Elsevier Editorial System(tm) for Computers and Mathematics with Applications
Manuscript Draft

Manuscript Number:

Title: Purposeful Model Parameters Genesis in Simple Genetic Algorithms

Article Type: SI: BioMath 2011

Keywords: genetic algorithms, purposeful genesis, model parameters, fermentation process, Saccharomyces cerevisiae

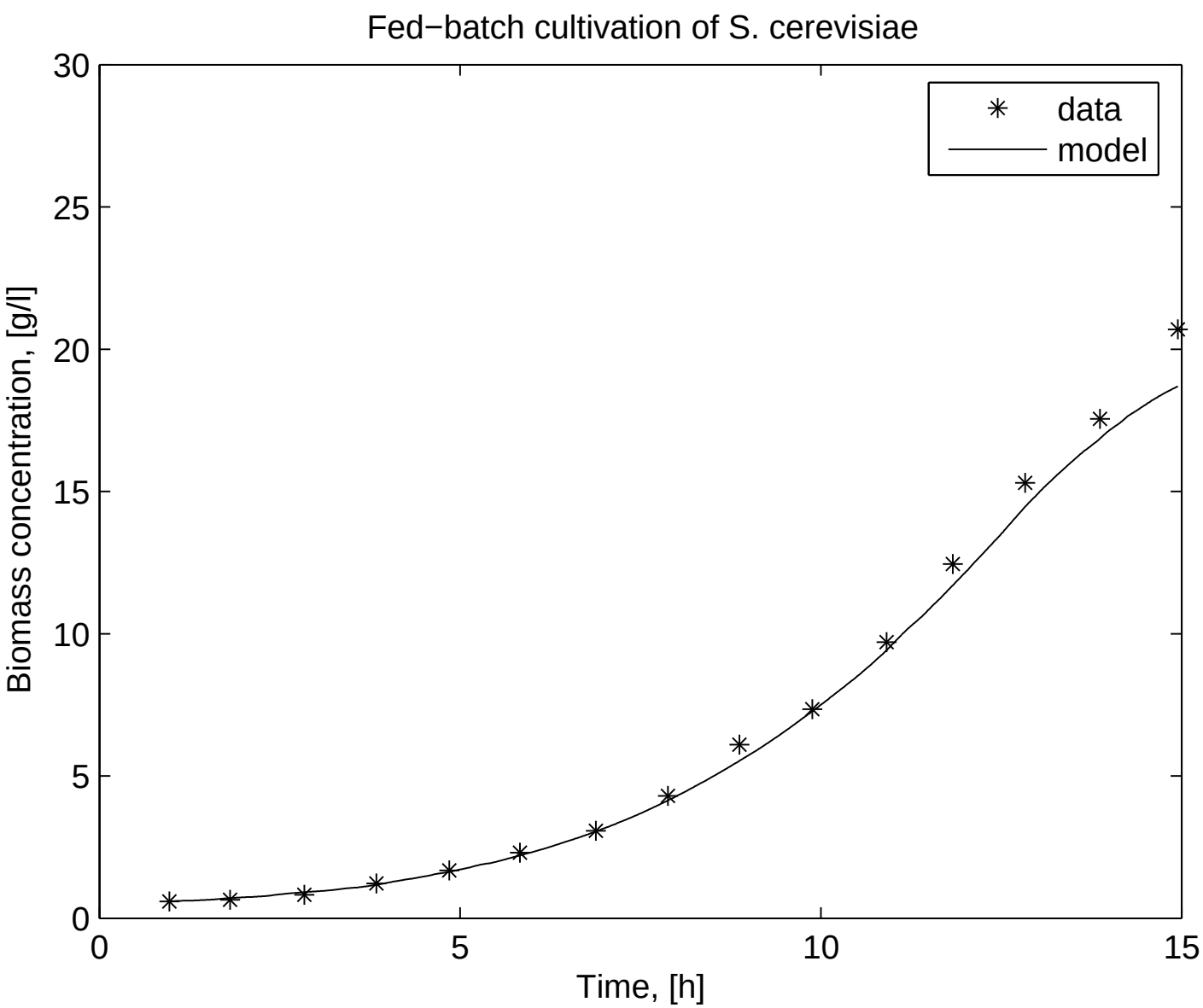
Corresponding Author: Dr.. Tania Pencheva, Ph.D.

