Genetic Algorithms based Parameter Identification of Yeast Fed-batch Cultivation

Maria Angelova, Stoyan Tzonkov, and Tania Pencheva

Institute of Biophysics and Biomedical Engineering - Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
maria.angelova@clbme.bas.bg, tzonkov@clbme.bas.bg,
tania.pencheva@clbme.bas.bg

Abstract. Different kinds of genetic algorithms have been investigated for a parameter identification of a fermentation process. Altogether eight realizations of genetic algorithms have been presented - four of simple genetic algorithms and four of multi-population ones. Each of them is characterized with a different sequence of implementation of main genetic operators, namely selection, crossover and mutation. A comparison of considered eight kinds of genetic algorithms is presented for a parameter identification of a fed-batch cultivation of *S. cerevisiae*. All kinds of multi-population algorithms lead to considerable improvement of the optimization criterion value but for more computational time. Among the considered multi-population algorithms, the best one has an operators' sequence of crossover, mutation and selection. Different kinds of considered simple genetic algorithms lead to similar values of the optimization criterion but the genetic algorithm with an operators' sequence of mutation, crossover and selection is significantly faster than the others.

1 Introduction

Fermentation processes (FP) are widely used in different branches of industry, i.e. in the production of pharmaceuticals, chemicals and enzymes, yeast, foods and beverages. Live microorganisms play an important role in these processes so their peculiarities predetermine some specific characteristics of FP as modeling and control objects. As complex, nonlinear, dynamic systems with interdependence and time-varying process variables, FP are a serious challenge for modelling and further high-quality control. An important step for adequate modeling of nonlinear models of FP is the choice of a certain optimization procedure for model parameter identification. The conventional optimization methods can not overcome the limitations of FP, while genetic algorithms (GA), as stochastic global optimization method, are quite promising. GA are a direct random search technique for finding global optimal solution in complex multidimensional search space. GA have a lot of advantages such as hard problems solving, noise tolerance, easy to interface and hybridize. All these properties predetermine GA as suitable and more workable for the optimization of highly non-linear problems, especially for a parameter identification of fermentation process models [1, 2, 8–10].

Simple (SGA) and multi-population (MpGA) genetic algorithm, as presented initially in Goldberg [5] search a global optimal solution using three main genetic operators in a sequence selection, crossover and mutation. For the purpose of this investigation, SGA and MpGA with such sequence are denoted respectively as SGA-SCM and MpGA-SCM. Many improved variations of the SGA and MpGA have been developed [1, 4, 7, 10]. Among them are the modified genetic algorithm [10] with a sequence crossover, mutation and selection, here denoted as SGA-CMS, and consequent modification of MpGA based on such exchange [1], here denoted as MpGA-CMS. In these algorithms selection operator has been processed after performing of crossover and mutation. The main idea for such operators' sequence is to prevent the loss of reached good solution by either crossover or mutation or both operators. SGA-CMS applied to a parameter identification of *E. coli* fed-batch cultivation [10] and further tested also for a parameter identification of *S. cerevisiae* fed-batch cultivation [1] improves the optimization capability of the algorithm, decreasing decision time. MpGA-CMS applied to a parameter identification of *S. cerevisiae* fed-batch cultivation decreases the algorithm calculation time and improves significantly the decision adequacy compared to SGA. Obtained promising results applying SGA-CMS and MpGA-CMS encourage more investigations to be performed in order further improvements of the algorithms to be found.

In GA as presented initially in Holland [6] and further in Goldberg [5], the operator mutation is usually applied after the operator crossover. The basic idea of GA is to imitate the mechanics of natural selection and genetics, so one can make an analogy with the processes occurring in the nature. As such, the probability mutation to come first and then crossover is comparable to the idea both processes to occur in a reverse order. The purpose of this study is to investigate the influence of the genetic operators' sequence selection, crossover and mutation in SGA and MpGA. Presented in [5] SGA-SCM and MpGA-SCM, as well as developed SGA-CMS [10] and MpGA-CMS [1] have been compared with four new proposed kinds with exchanged sequence of mutation and crossover operators. Obtained altogether eight kinds of the SGA and the MpGA have been compared in terms of accuracy and performance for a parameter identification of *S. cerevisiae* fed-batch cultivation.

2 Implementation of Exchanged Operators' Sequence of Crossover and Mutation in Simple and Multi-population Genetic Algorithms

The ideology of implementation of GA for the parameter identification purposes could be summarized as follows. The chromosomes represents the models parameters and corresponding objective function value is associated to each chromosome. The objective function is used to provide a measure of how individuals have performed in the problem domain. In the case of minimization problem, the fitted individuals will have the lowest numerical value of the associated objective function. This raw measure of fitness is only used as an intermediate stage in

determining the relative performance of individuals in genetic algorithms. The selection algorithm chooses individuals for reproduction on the basis of their relative fitness. Selected chromosomes, through reproduction, crossover and mutation, form a new population. Generated in that way population is used for a further run of the algorithm. The GA is terminated when a certain number of generations is fulfilled, a mean deviation in the population is satisfied, or when a particular point in the search space is encountered.

