tag{13}$$

The measurable rates (r_m) can be improved using relative errors, redundant measurements and reconciliation algorithms given in [33,41,42].

The specific conversion rates used in the estimation models are derived from the primary measurements. The primary measurements are converted to appropriate units (mol/L, C mol/L) and then applied to the estimation models. In this study, the O_2 consumption rate (r_o), CO_2 production rate (r_c), ethanol production/consumption rate (r_e), substrate consumption rate (r_s) and metabolic heat production rate (r_q) are used as measured conversion rates.

The oxygen uptake rate and carbon dioxide production rates are determined as below using inert gas balance:

$$r_o = \frac{F_n}{V}\left[\frac{P^{in}_{O_2}}{1 - P^{in}_{O_2} - P^{in}_{CO_2} - P^{in}_W} - \frac{P^{out}_{O_2}}{1 - P^{out}_{O_2} - P^{out}_{CO_2} - P^{out}_W}\right] \tag{14}$$

$$r_c = \frac{F_n}{V}\left[\frac{P^{in}_{CO_2}}{1 - P^{in}_{O_2} - P^{in}_{CO_2} - P^{in}_W} - \frac{P^{out}_{CO_2}}{1 - P^{out}_{O_2} - P^{out}_{CO_2} - P^{out}_W}\right]. \tag{15}$$

Substrate consumption rate, r_s, is determined at a quasi-steady state by measuring the molasses feed rate, thus:

$$r_s = \frac{FS_o}{V_L}. \tag{16}$$

Ethanol production/consumption rate, r_e, is determined by online measuring alcohol concentration of the broth, thus:

$$r_e = \frac{\Delta E}{V_L}. \tag{17}$$

Metabolic heat production rate is determined using dynamic energy balance around the fermenters as follows:

$$\rho C_{p,b}\frac{d(VT)}{dt} = q_{\text{feed}} + q_{\text{metabolic}} + q_{\text{exchanger}} + q_{\text{surface}} + q_{\text{evaporation}} + q_{\text{radiation}} + q_{CO_2} + q_{\text{acid}}. \tag{18}$$

The contribution of each term apart from $q_{\text{metabolic}}$ in Eq. (18) is computed from online primary measurements [37,38]. Metabolic heat production rate is calculated when the left hand side of Eq. (18) is set to zero since the temperature of the broth is controlled at a set value.

Once all the rates other than r_x, r_n and r_w are calculated from the primary measurements discussed above, the r_x, r_n and r_w are computed using Eq. (13). Then the biomass concentration is estimated by integrating volumetric biomass production rate r_x.

As seen above, there are many primary measurements (ethanol concentration, exhaust gas, substrate flow, cooling water flow rate, cooling water in–out temperature, substrate feed rate and temperature) in order to setup estimation structure in industrial environment, but they are reliable, cheap and easy to implement.

4.3. Estimation based on observer

The general state space dynamical model can be written for biotechnological processes as follows [43]:

$$E = \begin{array}{c} \begin{matrix} \text{biomass} & \text{ethanol} & CO_2 & H_2O & \text{glucose} & \text{oxygen} & \text{ammonia} & \text{heat} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1.83 & 3 & 0 & 2 & 2 & 0 & 3 & 0 \\ 0.56 & 0.5 & 2 & 1 & 1 & 2 & 0 & 0 \\ 0.17 & 0 & 0 & 0 & 0.0125 & 0 & 1 & 0 \\ 560 & 683 & 0 & 0 & 467 & 0 & 383 & -1 \end{bmatrix} \end{array} \begin{array}{l} \text{carbon balance} \\ \text{hydrogen balance} \\ \text{oxygen balance} \\ \text{ammonia balance} \\ \text{energy balance} \end{array} \quad (9)$$

Box I.

$$\frac{d\xi}{dt} = K\varphi(\xi) - D\xi + F - Q \quad (19)$$

where ξ, Q, F involves n components, φ involves m reaction rates and K is $(N \times M)$ size yield coefficient matrix. The general dynamical model equations for the fed-batch fermentation can be written based on the mass balance equations:

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -Y^{ox}_{X/S} & -Y^{red}_{X/S} & 0 \\ 0 & Y^{red}_{X/E} & -Y^{eth}_{X/E} \\ -Y^{ox}_{X/O} & 0 & -Y^{eth}_{X/O} \\ Y^{ox}_{X/C} & Y^{red}_{X/C} & Y^{eth}_{X/C} \end{bmatrix} \begin{bmatrix} \mu^{ox}_s \\ \mu^{red}_s \\ \mu^{ox}_e \end{bmatrix} X - D\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{in} \\ 0 \\ OTR \\ -CPR \end{bmatrix}. \quad (20)$$

The O_2, CO_2 in gas phase and ethanol in liquid phase can be measured in industrial environment. However, due to linear dependency of these states, the corresponding yield matrix will be ill-conditioned [44]. Since the inversion of this matrix is required, the numerical implementation is extremely sensitive to numerical errors [26]. In order to overcome this drawback, the partial model is used as suggested by Pomerleau and Perrier [44]:

The fermentative partial model is stated as:

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -Y^{ox}_{X/S} & -Y^{red}_{X/S} \\ 0 & Y^{red}_{X/E} \\ -Y^{ox}_{X/O} & 0 \\ Y^{ox}_{X/C} & Y^{red}_{X/C} \end{bmatrix} \begin{bmatrix} \mu^{ox}_s \\ \mu^{red}_s \end{bmatrix} X - D\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{in} \\ 0 \\ OTR \\ -CTR \end{bmatrix}. \quad (21)$$

