BAKER'S YEAST FERMENTATION PARAMETERS ESTIMATION: AN EVOLUTIONARY APPROACH

Celina P. Leão[1], Lino A. Costa[1], Filomena O. Soares[2]

[1]Dept. Systems and Production Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal, {cpl,lac}@dps.uminho.pt
[2]Dept. Industrial Electronics, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal, fsoares@dei.uminho.pt

Abstract: An initial understanding of the application of an evolutionary approach to baker's yeast fermentation parameters estimation is described. This problem has a nonlinear and multimodal nature. Thus, a global optimization technique, such as the evolution strategies, must be used in order to approximate the global optimum. The process is modelled by a set of differential-algebraic equations to obtain the six state variables concentration profiles, in a well-mixed reactor. The state variables profiles are obtained using parameters usually referred in literature: seven kinetic parameters and nine yield coefficients. A new problem of optimization can be formulated with the objective of minimizing the overall sum of the squared residuals that adjusts the simulation results to the experimental data. Using the evolutionary approach, not only the parameters can be estimated but also the most significant model parameters can be identified.

Keywords: Modelling, simulation and identification, Evolutionary algorithms, Baker's yeast fermentation.

1. INTRODUCTION

Baker's yeast industrial production is a fed-batch fermentation that employs a sugar solution feed. The mathematical model consists on a set of differential equations for the state variables, biomass X, glucose S, ethanol E, oxygen O, carbon dioxide, C and working volume V, together with the kinetics and the gas transfer rates algebraic relations (Feyo de Azevedo et. all, 1991, Soares, 1997; Leão and Soares, 2002). The kinetic and yield coefficients parameters can be obtained from the literature (Sonnleitner and Käppeli, 1986; Pommerleau, 1990). However, these parameters can be considered unknown and can be estimated fitting the model to the experimental data through an optimisation procedure. Experimental data was collected from a well-mixed fed-batch bioreactor, running for 20 hours.

Some studies have been done on optimization of the baker's yeast fermentation process (Soares, 1997; Leão and Soares, 2002; Leão and Soares, 2004) involving heuristic and multiple nonlinear regression analysis for parameters estimation. Multiobjective optimization with genetic algorithms has been applied to the optimization of antibiotic fermentation process (Potocnik and Grabec, 1999).

This work intends to investigate the possibilities of parameters estimation of the baker's yeast fermentation process using an evolutionary strategy. The main objective is to verify model and parameters validity. A multiple nonlinear regression can be performed by simultaneously fitting all differential equations to the data. This technique will enable to obtain the value of model parameters which minimise the overall sum of the squared residuals. In order to reduce the dimension of the optimisation procedure,

a heuristic sensitivity analysis was performed to identify the most significant model parameters, i.e., the parameters that give the most significant differences between experimental and simulated data (Leão and Soares, 2004). The maximum uptake rate for glucose and oxygen and the yield coefficients were identified as the most significant model parameters.

For the nonlinear regression analysis, a Marquardt algorithm (Wolfe, 1978) for multiresponse data, which uses an interpolation technique to combine the Gauss-Newton and Steepest Descent methods, was coded in Matlab. For the solution of the differential system of equations, a stiff Matlab solver was integrated in the developed code.

A set of statistical tests, which includes the correlation matrix of the parameters to determine the extent of the correlation between parameters, and a run test on the deviations between the state experimental and predicted values to ensure random distribution, was also performed along with the regression analysis.

The paper is organised as follows. Section 2 gives a brief description of the baker's yeast fermentation process and presents the model equations. In Section 3 a brief description of the evolutionary approach is presented and in Section 4 it is presented the Baker's yeast fermentation objective problem formulation. Section 5 shows some computational results. Finally, the conclusions and some suggestions for future work are referred in Section 6.