Simple genetic algorithm (denoted here as SGA-SCM) guides the mechanism of evaluation implementing the three main operators in a sequence selection, crossover and mutation. Presently four modifications of SGA and MpGA are elaborated and demonstrated, implementing the exchange of operators' sequence crossover and mutation. Newly presented modifications are as follows:

- SGA-SMC - a modification of the developed in [5] SGA-SCM;
- SGA-MCS - a modification of the developed in [10] SGA-CMS;
- MpGA-SMC - a modification of the developed in [5] MpGA-SCM;
- MpGA-MCS - a modification of the developed in [1] MpGA-CMS.

Since the MpGA are more complex than SGA, and as a case with most exchanges towards the originally presented by Goldberg GA, the elaboration of MpGA-MCS is shortly presented below. Multi-population genetic algorithm is a single population genetic algorithm, in which many populations, called subpopulations, evolve independently from each other for a certain number of generations. After a certain number of generations (isolation time), a number of individuals are distributed between the subpopulations. In the beginning, the MpGA generates a random population of n chromosomes, i.e. suitable solutions for the problem. In order to prevent the loss of reached good solution by either crossover or mutation or both operators, selection operator has been processed after performing of crossover and mutation [1]. The new modification presented here is that, the individuals are reproduced processing firstly mutation, followed by crossover. The elements of chromosome are a bit changed when a newly created offspring mutates, after that the genes from parents combine to form a whole new chromosome during the crossover. After the reproduction, the MpGA-MCS calculates the fitness values for the offspring and the best fitted individuals are selected to replace the parents. Then the algorithm evaluates the objective values (cost values) of the individuals in the current population and according to that the new chromosome is created. the MpGA is terminated when a certain number of generations is fulfilled.

Proposed exchange in a operators' sequence mutation and crossover has been also applied towards SGA-SCM, SGA-CMS and MpGA-SCM. This results in new algorithm modifications considered in this investigation and denoted as SGA-SMC, SGA-MCS and MpGA-SMC respectively.

3 Parameter Identification of *S. cerevisiae* Fed-batch Cultivation using Different Kinds of Simple and Multi-population Genetic Algorithms

Experimental data of *S. cerevisiae* fed-batch cultivation is 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 [8]. The initial liquid volume is 1.3 l. Glucose in feeding solution is 35 g/l. The temperature was controlled at 30C, the pH at 5.5. The stirrer speed was set to 1200 rpm. The aeration rate was kept at 300 l/h. Biomass and ethanol were measured off-line, while substrate (glucose) and dissolved oxygen were measured on-line.

Mathematical model of *S. cerevisiae* fed-batch cultivation is commonly described as follows, according to the mass balance [11]:

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

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

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

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

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

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

Considered here fed-batch cultivation of *S. cerevisiae* is characterized with keeping glucose concentration equal or below to its critical level ($S_{crit} = 0.05$ g/l) and with sufficient dissolved oxygen in the broth $O_2 \geq O_{2crit}$ ($O_{2crit} = 18\%$). This state corresponds to so called mixed oxidative state according to functional state modeling approach [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. Dissolved oxygen consumption rate is obtained as a sum of two terms, which are directly proportional to the specific glucose rate

and specific ethanol production rate, respectively. Hence, specific rates in Eqs. (1)-(5) are presented as follows:

$$\mu = \mu_{2S}\frac{S}{S+k_S} + \mu_{2E}\frac{E}{E+k_E}, q_S = \frac{\mu_{2S}}{Y_{SX}}\frac{S}{S+k_S},$$

$$q_E = -\frac{\mu_{2E}}{Y_{EX}}\frac{E}{E+k_E}, q_{O_2} = q_E Y_{OE} + q_S Y_{OS} \tag{6}$$

where μ_{2S}, μ_{2E} - maximum growth rates of substrate and ethanol, [1/h]; k_S, k_E - saturation constants of substrate and ethanol, [g/l]; Y_{ij} - yield coefficients, [g/g].

As an optimization criterion, mean square deviation between the model output and the experimental data obtained during cultivation has been used:

$$J_Y = \sum (Y - Y^*)^2 \rightarrow \min \tag{7}$$

where Y and Y^* are the experimental and model predicted data respectively, $Y = [X, S, E, O_2]$.

