

Development of a Generalized Net for Comparison of Different Models Obtained using Metaheuristic Algorithms

Dimitar Dimitrov¹, Olympia Roeva²

¹ Faculty of Mathematics and Informatics, Sofia University
5 James Bourchier Str., Sofia, Bulgaria
e-mail: dgdimitrov@fmi.uni-sofia.bg

² Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
e-mail: olympia@biomed.bas.bg

Abstract: The present paper proposes a developed generalized net model that compares different mathematical models of the process of *E. coli* fed-batch cultivation. A system of four ordinary differential equations describes the main variables of the considered cultivation process. For the purposes of model simulation, we use the software package GNLite. The GN model compares the simulated performance of the proposed set of mathematical models and selects the best performing one on the basis of a predefined criterion. During the simulation, GNLite calls the Matlab software environment to solve the process model, presented as a system of nonlinear differential equations, and plots the dynamics of the main process variables' of the model that has been computed to perform best.

Keywords and phrases: Generalized net, GNLite, GNTicker, GN IDE, *E. coli*, Fed-batch cultivation, Mathematical models.

2000 Mathematics Subject Classification: 03E72.

1 Introduction

The problem of model parameter estimation refers to situations where decisions have to be made depending on multiple conflicting criteria,

oftentimes involving imprecision or uncertainty. A wide range of optimization methods are available for solving such problems, each featuring specific properties making it appropriate for specific cases. Choosing an optimization method very much depends on the nature of the problem and the availability of appropriate software that suits the problem statement.

Up to now Generalized nets (GNs) are used as a tool for modelling of parallel processes in many areas as economics, transport, medicine, computer technologies etc. The theory of the GNs [1] proved to be quite successful when applied to the description of the functioning of expert systems, machine learning and different technological processes [6, 7]. Recently attempts have been made to describe various models, optimization procedures, control loops considering bioprocesses, based on GNs [10].

In this paper the apparatus of GNs is used to describe a GN model for comparison of different mathematical models of an *E. coli* fed-batch cultivation process. Proposed here GN model is an extension of presented in Dimitrov and Roeva [3]. The new model is more complex including equation for acetate dynamics and thus two more model parameters. The GN model is simulated using GNLite – a software package for modeling and simulation with GNs [11]. GNLite implements the current state of the most aspects of GN theory, compared to other software tools for GNs [4]. Its graphical environment has user-friendly interface which assists the user through the whole modeling process. Some of the features of GNLite are the visual (WYSIWYG) editing and simulation of GN models, the usage of XML format to define the GN models, the support of several programming language for predicates and characteristic functions, as well as the supported integration of Matlab code into GN models.

2 GNLite Software

GNLite is a component oriented software product for running simulations of generalized net models. It has a client-server architecture and its core is the software interpreter of GN models GNTicker [11]. Compared to previous implementations of GN simulators, this approach has multiple advantages like the ensured platform independence, not restricting developers of new GN applications to particular operating systems or programming languages. Applications on the client side, such as visual environments for graphic editing and user input, communicate with the simulation core via TCP/IP connection and the specially designed the GNTicker Trace Protocol.

Together with GNTicker, the GNTicker Characteristic Function Language programming language has been developed. The language can code the GN transition predicates, characteristic functions, rules for tokens splitting and merging, as well as additional utility functions. GNTicker has been integrated to call functions written in Matlab environment, thus enabling the performance of complex computations, like solving differential equations, during the GN simulation [2].

GNLite features an integrated modeling environment, called GN IDE, consisting of graphical user interface providing facilities for visual building of GN models, running simulations and tracing their execution, and many more. GN IDE can be run on any platform with Java Runtime Edition installed, as well as in any browser using Java Web Start.

So far, the operating systems, GNLite has been tested on, are Ubuntu Linux and Microsoft Windows (32- and 64-bit versions).