The respirative partial model is stated as:

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -Y^{red}_{X/S} & 0 \\ Y^{red}_{X/E} & -Y^{eth}_{X/E} \\ 0 & -Y^{eth}_{X/O} \\ Y^{red}_{X/C} & Y^{eth}_{X/C} \end{bmatrix} \begin{bmatrix} \mu^{ox}_s \\ \mu^{eth}_e \end{bmatrix} X - D\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{in} \\ 0 \\ OTR \\ -CTR \end{bmatrix}. \quad (22)$$

By means of asymptotic observer equations, the unmeasured process states and specifically biomass concentration can be calculated as:

Table 2
Parameters used in biomass estimation based on kinetic model of overflow metabolism and estimation based on observer [43].

$k_1, Y^{ox}_{X/S}$	3.65 C mol/mol	$k_2, Y^{red}_{X/S}$	0.36 C mol/mol
$k_3, Y^{red}_{X/E}$	0.19 C mol/mol	$k_4, Y^{eth}_{X/E}$	1.35 C mol/mol
$k_5, Y^{ox}_{X/O}$	1.56 C mol/mol	$k_6, Y^{eth}_{X/O}$	0.83 C mol/mol
$k_7, Y^{ox}_{X/C}$	1.45 C mol/mol	$k_8, Y^{red}_{X/C}$	0.2 C mol/mol
$k_9, Y^{eth}_{X/C}$	1.99 C mol/mol		

$$\frac{d}{dt}\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = -D\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{in} \\ 0 \end{bmatrix} - K_c K_m^{-1}\begin{bmatrix} OTR \\ CTR \end{bmatrix} \quad (23)$$

$$\begin{bmatrix} \hat{X} \\ \hat{S} \\ \hat{E} \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} + K_c K_m^{-1}\begin{bmatrix} O \\ C \end{bmatrix}. \quad (24)$$

Here, *OTR* and *CTR* are defined as the oxygen transfer rate and the carbon dioxide transfer rate respectively. The values of the parameters used in biomass estimation are based on known yield estimation and estimation based on observer presented in Table 2 [45].

In this method, only exhaust gas concentrations are measured and detailed mathematical operations are used.

4.4. Estimation based on Artificial Neural Network (ANN)

Among the techniques used for the development of a soft-sensor, ANN is a popular approach in the literature. Two different neuro-soft sensors to estimate biomass concentration and specific growth rate were examined previously in our group, with different input variables for the estimation of biomass concentration, and specific growth rate for large scale fermentations [17]. The inputs of the first type neuro-soft-sensor developed are fermentation time (t), respiratory quotient (RQ) and molasses feeding rate (F_s). These were appropriately chosen among available measurements. The output of the neural network is the total specific growth rate (μ). Three layered feed forward neural network architecture was used for the ANN estimator and the Levenberg–Marquardt (LM) training algorithm was used for obtaining suitable network parameters. Each training and test sets included 360 input/output sample data taken from the fermentation operated on process model with predetermined molasses feeding profiles. The initial conditions for fermentation were fixed for all training sets, but differed for some test sets. In order to observe the performance of the network, eight fermentation data sets were used from the production process. The inoculum sizes of the seven data sets were approximately the same but one had twice the amount of inoculum than the others. The estimation error for the last data was bad and could not be accepted. To overcome the estimation error occurring in the previous type estimator based on the process model, a new estimator (second type) was designed to estimate biomass concentration based on experimental data. The architecture of this soft biomass sensor is shown in Fig. 3.

As can be seen from the figure, input variables of the network were chosen as elapsed time (t), molasses feeding rate (F_s), $\%CO_2$ in outlet gas, %Ce in liquid phase and inoculum size (C_{xo}) and all

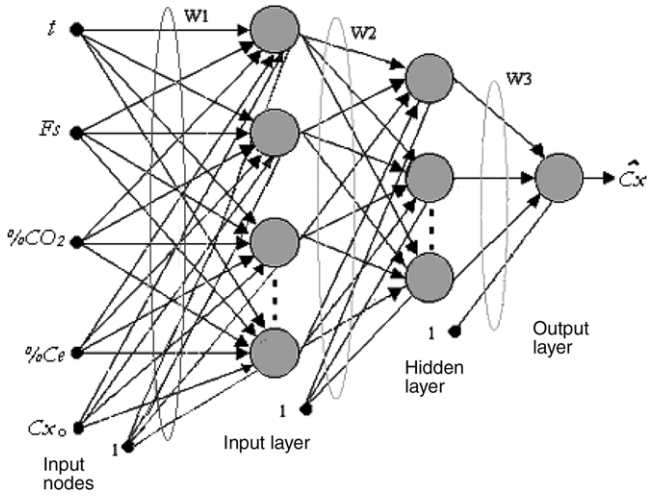

Fig. 3. Architecture of the three layered feed forward neural network estimator [17].

of them were available online. Normalizing to the [0–1] interval, all these variables were supplied to the network. The network has six input node and four neurons with logistic sigmoid activation functions in the input layer and the hidden layer respectively and one neuron in the output layer with piecewise linear activation function. The output of the network is the normalized increase in biomass concentration. Network output is denormalized and inoculum size (C_{xo}) is added to this output. Hence, the biomass concentration is computed at each time step. This network was trained with Levenberg–Marquardt (LM) training algorithm using five industrial scale fermentation data sets and validated with two data sets. As a result, second type ANN estimator was robust for the various initial inoculum sizes and used in this paper for comparison with the other estimation methods [17].