Parameter identification of the model (1)-(5) has been performed using *Genetic Algorithm Toolbox* in *Matlab 5.3* environment [3]. All the computations are performed using a PC Intel Pentium 4 (2.4 GHz) platform running Windows XP. Consequently eight kinds of genetic algorithms - four kinds of SGA and four kinds of MpGA, four of them newly presented here, have been applied for the purpose of a parameter identification of *S. cerevisiae* fed-batch cultivation. A comparison between performances of four kinds of SGA is presented in Table 1, while Table 2 presents results obtained using four kinds of MpGA.

Table 1. Table 1. Results from model parameter identification using different kinds of SGA

Parameter	**SGA-SCM**	**SGA-SMC**	**SGA-CMS**	**SGA-MCS**
J_Y	0.0223	0.0221	0.0225	0.0223
CPU time, s	73.8281	73.4688	64.8281	59.5156
μ_{2S}, 1/h	0.9616	0.9038	0.9211	0.9119
μ_{2E}, 1/h	0.0971	0.1320	0.0872	0.0966
k_S, g/l	0.1154	0.1119	0.1176	0.1109
k_E, g/l	0.7963	0.7990	0.7620	0.7987
Y_{SX}, g/g	0.4279	0.4072	0.4279	0.4316
Y_{EX}, g/g	1.2898	1.7699	1.2898	1.3170
$k_L^{O_2}a$, 1/h	38.5895	116.4160	127.2898	141.1076
Y_{OS}, g/g	313.8285	898.6292	989.8014	993.2537
Y_{OE}, g/g	234.7797	281.1797	62.6547	166.6377

As shown in Table 1, the optimization criterion values obtained with four types of standard genetic algorithms are very similar. Hopefully, there is no

loss of adequacy of the model when the operator mutation is performed before crossover. Moreover, proposed modification in the algorithm reduses time of reaching of a global minimum. While the implementation of SGA-SMC compared to SGA-SCM does not lead to significant decrease (<1%) of decision time, the use of SGA-MCS reduces the time with 9% compared to SGA-CMS. The fastest algorithm SGA-MCS achieves the global minimum for 24% less time than SGA-SCM. Presented here comparison shows that the implementation of the operators in a sequence of mutation, crossover and than selection is the most optimal in attitude of rate with reserved high adequacy of the decision.

Table 2. Table 2. Results from model parameter identification using different kinds of MpGA

Parameter	**MpGA-SCM**	**MpGA-SMC**	**MpGA-CMS**	**MpGA-MCS**
J_Y	0.0144	0.0145	0.0144	0.0145
CPU time, s	100.6563	98.0625	95.6094	100.4688
μ_{2S}, 1/h	0.9000	0.9012	0.9003	0.9073
μ_{2E}, 1/h	0.1447	0.0967	0.1342	0.0549
k_S, g/l	0.1500	0.1499	0.1500	0.1500
k_E, g/l	0.8000	0.7739	0.8000	0.7647
Y_{SX}, g/g	0.3944	0.4131	0.4076	0.4271
Y_{EX}, g/g	6.9156	4.8402	6.5616	2.7389
$k_L^{O_2}a$, 1/h	101.6394	71.5478	95.7177	98.3150
Y_{OS}, g/g	808.7495	569.6776	753.0205	772.7289
Y_{OE}, g/g	522.0352	759.6290	282.2053	449.7269

As it is seen from Table 2, the values of the optimization criterion obtained using multi-population genetic algorithms are also comparable. The expected improvement in CPU time of MpGA-SMC towards MpGA-SCM has been observed (with about 3%), while MpGA-MCS reaches the decision slowly than MpGA-CMS. Hence, the fastest MpGA has an operators sequence of crossover, mutation and selection and reaches the decision with 5% faster than MpGA-SCM.

The results presented in Table 1 for SGA and those in Table 2 for MpGA have been compared too. The value of the optiomization criterion in MpGAs is about 50% less than the criterion in SGA. Unfortunately, MpGA need more time to reach the global minimum. That is why it is up to the user to make a decision which type of GA to use as a compromise between the time consumption and model precision.

Due to the similarity of the results from the implementation of all considered here eight types of GA, only these obtained with MpGA-CMS (as the fastest and the most precise among the MpGA) are here presented. Fig. 1 presents results from experimental data and model prediction respectively for biomass, substrate, ethanol and dissolved oxygen.