3 *E. coli* Fed-batch Cultivation Process Model

The considered fed-batch cultivation could be described according to the mass balance as follows:

$$\frac{dX}{dt} = \mu_{max} \frac{S}{S + k_s} X - \frac{F}{V} X \quad (1)$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}} \mu_{max} \frac{S}{S + k_s} X + \frac{F}{V} (S_{in} - S) \quad (2)$$

$$\frac{dA}{dt} = -\frac{1}{Y_{A/X}} \mu_{max} \frac{A}{A + k_A} A + \frac{F}{V} A \quad (3)$$

$$\frac{dV}{dt} = F \quad (4)$$

where X – biomass concentration, [g/l]; S – substrate concentration, [g/l]; A – acetate concentration, [g/l]; F – influent flow rate, [l-1]; V – bioreactor volume, [l]; S_{in} – influent glucose concentration, [g/l]; μ_{max} – maximum specific growth rate, [h-1]; $Y_{S/X}$, $Y_{A/X}$ – yield coefficients; k_s , k_A – saturation constants, [g/l].

We are solving the model parameter estimation problem, using real experimental data of the *E. coli* MC4110 fed-batch cultivation process [9].

We have taken into consideration several different optimization techniques, namely, Genetic algorithms (GA), Tabu search (TS) and Firefly algorithm (FA), advancing some previous authors' research with mathematical models based on GA [9], TS [5] and FA [8]. The considered optimization techniques are applied to the mathematical model (1)-(4), assigned with the following values of the parameters:

- GA: $\mu_{max} = 0.52 \text{ h}^{-1}$, $k_S = 0.023 \text{ g/l}$, $k_A = 0.59 \text{ g/l}$, $1/Y_{S/X} = 0.5$, $1/Y_{A/X} = 0.013$;
- TS: $\mu_{max} = 0.49 \text{ h}^{-1}$, $k_S = 0.03 \text{ g/l}$, $1/Y_{S/X} = 0.5$ according [6], and $k_A = 0.46$, $1/Y_{A/X} = 0.01$ identified in this research;
- FA: $\mu_{max} = 0.46 \text{ h}^{-1}$, $k_S = 0.013 \text{ g/l}$, $k_A = 0.18 \text{ g/l}$, $1/Y_{S/X} = 0.5$, $1/Y_{A/X} = 0.03$;
- Test model: $\mu_{max} = 2.4 \text{ h}^{-1}$, $k_S = 6 \text{ g/l}$, $k_A = 0.18 \text{ g/l}$, $1/Y_{S/X} = 0.5$, $1/Y_{A/X} = 0.03$.

The fourth set of model parameters are defined in a way to test if the considered parameters belong to the defined range.

The model choice is done on the basis of the evaluation of a distance measure J :

$$J = \sum_{i=1}^n \sum_{j=1}^m \left\{ \left[y_{exp}(i) - y_{mod}(i) \right]_j \right\}^2 \rightarrow \min \quad (5)$$

where n – number of measurements for each state variable; m – number of state variables; y_{exp} – experimental data; y_{mod} – model predicted data.

4 Developed Generalized Net Model

In Figure 1 the graphical structure of the constructed GN is shown.

Here is its formal definition.

$$E = \langle \langle \{Z_1, Z_2, Z_3\} \rangle, \langle \{\alpha, \beta\} \rangle, \langle \{x_0^\alpha, x_0^\beta\} \rangle, \Phi \rangle,$$

where the three transitions are described consequently, as follows

$$Z_1 = \langle \{l_1\}, \{l_3, l_4, l_5, l_6\}, \frac{l_3 \quad l_4 \quad l_5 \quad l_6}{l_1 \mid \begin{array}{cccc} true & true & true & true \end{array}}, l_1 \rangle$$