2. BAKER'S YEAST FERMENTATION – THE MODEL

In the baker's yeast fermentation process three metabolic pathways can be distinguish: respirative growth on glucose (O), fermentative growth on glucose (r), and respirative growth on ethanol (OE). Respirative pathways occur in presence of oxygen and the fermentative pathway in its absence (with production of ethanol) (Soares, 1997). The metabolic pathways of fermentative growth on glucose and oxidative growth on ethanol are competitive. This competition is governed by the respiratory capacity of the cells. If the instantaneous oxygen uptake capacity exceeds the oxygen need for total respiratory glucose uptake, then, all sugar uptakes follow the respiratory pathway (O) with the remaining oxygen being spent on ethanol respiratory uptake (OE). Otherwise, if the instantaneous oxygen uptake capacity is not enough, part of glucose uptake follows the respiratory pathway (O) while the remaining follows the fermentative pathway (r).

The mechanistic model for the fed-batch fermentation is obtained from mass balances for all the components: biomass X, glucose S, ethanol E, oxygen O and carbon dioxide, C. It is assumed that the yield coefficients (defined as gram of biomass X per gram of glucose S, ethanol E, oxygen O or carbon dioxide C, in each pathway), $Y^O_{X/S}$, $Y^r_{X/S}$, $Y^r_{X/E}$, $Y^{OE}_{X/E}$, $Y^O_{X/O}$, $Y^{OE}_{X/O}$, $Y^O_{X/C}$, $Y^r_{X/C}$, and $Y^{OE}_{X/C}$, are constant and the dynamics of the gas phase is neglected. The kinetics equations for baker's yeast growth, in the three metabolic pathways, respirative growth on glucose μ^O_S, fermentative growth on glucose, μ^r_S, and respirative growth on ethanol, μ^O_E, are considered as Monod equations.

The set of differential-algebraic equations is then defined:

- mass balance for the biomass,

$$\frac{dX}{dt} = \left(\mu^O_S + \mu^r_S + \mu^O_E - D\right)X \tag{1}$$

- mass balance for the sugar,

$$\frac{dS}{dt} = \left(-\frac{\mu^O_S}{Y^O_{X/S}} - \frac{\mu^r_S}{Y^r_{X/S}}\right)X + \left(S_f - S\right)D \tag{2}$$

where S_f is the glucose concentration in the feed,

- mass balance for the ethanol,

$$\frac{dE}{dt} = \left(\frac{\mu^r_S}{Y^r_{X/E}} - \frac{\mu^O_E}{Y^{OE}_{X/E}}\right)X - DE \tag{3}$$

- mass balance for the oxygen,

$$\frac{dO}{dt} = \left(-\frac{\mu^O_S}{Y^O_{X/O}} - \frac{\mu^O_E}{Y^{OE}_{X/O}}\right)X - DO + OTR \tag{4}$$

- mass balance for the carbon dioxide,

$$\frac{dC}{dt} = \left(\frac{\mu^O_S}{Y^O_{X/C}} + \frac{\mu^r_S}{Y^r_{X/C}} + \frac{\mu^O_E}{Y^{OE}_{X/C}}\right)X - DC - CTR \tag{5}$$

- accumulation of the working volume, V, during the fed-batch process,

$$\frac{dV}{dt} = DV \tag{6}$$

where D is dilution rate (ratio feed rate/volume) defined by, $D = F/V$.

The gas transfer rates for oxygen and carbon dioxide are given, respectively, by

$$OTR = K^O_L a\left(O^* - O\right) \tag{7}$$

$$CTR = K^C_L a\left(C - C^*\right) \tag{8}$$

where $K^i_L a$ are overall mass transfer coefficients for oxygen and carbon dioxide and O^* and C^* are the corresponding equilibrium concentrations.