4.5. Estimation based on differential evaluation

Differential evaluation (DE) adds the weighted difference between two individuals randomly selected in population to a third individual [46]. The DE creates new candidate solutions by combining the parent individual and several other individuals of the same population. The candidate replaces the parent only if it has better fitness [47]. The DE comprises three important parameters: scaling factor *SF*, crossover probability constant *CR* and population size *PS*. In the proposed differential evolution based estimation method, the crossover probability constant *CR* can be calculated as a function of the generation number as given below:

$$CR = CR_0 \cdot e^{-(g-1)/10}. \quad (25)$$

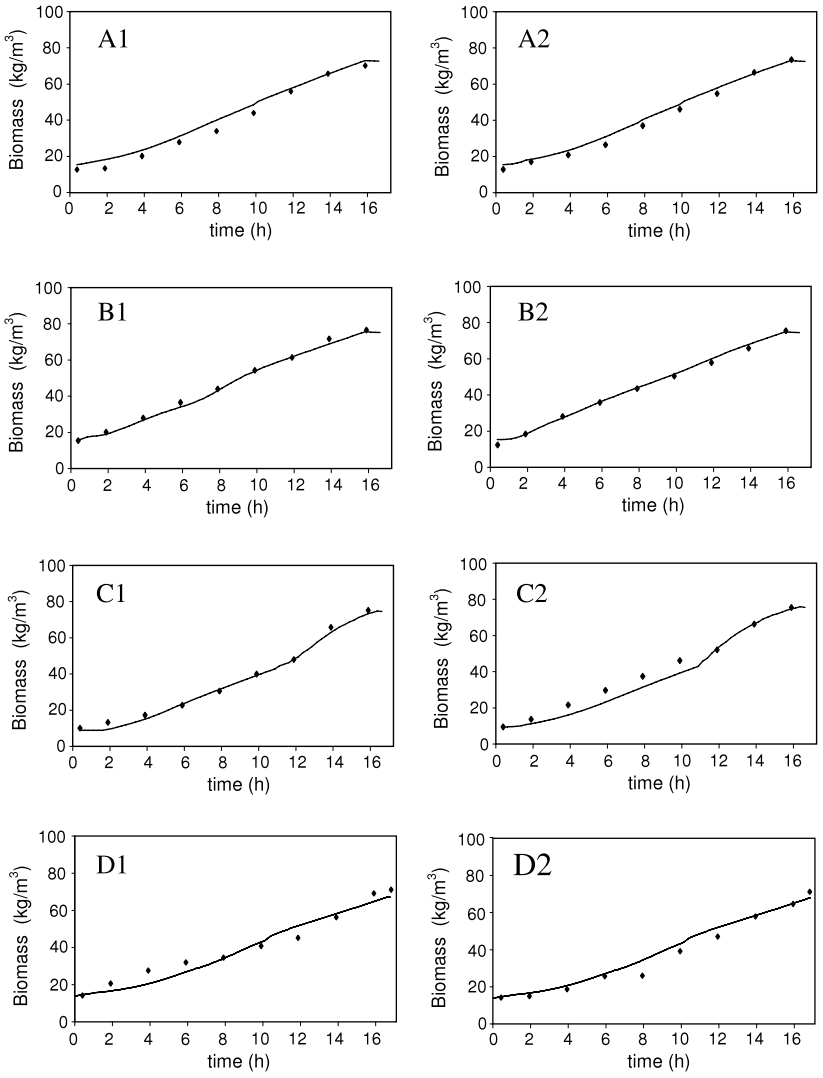

Fig. 4. Biomass curves obtained by the metabolic black-box model for type A, B, C and D fermentations and offline biomass measurements.

