Serdica J. Computing **4** (2010), 11–18

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

PARAMETER IDENTIFICATION OF A FED-BATCH CULTIVATION OF *S. CEREVISIAE* USING GENETIC ALGORITHMS*

Maria Angelova, Stoyan Tzonkov, Tania Pencheva

Abstract. Fermentation processes as objects of modelling and high-quality control are characterized with interdependence and time-varying of process variables that lead to non-linear models with a very complex structure. This is why the conventional optimization methods cannot lead to a satisfied solution. As an alternative, genetic algorithms, like the stochastic global optimization method, can be applied to overcome these limitations. The application of genetic algorithms is a precondition for robustness and reaching of a global minimum that makes them eligible and more workable for parameter identification of fermentation models. Different types of genetic algorithms, namely simple, modified and multi-population ones, have been applied and compared for estimation of nonlinear dynamic model parameters of fed-batch cultivation of *S. cerevisiae*.

1. Introduction. Genetic algorithms (GA) are directed random search techniques, based on the mechanics of natural selection and natural genetics, according to the Darwinian evolutionary theory [4]. GA finds the global optimal

ACM Computing Classification System (1998): G.1.6.

Key words: Genetic algorithms, Parameter identification, Fed-batch cultivation of *S. cerevisiae*.

*This work is partly supported by the National Science Fund Project MI – 1505/2005.

solution in complex multidimensional search spaces simultaneously evaluating many points in the parameter space. They require only information concerning the quality of the solution and do not require linearity in the parameters. Properties like hard problems solving, noise tolerance, ease of interface and hybridization make genetic algorithms suitable and more workable for the parameter identification of fermentation models [1], [8], [9], [10].

Three main operators, namely reproduction, crossover and mutation, guide the mechanisms of the simple genetic algorithms (SGA) [2], [3], [4], [5]. Through reproduction chromosomes representing better possible solutions are chosen from the population. Selected individuals are crossed over to form new offspring. Mutation is then applied with determinate probability. The mutation prevents falling of all solutions in the population into a local optimum of the solved problem. For the new individuals the objective function and fitness function values are again calculated. The new offspring is inserted into the population. Then the generated population is used for a further run of the algorithm.

The modified genetic algorithm (MGA) [9] has a structure similar to the simple genetic algorithm but the operator of reproduction is processed after both crossover and mutation. Thus the loss of the best chromosome from the last population never happens and the current generation will be superior to or at least the same as the parents.

A single population genetic algorithm that can be improved by introducing many populations, called subpopulations, is known as multi-population genetic algorithm (MpGA) [2], [3], [4]. These subpopulations evolve independently from each other for a certain number of generations (isolation time), after that a number of individuals are distributed between the subpopulation (migration).

The aim of this investigation is the abovementioned three types of genetic algorithms, namely simple, modified and multi-population GA, to be applied and compared for a fed-batch cultivation of *S. cerevisiae*.

2. Comparison of different types of genetic algorithms. There are many operators, functions, parameters and settings in the genetic algorithms that can be implemented differently in various problems [2], [3], [5]. The effect of genetic algorithms' parameters was investigated for the values shown in Table 1, according to the following statements [6], [9]. A very big generation gap value does not improve the performance of GA, especially regarding how fast the solution will be found. Mutation is randomly applied with low probability, typically in the range 0.01 to 0.1. Particularly important parameters of GA are the population size and number of generations. If there is too low a number of chromosomes, GA

has few possibilities to perform crossover and only a small part of search space is explored. On the other hand, if there are too many chromosomes, GA slows down. An analysis of the effect of genetic algorithms' parameters has been done for each parameter separately, investigating different values for the examined parameter when the others are kept constant. For the next steps, the best obtained values for the already examined parameters are used.

Table 1. Investigation of the range of genetic algorithm parameters.

GGAP	0.8	0.85	0.9	0.95	
MUTR	0.05	0.06	0.07	0.08	0.09
NIND	20	40	60	80	100
MAXGEN	100	200	500	1000	

In Table 1 GGAP means generation gap – how many new individuals are created; MUTR – mutation rate; NIND – number of individuals per subpopulations; MAXGEN – maximum number of generations.