Corresponding Author's Institution: Institute of Biophysics and Biomedical Engineering

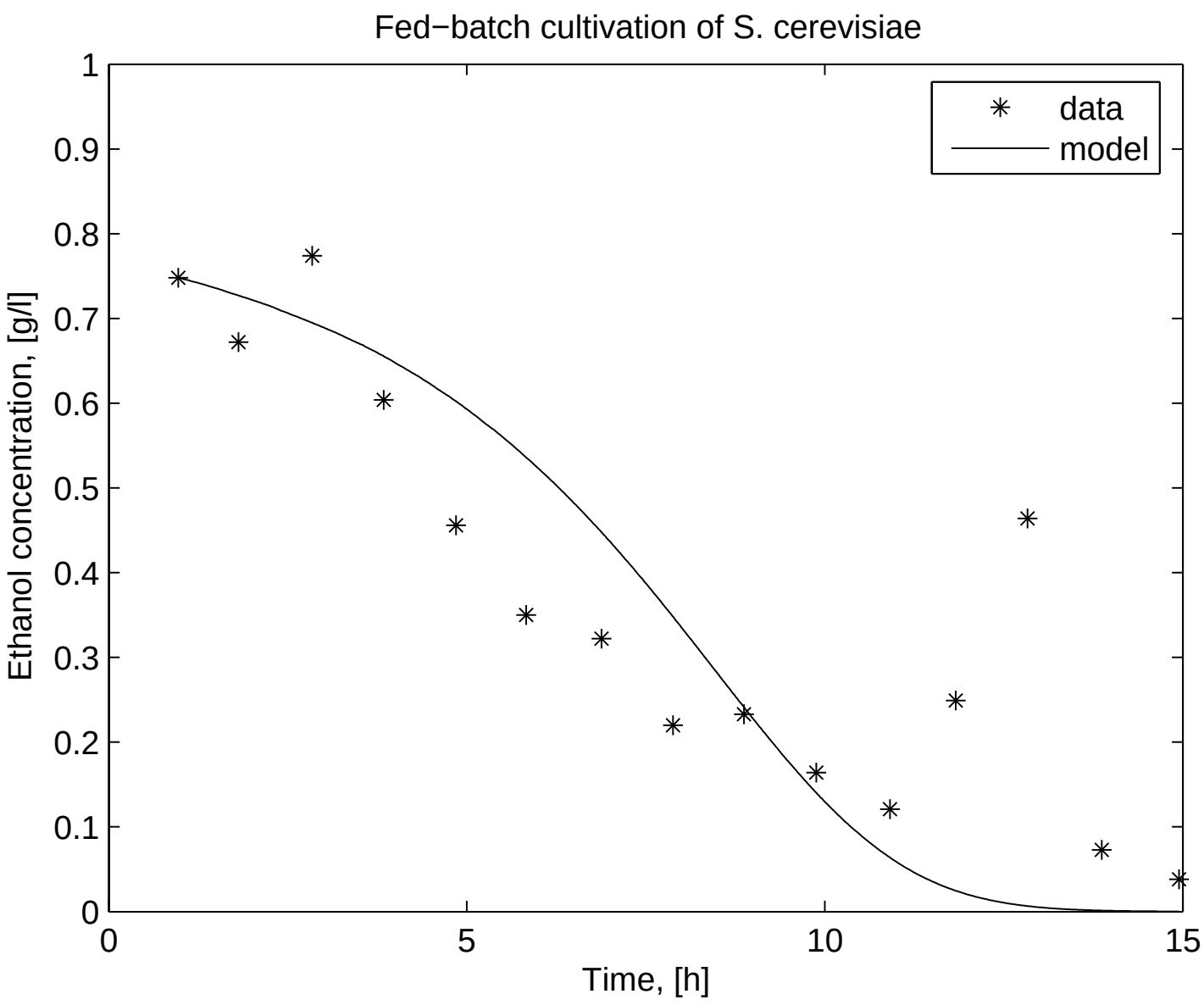
First Author: Maria Angelova

Order of Authors: Maria Angelova; Krassimir Atanassov, D. Sc.; Tania Pencheva, Ph.D.