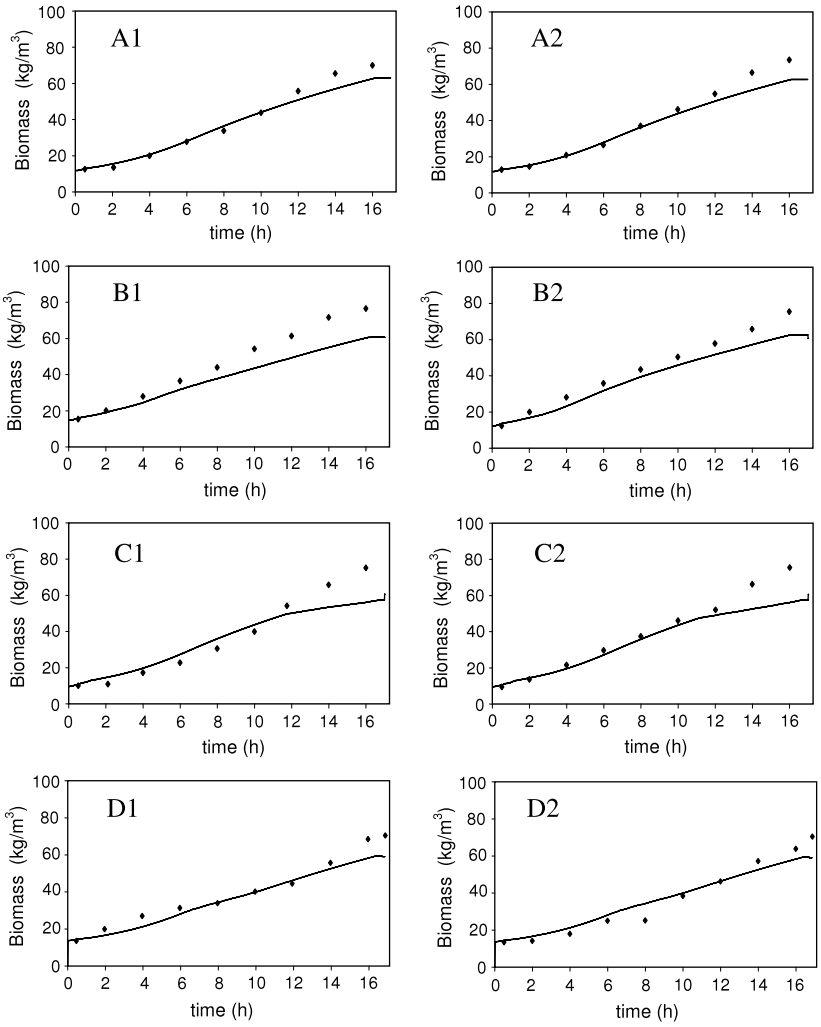

Fig. 5. Biomass curves obtained by the estimation based on kinetic model of overflow metabolism for type A, B, C and D fermentations and offline biomass measurements.

CR_0 represents the initial value of the crossover probability constant and g denotes the generation number. The DE algorithm consists of three genetic operators: mutation, crossover and selection. Mutation and crossover operators generate new individuals and selection operator determines suitable individuals, which have best fitness values and in this way population has the better individuals. There are several variations of DE [27]. In this proposed estimation method, the DE scheme often used in practice is considered as a strategy that is represented with notation of DE/rand-to-best/1/bin. The DE based estimator assumes that $\hat{\mu}$ is the estimate value of the total specific growth rate (μ) at time k and biomass concentration $X(k+1)$ and ethanol concentration $E(k+1)$ are estimated at time $k+1$ by equations given below [35]:

$$\hat{X}(k+1) = f\left(\hat{\mu}(k), \hat{X}(k), F(k), V(k)\right) \tag{26}$$

$$\hat{E}(k+1) = g\left(\hat{X}(k), \hat{E}(k), F(k), V(k), q_{e,pr}(k), q_{e,ox}(k)\right). \tag{27}$$

In this study, total fermentation time is divided into subintervals to overcome the heavy computing conditions, and the estimation problem is solved one by one for these calculation stages. The DE approach is run during a fixed time interval until a convergence criteria (ε) of the evaluation function is reached and at the end of each calculation stage biomass concentration is estimated using the best specific growth rate determined by the proposed approach. Table 3 gives the values of parameters used in biomass estimation based on DE.

Table 3
Parameters used in biomass estimation based on DE approach.

Parameter name	Symbol	Value
Number of population members	PS	15
Initial crossover probability constant	CR_0	0.75
Scaling factor	SF	0.1
Stop criteria value	ε	1e−12
Length of subintervals (h)	N	0.5

5. Results and discussions

The obtained biomass estimation curves and offline biomass measurements are given in Figs. 4–8 for each estimation method respectively. When the biomass curves are examined in Fig. 4, biomass estimations obtained by the metabolic black-box model closely match offline measurements for each fermentation type. Although there are differences in initial conditions and fermentation characteristic, the biomass estimations are very close to the offline estimations in metabolic black-box model estimation. The biomass estimations were obtained more accurately by means of reconciliated reaction rates. The main reasons are the precise description of process input/outputs, redundant measurements and reconciliation. In this study, energy balance is used as redundant measurement in order to obtain more accurate reaction rates and biomass concentration estimations.

In the estimation based on kinetic model of overflow metabolism, all the process parameters and yield coefficients are assumed