As referred above, regarding metabolic pathways, two situations may occur: excess of oxygen that implies no fermentative growth of biomass or lack of oxygen and consequently excess of glucose that implies no respiratory growth on ethanol. Then, the kinetics equations for the respirative regime are taken as:

$$\mu^O_S = Y^O_{X/S}.q_S \tag{9a}$$

$$\mu^r_S = 0 \tag{10a}$$

$$\mu^O_E = min\left(\mu^O_{E_1}, \mu^O_{E_2}\right) \tag{11a}$$

or for the respiro-fermentative regime as:

$$\mu^O_S = Y^O_{X/S}.\frac{q_O}{a} \tag{9b}$$

$$\mu^r_S = Y^r_{X/S}.\left(q_S - \frac{q_O}{a}\right) \tag{10b}$$

$$\mu^O_E = 0 \tag{11b}$$

where a is the stoichiometric coefficient of the oxygen in the respiratory pathway of glucose, q_s is

the specific glucose uptake rate and q_o is the specific oxygen uptake rate.

Equations, (12-15), must be added to the kinetics equations, (9a-11a) or (9b-11b), for the estimation of the specific growth rates:

$$\mu_{E_1}^{O} = \mu_E^{max} \frac{E}{E+K_E} \frac{K_i}{S+K_i} \quad (12)$$

$$\mu_{E_2}^{0} = \frac{Y_{X/O}^{OE}}{Y_{X/E}^{OE}} \left(q_O - aq_S\right) \quad (13)$$

$$q_S = q_S^{\max} \frac{S}{S+K_S} \quad (14)$$

$$q_O = q_O^{\max} \frac{O}{O+K_O} \quad (15)$$

where $\mu_E^{\max}$ is the maximal specific growth rate, K_i is the inhibition parameter, K_E is the saturation parameter, $q_S^{\max}$ is the maximal specific glucose uptake rate, K_S and K_O are saturation parameters and $q_O^{\max}$ is the maximal specific oxygen uptake rate.

This bottleneck is determined by calculating aq_S and comparing with q_0. If $aq_S \le q_0$ the respirative regime takes place; and the regime respiro-fermentative if $aq_S > q_0$.

The kinetic coefficients, $q_S^{\max}$, $q_O^{\max}$, $\mu_E^{\max}$, K_E, K_i, K_S and K_O, are considered constants.

The set of differential-algebraic equations, (1-8, 12-15, 9a-11a) or (1-8, 12-15, 9b-11b), defines the mathematical model for the baker's yeast fermentation process.

3. EVOLUTIONARY OPTIMIZATION

The formulation of the above problem is that of a nonlinear optimization problem with differential algebraic constraints. Due to the nonlinear and constrained nature of the problem, it is multimodal and nonconvex. Therefore, if the problem is solved using standard local optimization methods, as, for instance, the Levenberg-Marquardt method, it is possible that the solution obtained will be of local nature. Thus, it is important to use, for these problems, global optimization techniques in order to obtain an approximation to global optimum. Some past studies namely, the work of Moles *et al.* (2003) suggests that Evolutionary Algorithms and, in particular, Evolution Strategies might be the most competitive and promising global optimization techniques. Evolutionary algorithms mimic the natural evolution of the species in biological systems and they can be used as a robust global optimization tool. In past years, several distinct approaches have emerged. These algorithms do not require any continuity or convexity property of the problem being solved. Moreover, unlike conventional algorithms, only information regarding the objective function and constraints is required to perform the search.

Evolution Strategies (ESs), in the past, prove to be powerful single objective optimizers (Schwefel 1995). They start searching from an initial population (a set of points) and transition rules between generations are deterministic. The individuals of the population are vectors of real coded decision variables that are potential optimal solutions. Each generation, λ offspring are generated from μ progenitors by mutation and recombination. The best individuals are then selected for next generation. Important features of evolution strategies are the self adaptation of step sizes for mutation during the search and the recombination of individuals that is performed between ρ individuals. In general, this algorithm is referred as $(\mu/\rho+\lambda)$-ES.