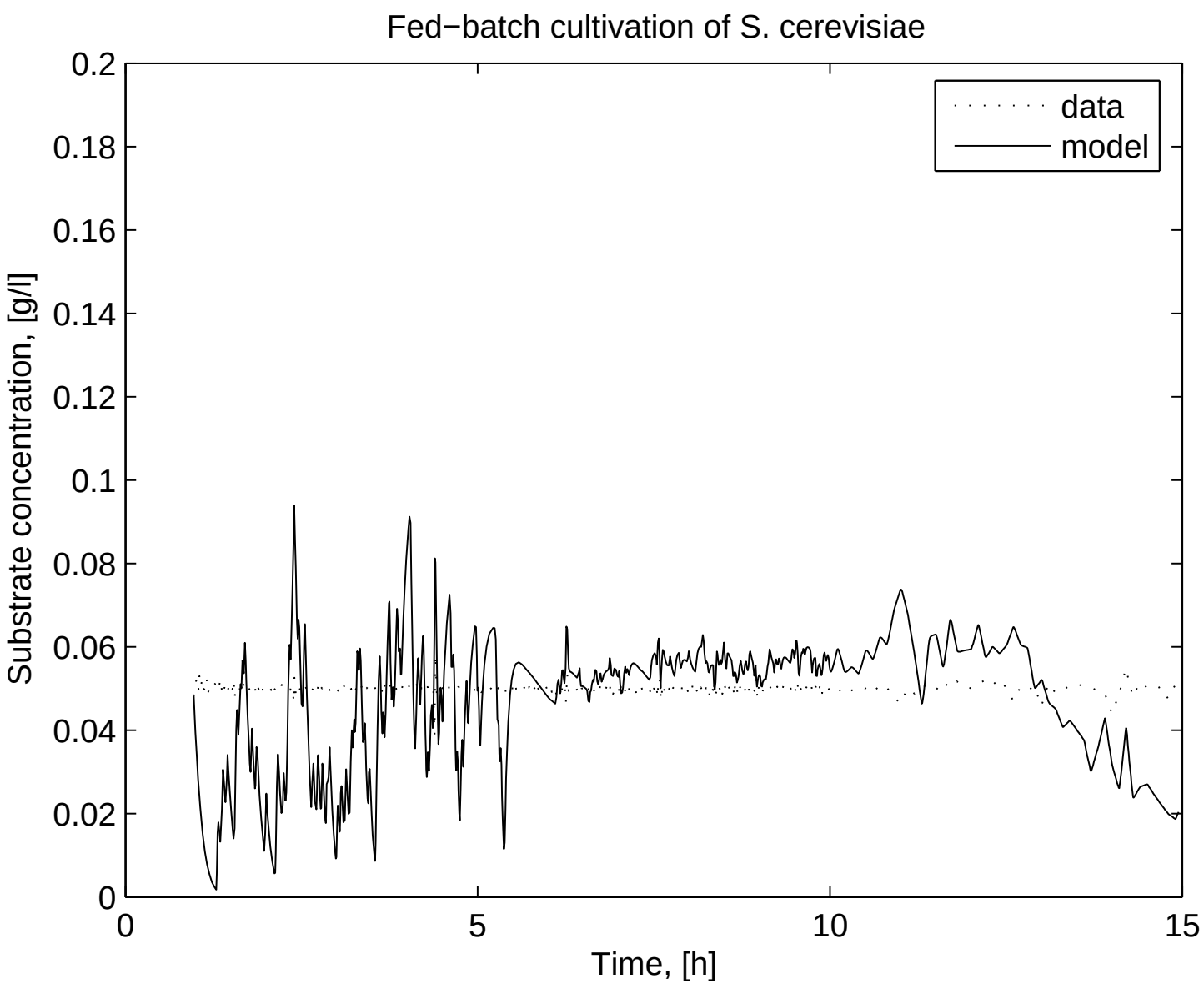
Figure


Fed−batch cultivation of S. cerevisiae
data
model
Biomass concentration, [g/l]
Time, [h]
0
5
10
15
20
25
30