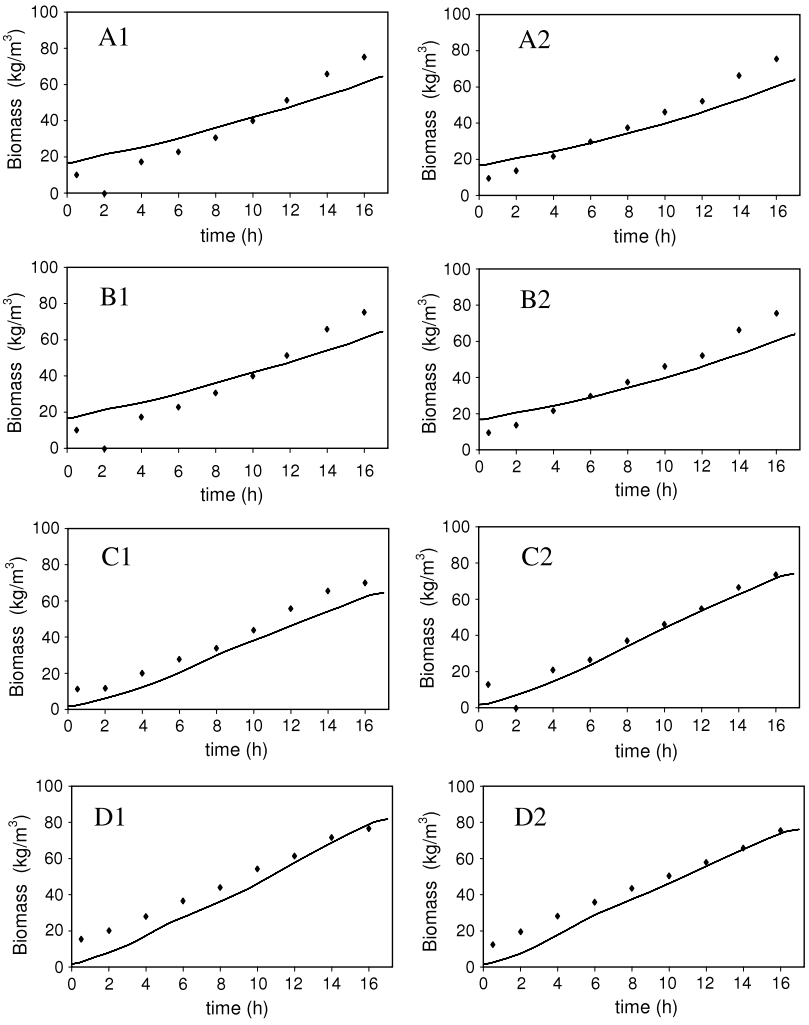

Fig. 6. Biomass curves obtained by the observer model for type A, B, C and D fermentations and offline biomass measurements.

known and not changed during fermentations. It can be seen in Fig. 5 that yeast cell model estimations were getting diverging during the second half of fermentations while matching closely at the beginning. Although there are differences in the structure of fermentations, the time varying process parameters could be the main source of the deviation in biomass estimations. Türker [37] has shown that for similar fermentations the yield coefficients varied along the fermentation. For the highly nonlinear fermentation process this static approach may not be suitable for biomass or specific growth rate estimations. In order to overcome this drawback, kinetic parameters and yield coefficients must be identified along the process. This can be done by online parameter identification algorithms or by employing look-up tables for the yield coefficients.

The asymptotic observer was implemented for the same fermentations data and the obtained biomass estimation curves are presented in Fig. 6. The main problem in observer based estimation was the low speed of convergence of estimation especially in first hour of fermentations. However, the advantage of the observer is that it requires less measurement compared to the other methods. As can be seen from the figure, the biomass curves are not as close to offline measurements as other model estimations. These differences may result from the structure of model equations and fixed yield coefficients.

The neural network estimator used in this study was formerly developed and operated with predetermined network parameters. The performance of the neural estimator is shown in Fig. 7. As can be noted from this figure, the performance is approximately similar for A and B type fermentations. The estimated biomass curves for type C fermentations are worse in comparison with the others. In fact, the main reason is that the network's training data set did not mainly include data like type C and D fermentations. In neural networks, it is well known that supervised training input/output samples should be distributed as homogeneously as possible in the input and output domain, otherwise, the trained neural network may give unexpected outputs for an area from which no samples are taken for training [17]. From the this point of view, it can be said that if the network could be formerly trained with the data obtained under as many different fermentation conditions as possible, the results obtained from type C and D fermentation would not occur.

The performance of the DE based estimator is given in Fig. 8. The biomass concentration values are fairly close to, especially, the experimental data of type A and type C and besides for the other data sets it can be seen from Fig. 8 that this has successful performances except for some parts of the fermentation process.

The kinetic model of overflow metabolism, observer and ANN based estimation results have a validity problem in comparison with the other estimation methods used in this study. Because ANN training is held on industrial scale fermenters data, the online and offline biomass estimations have significant differences when it is applied to data sets obtained from a pilot scale fermenter.