4. BAKER'S YEAST FERMENTATION - PROBLEM FORMULATION

The baker's yeast fermentation problem can be formulated as a single optimization problem and solved by an evolutionary approach, as follows,

· ProblemP1NW:

$$\min f_1\left(x_1,...,x_{16}\right) = SQR_X + SQR_E + SQR_S \quad (16a)$$

with

$$SQR_X = \sum_{j=1}^{m} r_j^T r_j = \sum_{j=1}^{m} \left(X_j^e - X_j\right)^T \left(X_j^e - X_j\right) \quad (16b)$$

where r_j is the residual vector of X_j and X_j^e is the vector of the m experimental observations of the dependent variable X_j. Similar formulas for E and S.

The boundaries of the decision variables x_i (i=1, …, n, n=16) are described in Table 1 for the initial conditions: X(0) = 0.3654 g l^{-1}, S(0) = 0.0361 g l^{-1}, E(0) = 0.3114 g l^{-1}, O(0) = 0.0065 g l^{-1}, C(0) = 0.0027 g l^{-1}, V(0) = 2 g l^{-1}, S_f = 30 g l^{-1}, final volume = 5l, glucose feed-rate F = 0.12 l h^{-1}. These different boundaries were defined taking into account the literature values. $K_L^i a$, O^* and C^*, were considered as constants and equals to 100, 0.007 and 0.0053, respectively.

The computational experiences were carried out using a (10/10+100)-ES. Table 2 describes the parameters used to perform the optimization.

The individuals of the initial population were generated randomly. Each individual consists on a vector of n decision variables x_i (real values) and n standard deviations σ_i (i=1, … , n). The parents and offspring population sizes were, respectively, 10 and 100. All members of the parent population are recombined using discrete recombination (for the decision variables and standard deviations). Mutation is applied according to a Normal distribution with mean zero and variance σ_i^2. The initial values for σ_i were given by equation $\sigma_i^{(0)} = \Delta x/\sqrt{n}$ with $\Delta x = x_u - x_l$ (where x_u and x_l are the upper and lower bounds of variable x_i). During the search, the standard deviations σ_i were actualized according to the equation $\sigma_i^{(k+1)} = \sigma_i^{(k)} e^{z_i} e^{z}$ as in Schwefel, 1995, where $z_i \sim N\left(0,\Delta\sigma^2\right)$, $z \sim N\left(0,\Delta\sigma'^2\right)$, $\Delta\sigma$ and $\Delta\sigma'$ are parameters of the algorithm. The values considered for these parameters were $\Delta\sigma = 1/\sqrt{2n}$. and $\Delta\sigma' = 1/\sqrt{2\sqrt{n}}$.

Table 1 Boundaries of the decision variables considered in the simulations

Kinetics parameters	Yield coefficient
$0.5 \le q_S^{\max} \le 8.0$	$0.1 \le Y_{X/S}^{O} \le 0.8$
g gluc^{-1}g biom^{-1}h^{-1}	g biomg gluc^{-1}
$0.01 \le q_O^{\max} \le 0.5$	$0.01 \le Y_{X/S}^{r} \le 0.2$
g O_2^{-1}g biom^{-1} h^{-1}	g biomg gluc^{-1}
$0.01 \le \mu_E^{\max} \le 0.5$	$0.01 \le Y_{X/E}^{r} \le 0.8$
h^{-1}	g biomg etha^{-1}
$0.01 \le K_E \le 0.5$	$0.1 \le Y_{X/E}^{OE} \le 1.0$
g l^{-1}	g biomg etha^{-1}
$0.01 \le K_i \le 0.5$	$0.5 \le Y_{X/O}^{O} \le 1.8$
g l^{-1}	g biomg O_2^{-1}
$0.01 \le K_S \le 0.5$	$0.1 \le Y_{X/O}^{OE} \le 0.8$
g l^{-1}	g biomg O_2^{-1}
$10^{-5} \le K_O \le 10^{-3}$	$0.1 \le Y_{X/C}^{O} \le 1.0$
mg l^{-1}	g biomg CO_2^{-1}
	$0.01 \le Y_{X/C}^{r} \le 0.8$
	g biomg CO_2^{-1}
	$0.5 \le Y_{X/C}^{OE} \le 1.5$
	g biomg CO_2^{-1}