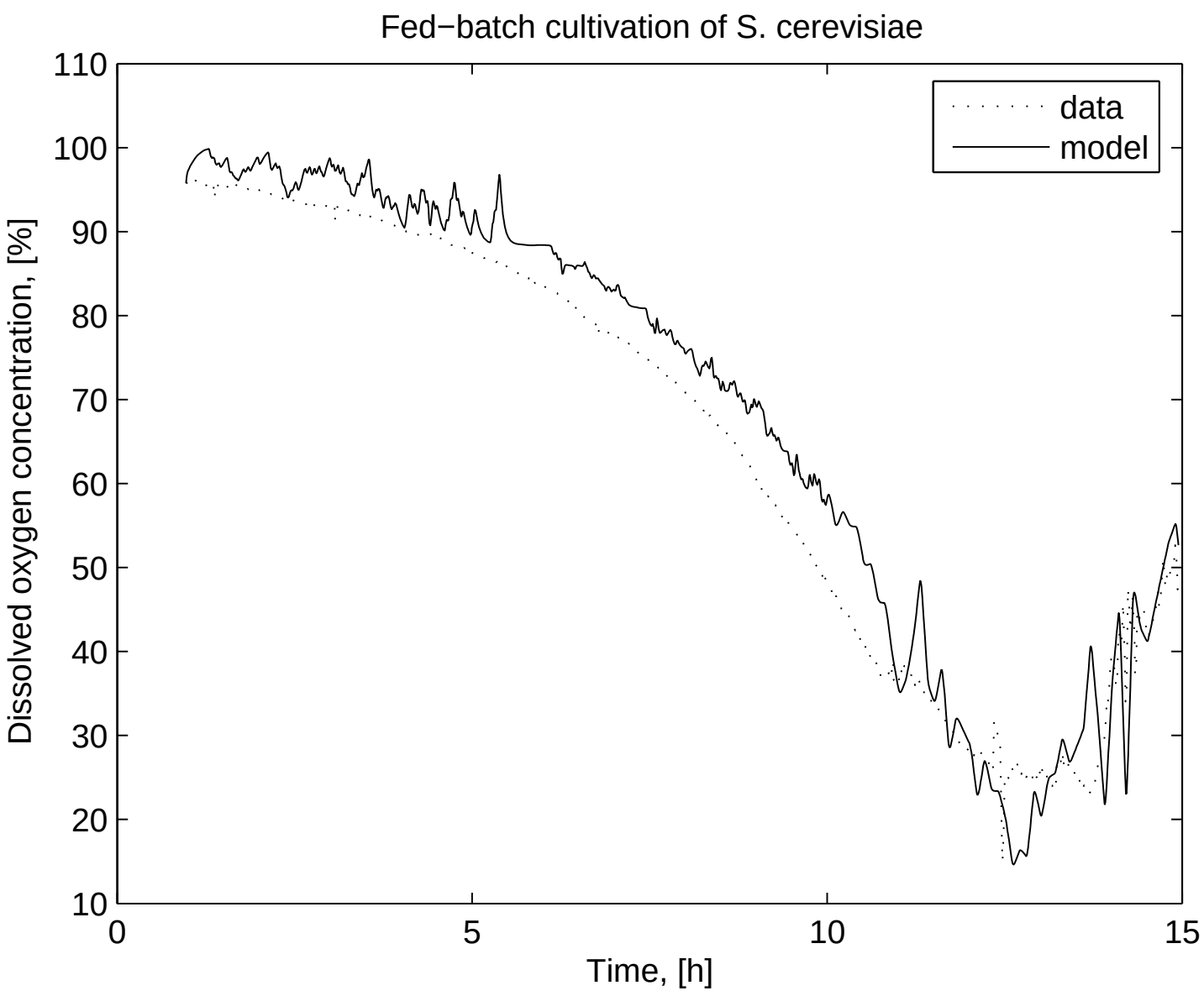
Figure


Fed−batch cultivation of S. cerevisiae
data
model
Ethanol concentration, [g/l]
Time, [h]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
5
10
15

Figure


Fed−batch cultivation of S. cerevisiae
data
model
Substrate concentration, [g/l]
Time, [h]
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
0
5
10
15

Figure


Fed−batch cultivation of S. cerevisiae
Dissolved oxygen concentration, [%]
Time, [h]
data
model
110
100
90
80
70
60
50
40
30
20
10
0
5
10
15

Purposeful Model Parameters Genesis in Simple Genetic Algorithms

Maria Angelova[a], Krassimir Atanassov[a], Tania Pencheva[a,]*

Institute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences
105 Acad. Georgi Bonchev Str., 1113 Sofia, Bulgaria
E-mails: maria.angelova@biomed.bas.bg, krat@bas.bg, tania.pencheva@biomed.bas.bg

* Corresponding author

Abstract: Simple genetic algorithms have been investigated aiming to improve the algorithm convergence time. Because of the stochastic nature of genetic algorithms, several runs have to be performed in order representative results to be achieved. A procedure for purposeful genesis concerning intervals of variations of model parameters is proposed for a standard simple genetic algorithm aiming to improve significantly the algorithm effectiveness. Such stepwise methodology is applied to a parameter identification of a fed-batch cultivation of *S. cerevisiae*. The procedure is further validated to a modified simple genetic algorithm with changed sequence of main genetic algorithm operators, namely mutation, crossover and selection proven as faster than the standard one. Results obtained from both applications show significantly improvement of algorithm convergence time saving the model accuracy.

Keywords: genetic algorithms, purposeful genesis, model parameters, fermentation process, *Saccharomyces cerevisiae.*

Introduction

Application of fermentation processes (FP) in different branches of industry makes their investigation a very topical question. Modelling and further optimal control of FP is a non-trivial task due to they are complex, dynamic systems with interdependence and time-varying process variables. An important step for FP adequate modeling is the choice of a certain optimization procedure for model parameter identification. Inability of conventional optimization methods such as Nelder-Mead's minimization, sequential quadratic programming, quasi-Newton algorithms (i.e. Broyden, Fletcher, Goldfarb and Shanno), etc. to reach to a satisfied solution [1,

2] provokes an idea some stochastic algorithms to be applied. As an alternative for solving such a complex problem evolutionary algorithms can be considered. Among them genetic algorithms (GA) [3], inspired by Darwin's theory of "survival of the fittest", is a stochastic global optimization technique with applications in different areas [4, 5, 6]. Properties like hard problems solving, noise tolerance, easy to interface and hybridize make GA suitable and more workable for parameter identification of fermentation models [2, 4, 6-9]. Obtained promising results using GA encourage their future investigation.

The effectiveness of a certain optimization technique can be evaluated by the model accuracy achieved and the convergence time needed. Due to the stochastic nature of GA obtained results might be quite diverse. That is why several runs have to be performed in order representative results to be reached. Accumulating data from different runs, an idea for purposeful genesis concerning intervals of variations of model parameters has been provoked. Such idea is going to be elaborated for standard simple genetic algorithms and further promptly applied for a modified one, previously proven as faster one [7].