The minimization error between experimental data and the model simulation decreases inessentially by 0.01% when the values of MAXGEN increase. But the total computation time increases. Therefore MAXGEN = 100 is accepted as a good compromise between the model precision and computation time. For the other genetic algorithm parameter the minimization error increases inessentially by 0.01% when the values of GGAP, MUTR and NIND increase. The total computation time increases too. Therefore GGAP = 0.8, MUTR = 0.05 and NIND = 20 are accepted as a good compromise between the model precision and calculation time. Table 2 and Table 3 present the values of genetic parameters and operators used in this investigation.

Table 2. Genetic parameters

Parameter	**Value**
GGAP	0.8
XOVR	0.95
MUTR	0.05
PRECI	20
NIND	20
MAXGEN	100
MIGR	0.2
INSR	0.95
SUBPOP	5
MIGGEN	20

Table 3. Genetic operators

Operator	**Type**
encoding	binary
reinsertion	fitness-based
crossover	double point
mutation	bit inversion
selection	roulette wheel selection
fiitness function	linear ranking

In Table 2 XOVR means crossover rate; PRECI – precision of binary representation; MIGR – migration rate; INSR – insertion rate; SUBPOP – number of subpopulations; MIGGEN – number of generation after which migration takes place between subpopulations.

3. Results and discussion. The comparison of the application of the three types of GAs, namely SGA, MGA and MpGA, is carried out based on a data set from a fed-batch cultivation of *S. cerevisiae*. The experimental data are obtained in the*Institute of Technical Chemistry – University of Hannover*, Germany. The cultivation of the yeast *S. cerevisiae* is performed in a 2 l reactor, using a Schatzmann medium [7]. The glucose in the feeding solution is 35 g/l. The temperature was controlled at 30 °C, the pH at 5.5. The stirrer speed was set to 1200 rpm. The aeration rate was kept at 300 l/h. The biomass and ethanol were measured off-line and the substrate (glucose) and dissolved oxygen were measured on-line.

The mathematical model of *S. cerevisiae* fed-batch cultivation is commonly described as follows, according to the mass balance:

$$\frac{dX}{dt} = \mu X - \frac{F}{V}X \tag{1}$$

$$\frac{dS}{dt} = -q_S X + \frac{F}{V}(S_{in} - S) \tag{2}$$

$$\frac{dE}{dt} = q_E X - \frac{F}{V}E \tag{3}$$

$$\frac{dO_2}{dt} = -q_{O_2} X + k_L^{O_2} a\,(O_2^* - O_2) \tag{4}$$

$$\frac{dV}{dt} = F, \tag{5}$$

where X is the concentration of biomass, [g.l^{-1}]; S – concentration of substrate (glucose), [g.l^{-1}]; E – concentration of ethanol, [g.l^{-1}]; O_2 – concentration of oxygen, [%]; O_2^* – dissolved oxygen saturation concentration, [%]; F – feeding rate, [l.h^{-1}]; V – volume of bioreactor, [l]; $k_L^{O2}a$ – volumetric oxygen transfer coefficient, [h^{-1}]; S_{in} – glucose concentration in the feeding solution, [g.l^{-1}]; μ, q_S, q_E and q_{O_2} are respectively specific rates of growth, substrate utilization, ethanol production and dissolved oxygen consumption, [h^{-}].

The considered here fed-batch cultivation of *S. cerevisiae* is characterized by keeping the glucose concentration equal or below to its critical level ($S_{crit} = 0.05$ g.l^{-1}), sufficient dissolved oxygen in the broth $O_2 \geq O_{2crit}(O_{2crit} = 18$ %),

as well as presence of ethanol in the broth. This state corresponds to the so - called mixed oxidative state according to the functional state modeling approach [7], [11]. As presented in [11], the specific growth rate is generally found to be a sum of two terms, one describing the contribution of sugar and the other the contribution of ethanol to yeast growth. Both terms have the structure of Monod model. Monod model is also used for the specific ethanol and sugar consumption rates. The dissolved oxygen consumption rate is obtained as a sum of two terms, which are proportional to the specific glucose rate and specific ethanol production rates, respectively. More precisely, the specific rates in Eqs. 1–4 are presented as follows:

$$\mu = \mu_{2S}\frac{S}{S+k_S} + \mu_{2E}\frac{E}{E+k_E} \tag{6}$$

$$q_S = \frac{\mu_{2S}}{Y_{SX}}\frac{S}{S+k_S} \tag{7}$$

$$q_E = -\frac{\mu_{2E}}{Y_{EX}}\frac{E}{E+k_E} \tag{8}$$

$$q_{O_2} = q_E Y_{OE} + q_S Y_{OS}, \tag{9}$$

where μ_{2S}, μ_{2E} denote maximum growth rates of substrate and ethanol, [h^{-1}]; k_S, k_E – saturation constants of substrate and ethanol, [$g.l^{-1}$]; Y_{SX}, Y_{EX}, Y_{OE}, Y_{OS} – yield coefficients, [$g.g^{-1}$].