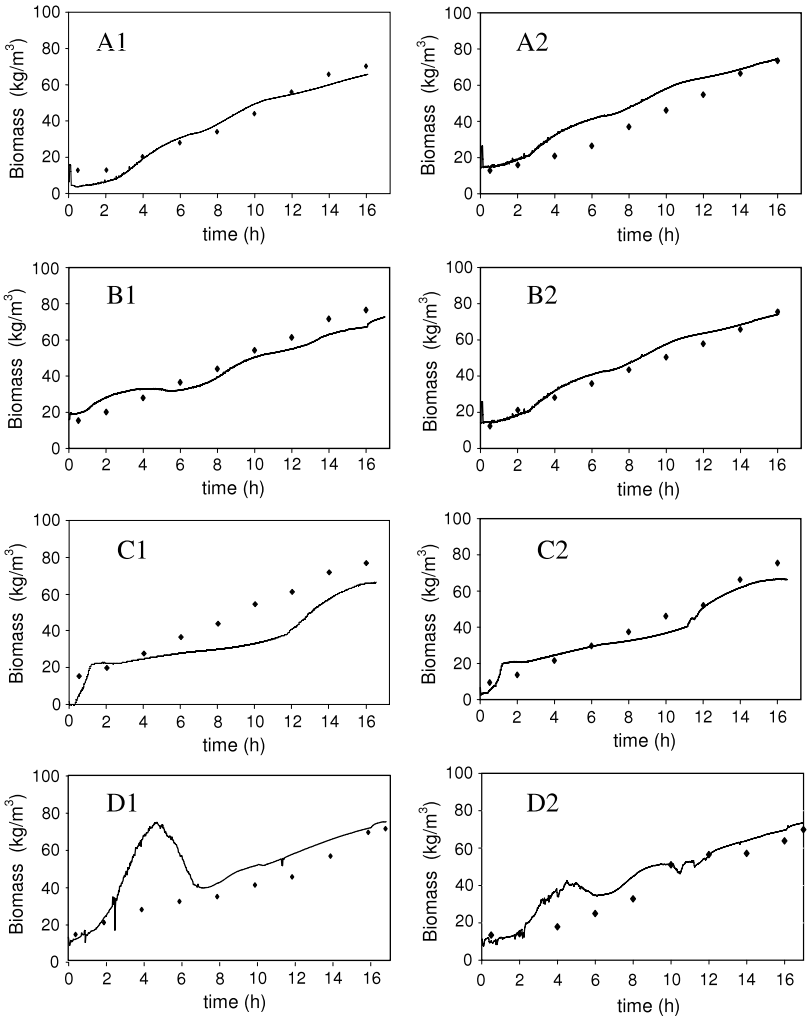

Fig. 7. Biomass curves obtained by the ANN model for type A, B, C and D fermentations and offline biomass measurements.

In order to evaluate the performances of the soft sensors the root mean square error of prediction is used:

$$RMSEP = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{y}_k - y_k)^2} \tag{28}$$

where K is the number of data points. The obtained results are given in Table 4. The lower RMSEP values indicate more accurate estimations. As can be seen in the table, the RMSEP values regarding the biomass estimations obtained by the metabolic black-box modelling are lower than the other modelling estimations. The performance results of the estimator using DE approach are generally better than those of other estimation techniques except the metabolic black-box based model. Furthermore, the RMSEP values of DE approach and the metabolic black-box model are almost close to one another for data of type A and type D. In short, the DE based estimator can be regarded as the estimator that has the closest performance to the metabolic black-box model based estimator.

The high performance of metabolic black-box modelling is the result of consistency of modelling, measurements and redundancy in measured process variables. Although the kinetic model of overflow metabolism and observer estimation methods use least number of process measurements (two measurements), their dependence on process kinetics and initial conditions give low estimation performance. The metabolic black-box estimation model uses more process measurements than the other methods, but redundant measurements and the simplicity of measurements give a competitiveness over the other models.

6. Conclusion

In this study, different estimation algorithms are compared with each other using experimental data obtained from industrial fermentations. The estimation methods are evaluated in terms of the number of measurements required, difficulty of implementation and required process knowledge. The evaluation results and required measurements presented in Tables 5 and 6. When the different estimation methods are compared in terms of primary measurements required, only metabolic black-box model requires more measurements that the others. However, the most of the measurements are simple and commonly available in industrial environment.

The estimation based on a kinetic model of overflow metabolism can be used for biomass estimation purposes if the kinetic model and yield coefficients are known but any changes in process conditions will give suspicious results. Observer based estimations show dependency on detailed mathematical knowledge and yield coefficients. The biomass estimations obtained by means of the metabolic black-box modelling shows good convergence with offline measurements with respect to other techniques. The elemental and heat balances are the basis of this modelling approach. In addition, this method does not require *a priori* process

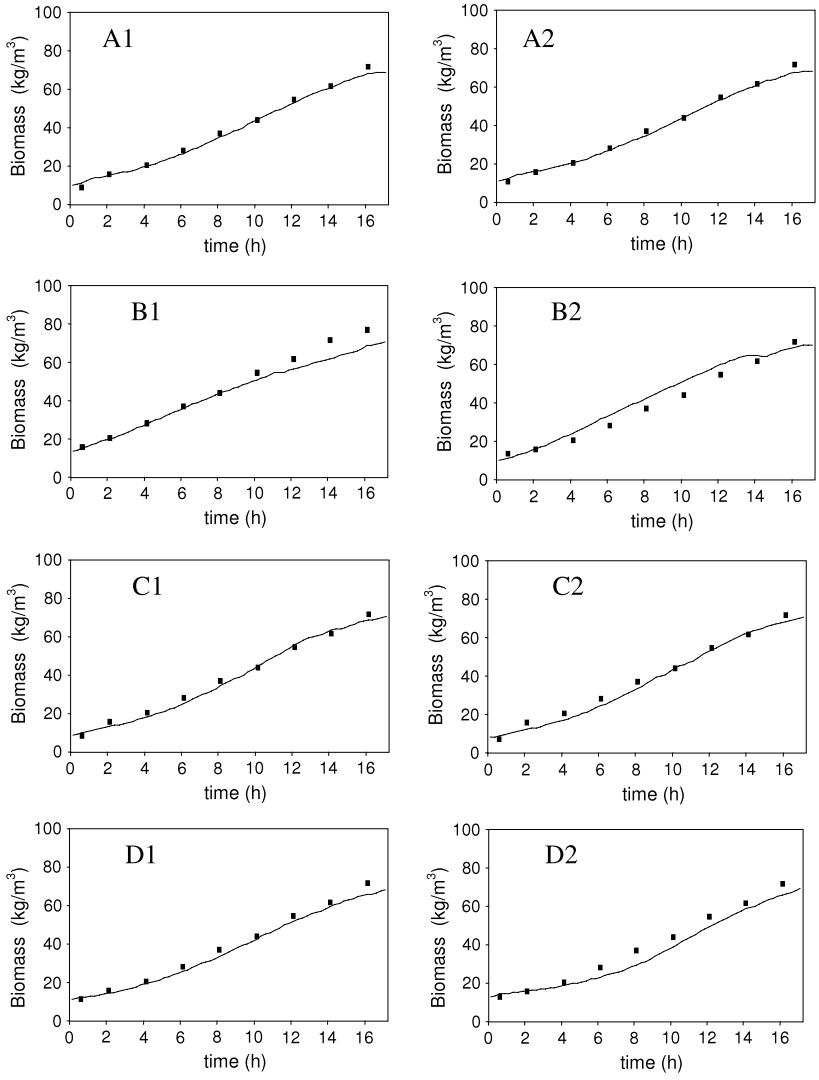

Fig. 8. Biomass curves obtained by the DE model for type A, B, C and D fermentations and offline biomass measurements.