Standard simple genetic algorithm (SGA), originally presented in [3], is here denoted as SGA-SCM, coming from an execution of main genetic operators in a sequence selection, crossover, mutation. In SGA-SCM chromosomes (coded parameter set) representing better possible solutions according to their own objective function values are chosen through selection from the population. After that, crossover proceeds in order to form new offspring. Mutation is then applied with determinate probability aiming to prevent falling of all solutions in the population into a local optimum of the solved problem. Many modifications of SGA-SCM, which differ from each other in the sequence of execution of those operators, have been elaborated aiming to improve the algorithm convergence time [7, 8]. Among them is SGA-MCS (mutation, crossover, selection), proposed and basically investigated in [7], in which selection operator has been processed after performing of crossover and mutation, in order to be prevented the loss of reached good solution by either crossover or mutation or both operators. After the reproduction, the SGA-MCS calculates the objective function for the offspring and the best fitted individuals from the offspring are selected to replace the parents, according to their objective function values. In both considered here SGA the calculations stop when a certain number of generations have been performed.

The aim of the study is simple genetic algorithms to be applied to a purposeful model parameters genesis for parameter identification of *S. cerevisiae* fed-batch cultivation. Aiming that a stepwise procedure is elaborated for SGA-SCM and further validated to SGA-MCS.

Procedure for Purposeful Model Parameters Genesis

Due to stochastic nature of GA a great number of algorithms runs have to be executed in order to be obtained reliable results in parameter identification of a fermentation process model. In the beginning GA searches for solutions of model parameters in wide, but reasonably chosen boundaries according to [10] statements. When results have been analyzed they showed that the values of model parameters can be assembled and preliminary defined boundaries could be straitened. Thus it has been generated an idea for developing a procedure of purposeful model parameters genesis. Such procedure is going to result in a definition of more appropriate boundaries for variation of model parameters values aiming to decrease convergence time while saving model accuracy.

The proposed procedure consists of six steps given further in details:

Step 1: Performance of *N* runs of genetic algorithms

Step 2: Determination of minimum and maximum of the objective function

Step 3: Determination of top level (TL), middle level (ML) and low level (LL) of performance with corresponding low boundary (LB) and up boundary (UB) following the scheme:

3.1. Determination of discrimination number Δ by

$$\Delta = \frac{maxJ - minJ}{3}$$

3.2.

Top level low boundary (TL_LB)	$min\ J$
Top level up boundary (TL_UB)	$min\ J + \Delta - \varepsilon$
Middle level low boundary (ML_LB)	$min\ J + \Delta$
Middle level up boundary (ML_UB)	$min\ J + 2\Delta - \varepsilon$
Low level low boundary (LL_LB)	$min\ J + 2\Delta$
Low level up boundary (LL_UB)	$max\ J$

where ε is a small number, ensuring the difference between levels.

Step 4: Determination of minimum, maximum and average value for each parameter at each level

Step 5: Based on averaged values, determination of new intervals of model parameters variations

Step 6: Run of genetic algorithm with intervals, determined in *Step 5*.

Model Parameters Genesis for *S. cerevisiae* Fed-batch Cultivation

Experimental data of *S. cerevisiae* fed-batch cultivation is obtained in the *Institute of Technical Chemistry – University of Hannover*, *Germany* [2]. The cultivation of the yeast *S. cerevisiae* is performed in a 1.5 l reactor, using a Schatzmann medium. Glucose in feeding solution is 50 g/l. The temperature was controlled at 30°C, the pH at 5.7. The stirrer speed was set to 500 rpm. 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 [2]:

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

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

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

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

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

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_{O2} – specific growth/utilization rates of biomass, substrate, ethanol and dissolved oxygen, [1/h]. All functions are continuous and differentiable.

Considered here fed-batch cultivation of *S. cerevisiae* is characterized with keeping glucose concentration equal to or below to its critical level (S_{crit} = 0.05 g/l), sufficient dissolved oxygen $O_2 \geq O_{2crit}$ (O_{2crit} = 18%) and availability of ethanol in the broth. This state corresponds to the so called *mixed oxidative state* (FS II) according to functional state modeling approach [2]. Hence, specific rates in Eqs. (1)-(5) are:

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

where μ_{2S}, μ_{2E} are the 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]; all model parameters fulfill the non-zero division requirement.

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

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

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

Developed procedure for a purposeful model genesis has been applied to parameter identification of *S. cerevisiae* fed-batch cultivation. The values of GA parameters and type of genetic operators in both considered here GA are tuned according to [7, 9]. Parameter identification of the model (1)-(6) has been performed using *Genetic Algorithm Toolbox* [11] in *Matlab 7* environment. All the computations are performed using a PC Intel Pentium 4 (2.4 GHz) platform running Windows XP.