As an optimization criterion, least square deviation between the model output and the experimental data obtained during cultivation was used:

$$J_Y = \sum (Y - Y*)^2 \rightarrow \min, \tag{10}$$

where Y is the experimental data, $Y*$ is model predicted data, $Y = [X,\ S,\ E,\ O_2]$.

The identification of the model parameters of a fed-batch cultivation of *S. cerevisiae* with simple, modified and multi-population genetic algorithms were done using *Matlab 7.0*, *Genetic Algorithm Toolbox* [2], [3], [5]. The results are presented in Table 4.

As can be seen from the results presented in the Table 4, the best value of the optimization criterion is obtained using MpGA, with about 50% less error than the one obtained with SGA and MGA. At the same time, the best value of the total computation time is obtained using MGA, which reaches the global minimum in time comparable to SGA, but more that two times faster than MpGA.

Table 4. Results from the model parameters identification

Parameter	**SGA**	**MGA**	**MpGA**
J	0.0223	0.0225	0.0147
CPU time, s	74.7656	67.5313	146.2969
μ_{2S}, h^{-1}	0.9616	0.9211	0.9001
μ_{2E}, h^{-1}	0.0971	0.0872	0.1192
k_S, $g.l^{-1}$	0.1154	0.1176	0.1200
k_E, $g.l^{-1}$	0.7963	0.7620	0.7607
Y_{SX}, $g.g^{-1}$	0.4279	0.4279	0.4088
Y_{EX}, $g.g^{-1}$	1.2898	1.2898	6.0204
k_{la}^{o2}, h^{-1}	38.5895	127.2898	90.5778
Y_{OS}, $g.g^{-1}$	313.8285	989.8014	716.0857
Y_{OE}, $g.g^{-1}$	234.7797	62.6547	178.6444

Since the results from the application of three types of GA are too similar, only the results obtained with the most precise MpGA are presented here. The following figures demonstrate simulation results for experimental and model data for biomass (Fig. 1), ethanol (Fig. 2), substrate (Fig. 3) and dissolved oxygen (Fig. 4).

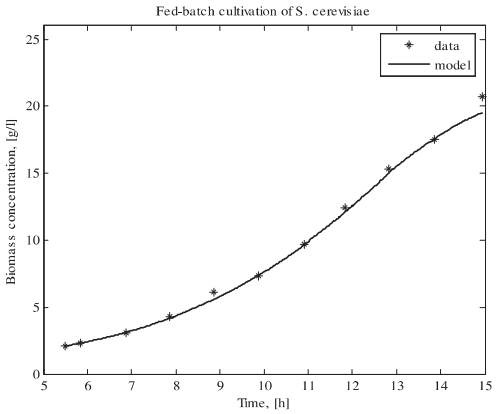

Fig. 1. Experimental and model data for biomass

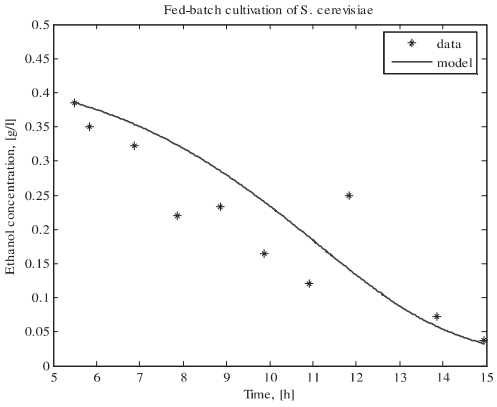

Fig. 2. Experimental and model data for ethanol

The results from MpGA application for parameter identification of *S. cerevisiae* fed-batch cultivation presented in the figures show the effectiveness of this type of GA.