Table 4
The obtained root mean square errors of predictions.

	Overflow metabolism	Observer	ANN	Metabolic black-box	DE
A	0.1849	0.3777	0.2638	0.1296	0.1213
B	0.4108	0.3580	0.2608	0.0697	0.2107
C	0.3315	0.3328	0.2964	0.0875	0.1725
D	0.2597	0.6145	0.4311	0.2181	0.2010

Table 5
Practical evaluation of the estimation methods.

Methods	Advantages	Disadvantages
Estimation based on kinetic model of overflow metabolism	Simple model to implement, few measurements required	Need detailed process knowledge, yield coefficient may not be constant throughout fermentation, bad convergence for different fermentations
Metabolic black-box model based estimation	Simple model to implement, high estimation accuracy, no need to know any process parameters	More measurements required
Observer based estimation	Few measurements required	Yield coefficient must be known, detailed mathematical knowledge, bad convergence for different fermentations
Artificial neural network based estimation	No need of process knowledge	Need much training process data, sensitive to the initial conditions
Differential evolution based estimation	Few measurements required, good convergence for different fermentations	Long response time as depending on high value parameters selected for DE

Table 6
The measurements used in models.

Measurements	Methods				
	Estimation based on kinetic model of overflow metabolism	Metabolic black-box model based estimation	Observer based estimation	Differential evolution	Artificial neural network based estimation
Molasses flow	+	+		+	+
Ethanol	+	+		+	+
O_2		+	+	+	
CO_2		+	+		+
Air temp.		+			
Airflow		+	+		
Temperature		+			
Flow of cooling water		+			
Inlet temp. of cooling water		+			
Outlet temp. of cooling water		+			
Molasses temp.		+			
Elapsed time					+

knowledge such as kinetic and yield parameters. However, it requires more measurements compared to other methods but these measurements are simple and easy to implement in practice. ANN based estimation requires large number of training data to cover different process conditions. It can be implemented in processes where quantitative process knowledge is not available. In DE based estimation, few measurements are required relating to the process, such as ethanol, oxygen concentration and molasses flow. The DE based estimations are satisfactory for both of the reactors having different dimensions and are irrelevant of the reactor size.

Acknowledgements

The authors would like to thank to late Prof. Dr. Rıdvan Berber (Ankara University) and Yaşar Becerikli (Kocaeli University) for a critical reading of the manuscript.