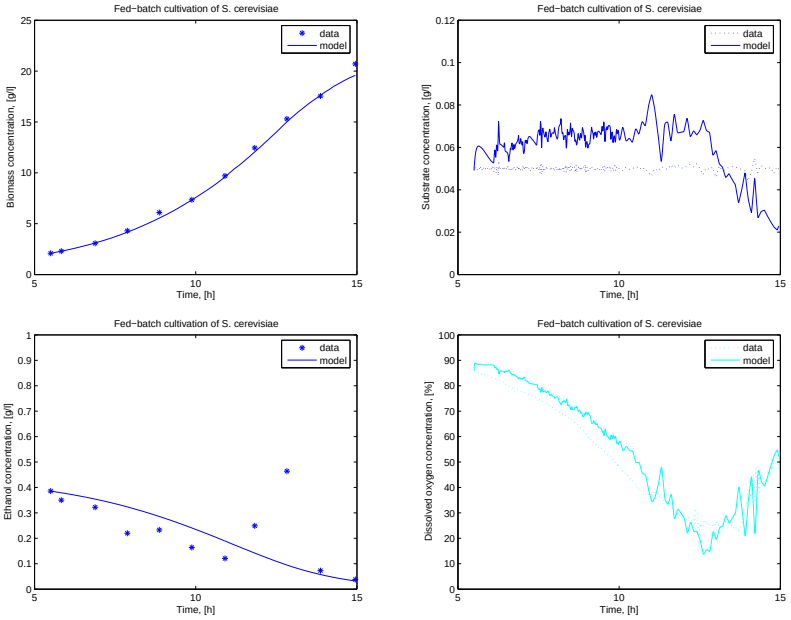

Fig. 1. Experimental data and model prediction for biomass, substrate, ethanol and dissolved oxygen concentrations

4 Analysis and Conclusions

In this investigation altogether four modifications two of SGA and two of MpGA have been proposed, implementing the exchanged operators sequence of mutation and crossover. Newly suggested SGA-SMC, SGA-MCS, MpGA-SMC and MpGA-MCS have been developed and compared respectively to SGA-SCM, SGA-CMS, MpGA-SCM and MpGA-CMS for the purposes of a parameter identification of a fed-batch cultivation of *S. cerevisiae*. Implementation of the main genetic operators in order mutation, crossover and selection in SGA significantly improves calculation time of the algorithm without affecting to the model adequacy. SGA-MCS solves the optimization problem 9% faster than SGA-CMS and 24% than SGA-SCM. Four kinds of MpGA lead to significant improvement of about 50% of the optimization criterion value but for more computational time. Among the considered MpGA, the fastest and the most precise one implements an operators sequence of crossover, mutation and selection. Finally, comparing SGA and MpGA it is up to the user to make a decision which type of GA to use as a compromise between the time consumption and model precision.

5 Acknowledgements

This work is partially supported by the European Social Fund and Bulgarian Ministry of Education, Youth and Science under Operative Program "Human Resources Development", grant BG051PO001-3.3.04/40 and National Science Fund of Bulgaria, grant number DID 02-29 "Modeling Processes with Fixed Development Rules".

References

1. Angelova, M., Tzonkov, St., Pencheva, T.: Modified multi-population genetic algorithm for yeast fed-batch cultivation parameter identification. Int. J. Bioautomation **13(4)** (2009) 163–172
2. Carrillo-Ureta, G. E., Roberts, P. D., Becerra, V. M.: Genetic algorithms for optimal control of beer fermentation. Proc. of the 2001 IEEE Int. Symp. on Intelligent Control, Mexico City, Mexico (2001) 391–396
3. Chipperfield, A. J., Fleming, P., Pohlheim, H., Fonseca, C. M.: Genetic algorithm toolbox for use with MATLAB. Users guide, version 1.2. Dept. of Automatic Control and System Engineering, University of Sheffield, UK (1994)
4. Cordon, O., Herrera, F.: Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems. Fuzzy Sets and Systems **118** (2001) 235–255
5. Goldberg, D.: Genetic algorithms in search, optimization and machine learning. Addison-Wiley Publishing Company, Massachusetts (1989)
6. Holland, J.: Adaptation in natural and artificial systems. MIT Press (1975)
7. Kuo, R. J., Chen, C. H., Hwang, Y. C.: An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. Fuzzy Sets and Systems **118** (2001) 21–45
8. Pencheva, T., Roeva, O., Hristozov, I.: Functional state approach to fermentation processes modelling. Tzonkov St., B. Hitzmann (Eds.). Prof. Marin Drinov Academic Publishing House, Sofia (2006)
9. Ranganath, M., Renganathan, S., Gokulnath, C.: Identification of bioprocesses using genetic algorithm. Bioprocess Engineering **21** (1999) 123–127
10. Roeva, O., A modified genetic algorithm for a parameter identification of fermentation processes. Biotechnol. and Biotechnol. Equip. **20** (2006) 202–209
11. Zhang, X.-C., Visala, A., Halme, A., Linco, P.: Functional state modelling approach for bioprosesses: local models for aerobic yeast growth processes. J. Proc. Contr. **4(3)** (1994) 127–134