Following model equations (1)-(6) altogether nine parameters for *S. cerevisiae* fed-batch cultivation model have been estimated firstly applying standard simple genetic algorithm, denoted here as SGA-SCM. As presented in Table 1, the algorithm has been investigated for four different values of the most sensitive towards algorithm convergence time parameter, namely the generation gap (GGAP) [9]. For each value of GGAP several runs of SGA-SCM have been executed. Obtained results are analyzed according to achieved objective function values. For each GGAP value minimum and maximum of the objective function are determined and the discrimination number is assigned according to *Step 3* of the procedure. After that the determination of top, middle and low level of performance is going on. The best results hit the interval [*min J*; *min J* + Δ – ε]. These ones classified in the middle level have an objective function varying in terms [*min J* + Δ; *min J* + 2Δ – ε]. The worst solutions for the objective function fall in the interval [*min J* + 2Δ; *max J*].

For each of the constructed in such a way levels, the minimum, maximum and average value of each model parameter have been determined. Table 2 presents these values only for the top levels according to Table 1.

Table 1

SGA_SCM	**Objective function**		**Levels of performance**		**Average convergence time**
GGAP = 0.9	*min J*	0.0221	TL_LB	0.0221	81.6693
			TL_UB	~0.0222	
	avrg J	0.0222	ML_LB	0.0222	
			ML_UB	0.0222	
	max J	0.0223	LL_LB	0.0222	
			LL_UB	0.0223	
GGAP = 0.8	*min J*	0.0221	TL_LB	0.0221	70.3386
			TL_UB	~0.0222	
	avrg J	0.0222	ML_LB	0.0222	
			ML_UB	0.0222	
	max J	0.0223	LL_LB	0.0222	
			LL_UB	0.0223	
GGAP = 0.67	*min J*	0.0221	TL_LB	0.0221	59.1654
			TL_UB	~0.0222	
	avrg J	0.0222	ML_LB	0.0222	
			ML_UB	0.0222	
	max J	0.0223	LL_LB	0.0222	
			LL_UB	0.0223	
GGAP = 0.5	*min J*	0.0222	TL_LB	0.0222	46.9997
			TL_UB	~0.0223	
	avrg J	0.0223	ML_LB	0.0223	
			ML_UB	~0.0225	
	max J	0.0226	LL_LB	0.0225	
			LL_UB	0.0226	

Table 2

SGA-SCM		μ_{2S}	μ_{2E}	k_S	k_E	Y_{SX}	Y_{EX}	$k_L^{O_2}a$	Y_{OS}	Y_{OE}
GGAP = 0.9	*min*	0.94	0.14	0.13	0.80	0.40	2.00	40.42	333.03	35.73
	max	0.99	0.15	0.14	0.80	0.39	1.81	95.53	785.10	96.73
	avrg	0.97	0.14	0.13	0.80	0.39	1.92	63.24	515.78	61.78
GGAP = 0.8	*min*	0.91	0.12	0.12	0.79	0.41	1.84	52.20	415.38	102.70
	max	1.00	0.14	0.13	0.80	0.40	1.63	126.44	990.90	201.39
	avrg	0.95	0.13	0.13	0.80	0.40	1.78	96.82	769.11	155.37
GGAP = 0.67	*min*	0.91	0.12	0.11	0.80	0.41	1.98	48.98	379.71	30.46
	max	0.97	0.15	0.14	0.80	0.39	1.67	126.78	987.88	149.45
	avrg	0.95	0.14	0.13	0.80	0.40	1.84	101.39	801.98	92.02
GGAP = 0.5	*min*	0.91	0.11	0.11	0.79	0.40	2.04	99.59	768.66	102.84
	max	1.00	0.15	0.14	0.80	0.39	1.47	126.78	983.37	261.13
	avrg	0.95	0.14	0.13	0.80	0.40	1.84	108.41	853.07	216.27

The new boundaries of model parameters are constructed following that the new minimum is the minimum between the averages, and the new maximum is the maximum between the averages. Table 3 presents previously used "wide" boundaries for each model parameter according to [10] as well as new boundaries proposed based on the procedure for purposeful model parameter genesis.

Table 3

SGA-SCM		μ_{2S}	μ_{2E}	k_S	k_E	Y_{SX}	Y_{EX}	$k_L^{O_2}a$	Y_{OS}	Y_{OE}
previously used	**LB**	0.9	0.05	0.08	0.5	0.1	0.1	0.001	0.001	0.001
	UB	1	0.15	0.15	0.8	3	10	1000	1000	300
advisable after procedure application	**LB**	0.94	0.13	0.12	0.7	0.38	1.7	60	500	50
	UB	0.97	0.15	0.14	0.8	0.42	2.5	120	900	220

Investigated SGA-SCM has been again applied involving newly proposed boundaries at GGAP = 0.5. Several runs have been performed to obtain reliable results. Table 4 presents the average values of objective function, calculation time and model parameters.