Table 2 The ES parameters

Parameter	
Maximum number of generations	100
Parents population size	10
Offspring population size	100
Number of recombinants	10

5. COMPUTATIONAL RESULTS AND DISCUSSION

Figures 1 to 3 show the results for the most important state variables: biomass, sugar and ethanol. The continuous line reflects the estimated values using the 16 decision variables or parameters values achieved from the evolutionary approach (Table 3). The dashed line illustrates the results using the parameters from the literature and the points are the experimental data obtained under the initial conditions previously mentioned.

Using the evolutionary approach, it was possible to find a set of parameters (decision variables) that, with the model, reasonably approximates the experimental data. The highest error values are observed in the beginning of the process. At the end of the simulation process, the error increases. The results suggest the use of a weighted objective function since the magnitude of the residuals is different.

Figure 4 shows the objective function (16) values (the overall sum of the squared residuals) of the best solutions found along the search. After the 60th generation the objective function change very gradually, converging to 1.50. It is clear the

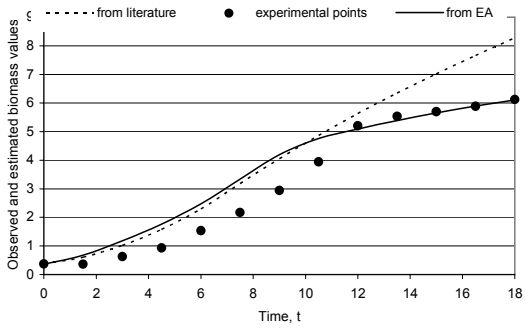

Fig. 1. Biomass observed (♦) and estimated values from literature (--) and from the Evolutionary Approach (—).

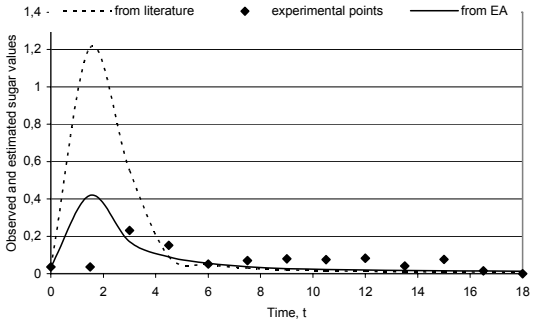

Fig. 2. Sugar observed (♦) and estimated values from literature (--) and from the Evolutionary Approach (—).

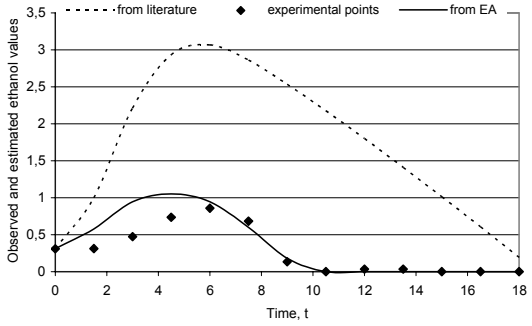

Fig. 3. Ethanol observed (♦) and estimated values from literature (--) and from the Evolutionary Approach (—).

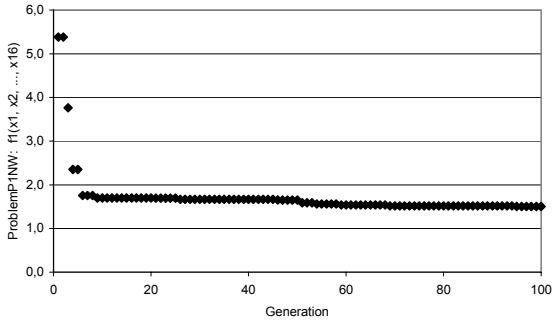

Fig. 4. Convergence of the Evolutionary Approach for ProblemP1NW.

convergence of the algorithm. Thus, this algorithm seems to be a valid approach to the optimization of the baker's yeast fermentation process.
Most of the parameters obtained by the evolutionary approach are very different when compared with the values given in literature (Table 3). However, they allow the achievement of good predictions values.