References

[1] James S, Legge R, Budman H. On-line estimation in bioreactors: a review. Reviews in Chemical Engineering 2000;16(4):311–40.
[2] Olsson L, Nielsen J. On-line and in-situ monitoring of biomass in submerged cultivations. Trends in Biotechnology 1997;5:517–22.
[3] Kornmann H, Rhiel M, Cannizzaro C, Marison I, von Stockar U. Methodology for real-time, multianalyte monitoring of fermentations using an *in-situ* mid-infrared sensor. Biotechnology and Bioengineering 2003;82(6):702–9.
[4] Hatch RT, Veilleux BG. Monitoring of *Saccharomyces cerevisiae* in commercial baker's yeast fermentation. Biotechnology and Bioengineering 1995;46: 371–4.
[5] Kiviharju K, Salonen K, Moilanen U, Eerikäinen T. Biomass measurement online: the performance of in situ measurements and software sensors. Journal of Industrial Microbiology and Biotechnology 2008;35:657–65.
[6] Cheruy A. Software sensors in bioprocess engineering. Journal of Biotechnology 1996;52:193–9.
[7] Dochain D, Perrier M. Dynamical modelling analysis, monitoring and control design for non-linear bioprocesses. Advances in Biochemical Engineering 1997;56:149–97.
[8] Dochain D. State and parameter estimation in chemical and biochemical processes: a tutorial. Journal of Process Control 2003;13:801–18.
[9] Komives C, Parker RS. Bioreactor state estimation and control. Current Opinion in Biotechnology 2003;14:468–74.
[10] O'Connor GM, Sanchez-Riera F, Cooney CL. Design and evaluation of control strategies for high cell density fermentations. Biotechnology and Bioengineering 1992;39:293–304.
[11] Wang HY, Cooney CL, Wang DIC. Computer aided baker's yeast fermentations. Biotechnology and Bioengineering 1977;19:69–86.
[12] Wang HY, Cooney CL, Wang DIC. Computer control of baker's yeast production. Biotechnology and Bioengineering 1979;21:975–95.
[13] Jobe AM, Herwig C, Surzyn M, Walker B, Marison I, Von Stockar U. Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing. Biotechnology and Bioengineering 2003;82(6):627–39.
[14] Zabriskie DW, Humphrey AE. Real time estimation of aerobic batch fermentation biomass concentration by component balancing. AIChE Journal 1978;24(1):138–46.
[15] Mou DG, Cooney CL. Growth monitoring and control through computer aided on-line mass balancing in a fed-batch penicillin fermentation. Biotechnology and Bioengineering 1983;25:225–55.
[16] Mou DG, Cooney CL. Growth monitoring and control in complex medium: a case study employing fed-batch penicillin fermentation computer aided on-line mass balancing. Biotechnology and Bioengineering 1983;25:257–69.
[17] Karakuzu C, Türker M, Öztürk S. Modelling, on-line state estimation and fuzzy control of production scale fed-batch baker's yeast fermentation. Control Engineering Practice 2006;14:959–74.
[18] Kitsuta Y, Kishimoto M. Fuzzy supervisory control of glutamic acid production. Biotechnology and Bioengineering 1994;44:87–94.
[19] Siimes T, Linko P, Numers C, Nakajima M, Endo I. Real time fuzzy knowledge based control of baker's yeast. Biotechnology and Bioengineering 1995;45: 135–43.
[20] Azevedo FS, Dham B, Oliveira F. Hybrid modelling of biochemical processes: a comparison with the conventional approach. Computers & Chemical Engineering 1997;21:751–6.
[21] Schubert J, Simutus R, Dors M, Havlik I, Lübbert A. Hybrid modelling of yeast production processes. Chemical Engineering & Technology 1994;17:10–20.
[22] James S, Legge R, Budman H. Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. Journal of Process Control 2002;12:113–21.
[23] Keulers M. Identification and control of a fed-batch process; application to culture of *Saccharomyces cerevisiae*. Ph.D. thesis. Technische Universiteit Eindhoven; 1993.
[24] Claes JE, Van Impe JF. On-line estimation of the specific growth rate based on viable biomass measurements: experimental validation. Bioprocess Engineering 1999;21:389–95.
[25] Neeleman R, Van Boxtel AJ. Estimation of specific growth rate from cell density measurements. Bioprocess and Biosystems Engineering 2001;24:179–85.
[26] Oliveira RMF. Supervision, control and optimization of biotechnological processes based on hybrid models. Ph.D. thesis. Department of Mathematic and Natural Sciences, Martin-Luther Universitat; 1998.
[27] Price K, Storn R. Differential evolution—a simple evolution strategy for fast optimization. Dr. Dobb's Journal 1997;22:18–24.
[28] Karaboga N, Çetinkaya B. Design of digital FIR filters using differential evolution algorithm. Circuits, Systems and Signal Processing 2006;25(5): 649–60.
[29] Subudhi B, Jena D. Differential evolution and Levenberg–Marquardt trained neural network scheme for nonlinear system identification. Neural Processing Letters 2008;27(3):285–96.
[30] Babu BV, Sastry KKN. Estimation of heat-transfer parameters in a trickle-bed reactor using differential evolution and orthogonal collocation. Computers and Chemical Engineering 1999;23:327–39.
[31] Babu BV, Pallavi GC, Syed Mubeen JH. Multiobjective differential evolution (MODE) for optimization of adiabatic styrene reactor. Chemical Engineering Science 2005;60(17):4822–37.
[32] Lee MH, Han C, Chang KS. Dynamic optimization of a continuous polymer reactor using a modified differential evolution algorithm. Industrial and Engineering Chemistry Research 1999;38(12):54825–31.
[33] Hocalar A, Türker M, Öztürk S. State estimation and error diagnosis in industrial fed-batch yeast fermentation. AIChE Journal 2006;52(11):3967–80.
[34] Hocalar A. Model based control of industrial fed-batch baker's yeast fermentation process. Ph.D. thesis. Department of Electronics and Communication, University of Kocaeli; 2007.
[35] Yüzgeç U, Türker M, Hocalar A. On-line evolutionary optimization of an industrial fed-batch yeast fermentation process. ISA Transactions 2009;48: 79–92.
[36] Hocalar A, Türker M. Model based control of minimal overflow metabolite in technical scale fed-batch yeast fermentation. Biochemical Engineering Journal 2010;51:64–71.

[37] Türker M. Measurement of metabolic heat in a production scale bioreactor by continuous and dynamic calorimetry. Chemical Engineering Communications 2003;190:573–98.
[38] Türker M. Development of biocalorimetry as a technique for process monitoring and control in technical scale fermentations. Thermochimica Acta 2004;419:73–81.
[39] Sonnleitner B, Kappeli O. Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. Biotechnology and Bioengineering 1986;28:927–38.
[40] Stephanopoulos GN, Aristidou AN, Nielsen J. Metabolic engineering: principles and methodologies. Academic Press; 1998.
[41] Van der Heijden RTJM, Heijnen JJ, Hellinga C, Romein B, Luyben KChAM. Linear constraint relations in biochemical reaction systems: I. Classification of calculability and the balanceability of conversion rates. Biotechnology and Bioengineering 1994;43:3–10.
[42] Van der Heijden RTJM, Heijnen JJ, Hellinga C, Romein B, Luyben KChAM. Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors. Biotechnology and Bioengineering 1994;43:11–20.
[43] Bastin G, Dochain D. On-line estimation and adaptive control of bioreactors. New York: Elsevier; 1990.
[44] Pomerleau Y, Perrier M. Estimation of multiple specific growth rates in bioprocess. AIChE Journal 1991;36(2):207–15.
[45] Beşli N, Türker M, Gül E. Design and simulation of a fuzzy controller for fed-batch yeast fermentation. Bioprocess Engineering 1995;13:141–8.
[46] Storn R, Price K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 1997;11:341–59.
[47] Brest J, Greiner S, Boskovic B, Mernik M, Zumer V. Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems. IEEE Transactions on Evolutionary Computation 2006; 10:646–57.