Table 4

Parameter/ GGAP = 0.5	**SGA-SCM**	
	before procedure application	after procedure application
J	0.0222	0.0221
CPU time, s	46.9997	29.25
μ_{2S}, 1/h	0.95	0.95
μ_{2E}, 1/h	0.14	0.14
k_S, g/l	0.13	0.13
k_E, g/l	0.80	0.80
Y_{SX}, g/g	0.40	0.40
Y_{EX}, g/g	1.84	1.87
$k_L^{O_2}a$, 1/h	108.41	77.21
Y_{OS}, g/g	853.07	608.53
Y_{OE}, g/g	216.27	212.56

Applied procedure leads to saving up to 38% of computational time without loss of model accuracy, thus showing good effectiveness of proposed procedure for purposeful model parameter genesis.

Further proposed procedure is verified when applied to the one of the modified SGA, namely SGA-MCS [7] since it has been distinguished as faster than SGA-SCM. Observed tendency in SGA-SCM that the lower value of GGAP ensures saving the model accuracy for less computational time has been confirmed in SGA-MCS. Up to almost 44% of calculation time is saved in case of SGA-MCS application using GGAP = 0.5 instead of 0.9 with a such model

accuracy. Applying again proposed here procedure for purposeful model parameters genesis for SGA-MCS, new boundaries for model parameters are proposed (Table 5).

Table 5

SGA-MCS		μ_{2S}	μ_{2E}	k_S	k_E	Y_{SX}	Y_{EX}	$k_L^{O_2}a$	Y_{OS}	Y_{OE}
advisable after procedure application	**LB**	0.9	0.1	0.11	0.75	0.4	1	60	500	80
	UB	0.93	0.13	0.13	0.8	0.43	2	120	800	220

New boundaries applicability has been tested in several runs of SGA-MCS aiming to obtain reliable results. Table 6 presents the average values of objective function, calculation time and model parameters when SGA-MCS has been executed at GGAP = 0.5. Applied procedure for model parameter genesis leads to almost 30% reduction of computational time of SGA-MCS without loss of model accuracy.

Table 6

Parameter/ GGAP = 0.5	**SGA-MCS**	
	before procedure application	after procedure application
J	0.0224	0.0221
CPU time, s	40.5261	28.6090
μ_{2S}, 1/h	0.92	0.91
μ_{2E}, 1/h	0.12	0.11
k_S, g/l	0.12	0.11
k_E, g/l	0.79	0.80
Y_{SX}, g/g	0.40	0.42
Y_{EX}, g/g	1.58	1.51
$k_L^{O_2}a$, 1/h	90.23	69.23
Y_{OS}, g/g	708.44	550.75
Y_{OE}, g/g	203.38	194.26

As it could be seen from Tables 4 and 6 the value of the objective function is further reduced when proposed here procedure is applied to modified SGA-MCS. This algorithm is accurate as SGA-SCM, even more for considerably less CPU time. Due to the similarity of the results achieved, here only the promising results applying SGA-MCS are presented. Fig. 1 shows results from experimental data and model prediction respectively for biomass, ethanol, substrate and dissolved oxygen.