Table 3 Comparison between the evolutionary approach, P1NW, final results and literature (Sonnleitner and Käppeli, 1986; Pomerleau, 1990)

Kinetics parameters		Yield coefficient	
model	literature	model	literature
q_S^{max} =		$Y_{X/S}^{O}$ =	
4.23	3.5	0.80	0.49
g gluc^{-1}g biom^{-1}h^{-1}		g biomg gluc^{-1}	
q_O^{max} =		$Y_{X/S}^{r}$ =	
0.18	0.256	0.20	0.05
g O_2^{-1}g biom^{-1} h^{-1}		g biomg gluc^{-1}	
μ_E^{max} =		$Y_{X/E}^{r}$ =	
0.52	0.17	0.44	0.10
h^{-1}		g biomg etha^{-1}	
K_E =		$Y_{X/E}^{OE}$ =	
0.005	0.10	1.0	0.72
g l^{-1}		g biomg etha^{-1}	
K_i =		$Y_{X/O}^{O}$ =	
0.20	0.10	1.60	1.20
g l^{-1}		g biomg O_2^{-1}	
K_S =		$Y_{X/O}^{OE}$ =	
0.80	0.20	0.80	0.64
g l^{-1}		g biomg O_2^{-1}	
K_O =		$Y_{X/C}^{O}$ =	
0.001	0.0001	0.10	0.81
mg l^{-1}		g biomg CO_2^{-1}	
		$Y_{X/C}^{r}$ =	
		0.027	0.11
		g biomg CO_2^{-1}	
		$Y_{X/C}^{OE}$ =	
		0.50	1.11
		g biomg CO_2^{-1}	

From the analysis of these results, $Y_{X/E}^{OE}$ and $Y_{X/S}^{O}$ were identified as parameters that have a significant influence in the fermentation process, *i.e.*, the values of these two parameters have influence on the objective function. Figure 5 shows the values of these two coefficients found along the search for the best solutions of the objective function. The objective function is sensible to the variations of the values of these parameters: the first different values correspond to the highest objective function values. However, this analysis can not be made individually. Others parameters must also identified: $Y_{X/O}^{O}$, $Y_{X/O}^{OE}$, $Y_{X/S}^{r}$ and q_O^{max}. They also play an important role in the fermentation process since they represent yield coefficients that define conversion proportionality between two state variables: biomass and other variable, namely glucose, ethanol and oxygen, in the three metabolic pathways (O, OE, r). One can not forget that some dependence and/or relationship can exist between them.
With opposite behaviour can be identified some parameters that have no significant influence on the objective function, *e.g.*, the kinetic parameter K_O (see Figure 6), independently of their values the objective function remains unaffected.

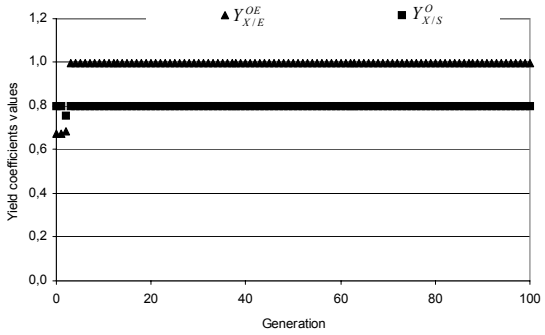

Fig. 5. $Y_{X/E}^{OE}$ and $Y_{X/S}^{O}$ yield coefficient values found along the search for ProblemP1NW.