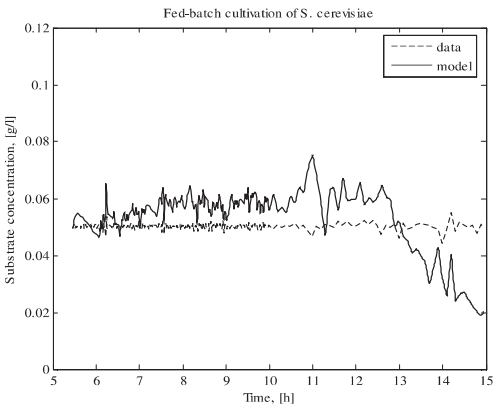

Fig. 3. Experimental and model data for substrate

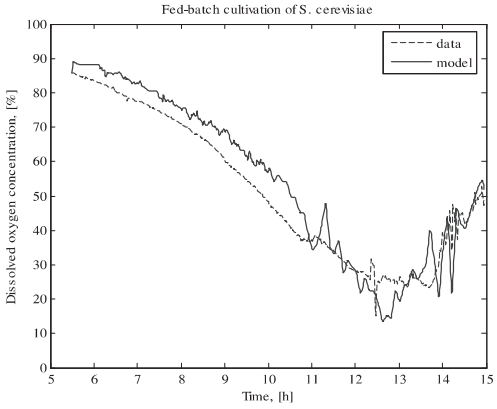

Fig. 4. Experimental and model data for dissolved oxygen

4. Analysis and conclusion. A comparison of three different types of genetic algorithms, namely simple, modified and multi-population GA, is presented here for a fed-batch cultivation of *S. cerevisiae*. The best value of the optimization criterion is obtained using MpGA, with about 50% less error than obtained with SGA and MGA. At the same time, the best value of the total computation time is obtained using MGA, which reaches the global minimum in time comparable to SGA, but more than two times faster than MpGA. This is why it is up to the user to decide which type of GA to use as a compromise between the time consumption and the model.

REFERENCES

[1] Carrillo-Ureta G. E., P. D. Roberts, V. M. Becerra. Genetic algorithms for optimal control of beer fermentation. In: Proceeding of the 2001 IEEE International Symposium on Intelligent Control, Mexico City, Mexico, 2001, 391–396.

[2] Chipperfield A. J., P. Fleming, H. Pohlheim, C. M. Fonseka. Genetic algorithm toolbox for use with MATLAB. Users's Guide. Version 1.2, Dept. of Automatic Control and System Engineering, University of Sheffield, U. K., 1994.

[3] Chipperfield A. J., P. Fleming. The Matlab Genetic Algorithm Toolbox, 1995.

[4] Goldberg D. Genetic algorithms in search, optimization and machine learning. Addison-Weslcy Publishing Company, Massachusetts, 1989.

[5] Genetic Algorithm Toolbox. Users's Guide, MatWorks Inc, 1999.

[6] Obittko M. Genetic algorithm. `http://cs.felk.cvut.cz/~xobitko/ga/main.html`, 2005.

[7] Pencheva T., O. Roeva, I. Hristozov. Functional state approach to fermentation processes modeling. (Eds St. Tzonkov, B. Hitzmann.), Prof. Marin Drinov Academic Publishing House, Sofia, 2006.

[8] Ranganath M., S. Renganathan, C. Gokulnath. Identification of bioprocesses using genetic algorithm. *Bioprocess Engineering*, **21** (1999), 123–127.

[9] Roeva O. A modified genetic algorithm for a parameter identification of fermentation processes. *Biotechnology and Biotechnological Equipment*, **20** (2006), No 1, 202–209.

[10] Roeva O. Genetic algorithm for a parameter estimation of a fermentation process model: A comparison. *Bioautomation*, **3** (2005), 19–28.

[11] Zhang X.-C., A. Visala, A. Halme, P. Linco. Functional state modelling approach for bioprosesses: Local models for aerobic yeast growth prosesses. *J. Proc. Contr.*, **4** (1994), No 3, 127–134.

Centre of Biomedical Engineering
Bulgarian Academy of Sciences
105, Acad. G. Bonchev Str.
1113 Sofia, Bulgaria
e-mail: `maria@clbme.bas.bg`
e-mail: `tzonkov@clbme.bas.bg`
e-mail: `tania.pencheva@clbme.bas.bg`

Received November 3, 2009
Final Accepted February 4, 2010