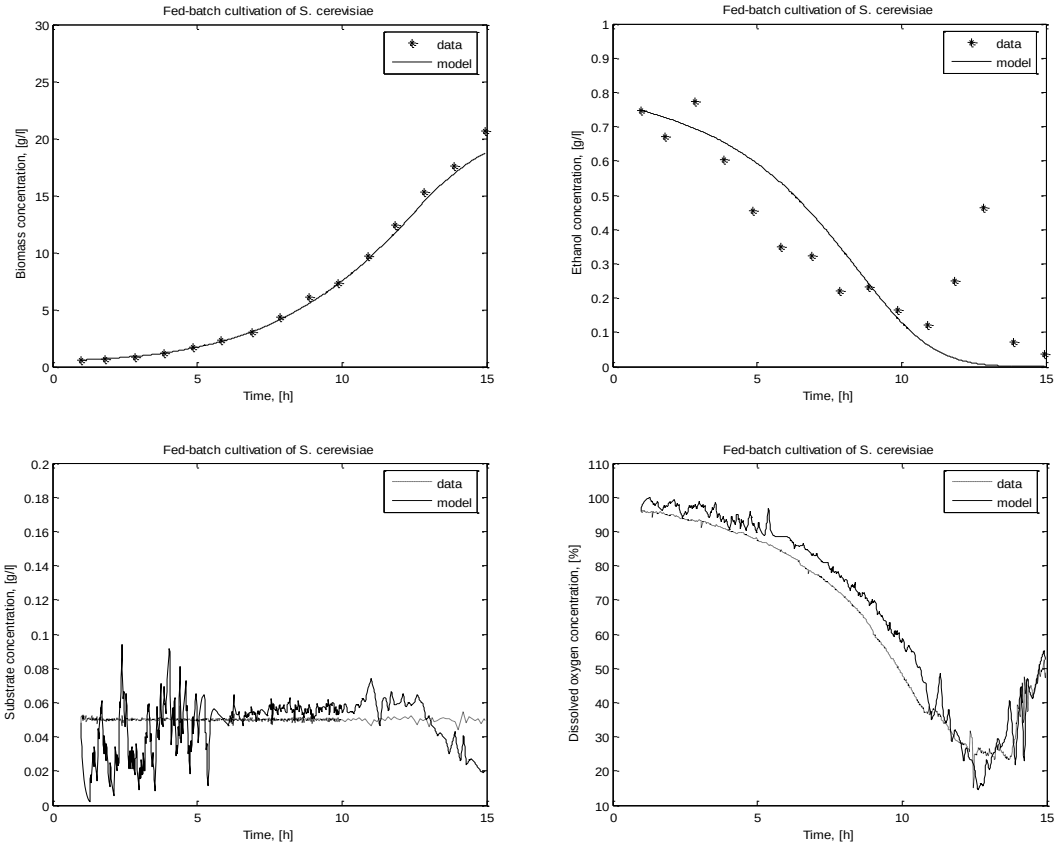


Fig. 1 Model prediction compared to experimental data

Obtained results show the workability of the proposed procedure for purposeful model parameters genesis and its effectiveness when applied to SGA-SCM and SGA-MCS.

Analysis and Conclusions

In this investigation a stepwise procedure for purposeful model parameters genesis for parameter identification of *S. cerevisiae* fed-batch cultivation has been elaborated and further applied towards two kinds of simple genetic algorithms. SGA-SCM, as a standard GA and mostly examined, has been used as a test example. Obtained promising results have been confirmed when proposed procedure has been validated in modified SGA-MCS with exchanged sequence of genetics operators, namely mutation, crossover and selection. Applying consequently proposed stepwise procedure in such algorithms, intervals of variations of model parameters have been straitened thus the GA performance has been improved. After the application of proposed procedure SGA-SCM has become 38% and SGA-MCS 29% faster, thus showing significantly improvement of algorithm convergence time saving the model accuracy. But if one compares obtained results for SGA-SCM when GGAP = 0.9 to these when GGAP = 0.5 in SGA-MCS with applied procedure of purposeful model parameters genesis, it could be seen that up to 65% of calculation time are saved without affecting the model accuracy. It is worth to note,

that the proposed here procedure of purposeful model parameters genesis is universal remedy and could be appropriately and successfully implemented to another stochastic optimization algorithms, as well as to different objects of model parameter identification.

Acknowledgements

This work is partially supported by National Scientific Fund of Bulgaria, grant DID 02-29 and grant DMU 02/4.

References

[1] I. Hristozov, T. Pencheva, B. Iliev, St. Tzonkov, A comparative analysis of genetic algorithms and conventional optimization procedures, Scient. Proc. of Scient.-Techn. Union of Mech. Eng. V(3-25) (1998) 187-192 (in Bulgarian).

[2] T. Pencheva, O. Roeva, I. Hristozov, Functional State Approach to Fermentation Processes Modelling, Tzonkov St., B. Hitzmann (Eds.), Prof. Marin Drinov Academic Publishing House, Sofia, 2006.

[3] D. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wiley Publishing Company, Massachusetts, 1989.

[4] J. Adeyemo, A. Enitian, Optimization of fermentation processes using evolutionary algorithms – a review, Scient. Res. and Ess. 6(7) (2011) 1464-1472.

[5] P. Vasant, N. Barsoum, Hybrid genetic algorithms and line search method for industrial production planning with non-linear fitness function, Eng. Appl. of Art. Intell. 22 (2009) 767-777.

[6] K. Jones, Comparison of genetic algorithms and particle swarm optimization for fermentation feed profile determination, Proc. CompSysTech'2006, Veliko Tarnovo, Bulgaria, June 15-16, 2006, IIIB.8-1–IIIB.8-7.

[7] M. Angelova, S. Tzonkov, T. Pencheva, Genetic algorithms based parameter identification of yeast fed-batch cultivation, Lect. Notes Comp. Sci. 6046 (2011) 224-231.

[8] O. Roeva, A modified genetic algorithm for a parameter identification of fermentation processes, Biotechn. & Biotechn. Equipm. 20 (2006) 202-209.

[9] M. Angelova, T. Pencheva, Tuning genetic algorithm parameters to improve convergence time, Int. J. Chem. Eng. (2011) Article ID 646917, doi:10.1155/2011/646917, available at http://www.hindawi.com/journals/ijce/2011/646917/cta/

[10] K. Schuegerl, K.-H. Bellgardt (Eds.), Bioreaction Engineering, Modeling and Control, Springer-Verlag, Berlin Heidelberg New York, 2000.

[11] A. J. Chipperfield, P. Fleming, H. Pohlheim, C. M. Fonseca, Genetic algorithm toolbox for use with MATLAB, User's guide, version 1.2. Dept. of Automatic Control and System Engineering, University of Sheffield, UK, 1994.