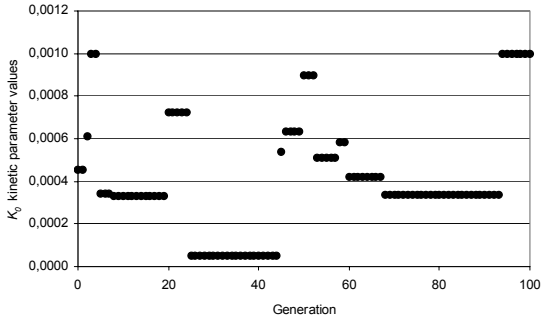

Fig. 6. K_O kinetic parameter values found along the search for ProblemP1NW.

6. CONCLUSIONS AND FUTURE WORK

The method used in this paper, an evolutionary approach based on evolution strategies, enables us to fit a baker's yeast model, consisting on multiple dependent variables, to multi-response experimental data in order to obtain the best values for the most significant parameters, which minimize the overall sum of squared residuals between the data and the model.
The parameters values obtained in most cases, were very different from the literature, however they allow obtaining good predictions values.
Yield coefficients, due to their definition (conversion factor stated as gram of biomass per gram of glucose, ethanol or oxygen), were identified as the most relevant parameters.
The results point out for a next study, consisting on the application of a weighted objective function since the magnitude of the residuals is different. Multiobjective formulations of the problem will be also investigated since the components of the overall

sum of squared residuals can be seen as distinct objectives to minimize.

REFERENCES

Feyo de Azevedo, S., P. Pimenta, F.O. Soares and E.C. Ferreira (1992). *Studies on on-line State and Parameter Estimation through a Real-time Process Simulato*r, Modeling and Control of Biotechnical Processes (2nd IFAC Symp. and 5th Int. Conf. Computer Applications in Fermentation Technology, Keystone USA, 1992, M. Nazmul Karim, G. Stephanopoulos, Eds.), Pergamon Press, Oxford, 453-458.

Leão, C. P. and F. O. Soares (2002). Two Different Strategies for Baker's Yeast Fermentation Process Simulation, In *Recent Advances in Simulation, Computational Methods and Soft Computing* (Nikos E. Mastorakis (Ed)), 11-16. WSEAS Press.

Leão, C. P., F. O. Soares and E. M. G. P. Fernandes (2003). Multiple Nonlinear Regression Analysis for the Baker's Yeast Fermentation Parameters Estimation, In *Modelling and Simulation in Chemical Engineering* (A. E. Rodrigues, P. Oliveira, J. A. Castro, J. A. Ferreira and M. C. Coimbra (Eds)), 1-8. CIM.

Leão, C. P. and F. O. Soares (2004). Heuristic Sensivity Analysis for Baker's Yeast Kinetics Parameters, *Investigação Operacional Journal*, **24**, 1-16.

Moles, C. G., P. Mendes and J.R. Banga (2003), Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods, *Genome Research*, **13**, 2467-2747.

Pomerleau, Y. (1990). *Modélisation et Contrôle d'un Procédé Fed-Batch de Culture des Levures à Pain (Saccharomyces cerevisiae)*, PhD Thesis, Université de Montreal.

Potocnik, P. and I. Grabec (1999), Empirical Modeling of Antibiotic Fermentation Process using Neural Networks and Genetic Algorithms. *Mathematics and Computers in Simulation*, **49**, 363-379.

Sanders, D.H., R.J. Eng and A.F. Murph (1985). *Statistics. A Fresh Approach*, Third Edition, McGraw-Hill Book Co.

Schwefel, H.-P. (1995). *Evolution and Optimal Seeking*, John Wiley & Sons.

Soares, F. O. (1997). *Monitorização e controlo de fermentadores – Aplicação ao fermento de padeiro*, PhD. Thesis, Universidade do Porto, (in portuguese).

Sonnleitner, B. and O. Käppeli (1986). Growth of Saccharomyces cerevisiae is controlled by its limites respiratory capacity: formulation and verification of a hypothesis, *Biotech. Bioeng.*, **28**, 927-937.

Wolfe, M. A. (1978). *Numerical Methods for Unconstrained Optimization - An Introduction.*, Van Nostrand Reinhold Company.