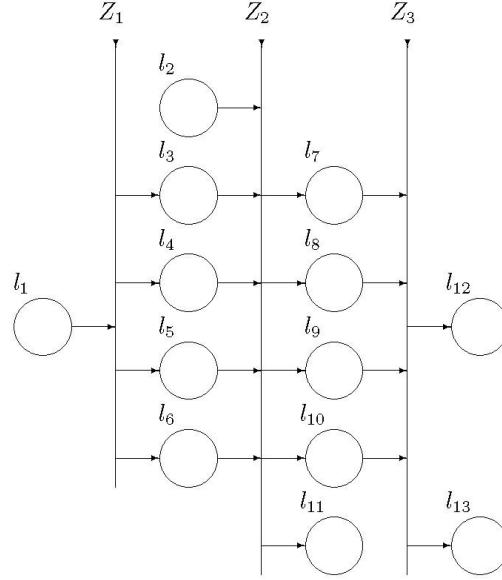


Figure 1. Graphical structure of the GN model

Token α enters place l_1 with four sets of parameter values as initial characteristic: $(\mu_{max_j}, k_{s_j}, k_{A_j}, 1/Y_{S/X_j}, 1/Y_{A/X_j})$. Token β enters place l_2 with values for $t, S_{in}, V_0, X_0, S_0, A_0, F, X_d, S_d$ and A_d .

After the token α is split into four tokens, characteristic function $\Phi_{l_i}, 3 \leq i \leq 6$ keeps only $(i - 2)$ -th set of parameter values. Token β is split into four tokens too. $\Phi_{l_i}, 7 \leq i \leq 10$ merges α and β tokens in its corresponding position. Then it calls a Matlab function named *simul* with the following declaration:

$$function[J, X_{mod}, S_{mod}, A_{mod}] = simul(t, \mu_{max}, k_s, k_A, 1/Y_{S/X}, \\ 1/Y_{A/X}, S_{in}, V_0, X_0, S_0, A_0, F, X_d, S_d, A_d).$$

$$Z_2 = \langle \{l_2, l_3, l_4, l_5, l_6\}, \{l_7, l_8, l_9, l_{10}, l_{11}\},$$

	l_7	l_8	l_9	l_{10}	l_{11}
l_2	$check(l_3)$	$check(l_4)$	$check(l_5)$	$check(l_6)$	$false$
l_3	$check(l_3)$	$false$	$false$	$false$	$\neg check(l_3)$
l_4	$false$	$check(l_4)$	$false$	$false$	$\neg check(l_4)$
l_5	$false$	$false$	$check(l_5)$	$false$	$\neg check(l_5)$
l_6	$false$	$false$	$false$	$check(l_6)$	$\neg check(l_6)$

$$\wedge (l_2, l_3, l_4, l_5, l_6)),$$

where $check(l)$ is a predicate that checks if the parameters in the token at the given place l are correct, as described in the previous section:

$$check(l) = pr_1 x^{tokenAt(l)} \in [0, 0.9] \wedge pr_2 x^{tokenAt(l)} \in [0, 2] \wedge pr_3 x^{tokenAt(l)} \in [0, 1],$$

where $tokenAt(l)$ returns the token at place l .

$$Z_3 = \langle \{l_7, l_8, l_9, l_{10}\}, \{l_{12}, l_{13}\},$$

	l_{12}	l_{13}
l_7	$isMinError(l_7)$	$\neg isMinError(l_7)$
l_8	$isMinError(l_8)$	$\neg isMinError(l_8)$
l_9	$isMinError(l_9)$	$\neg isMinError(l_9)$
l_{10}	$isMinError(l_{10})$	$\neg isMinError(l_{10})$

$$\vee (l_7, l_8, l_9, l_{10})),$$

where $isMinError(l)$ is a predicate that checks whether the error characteristic of the token at the given place l is minimal among the characteristics of the four tokens.

Token splitting and merging must be enabled for the net, i.e. operators $D(2,1)$ and $D(2,4)$ must be defined for E .

For the simulation of the presented here model, is used the GNLite software package, run on Windows 32-bit operating system.

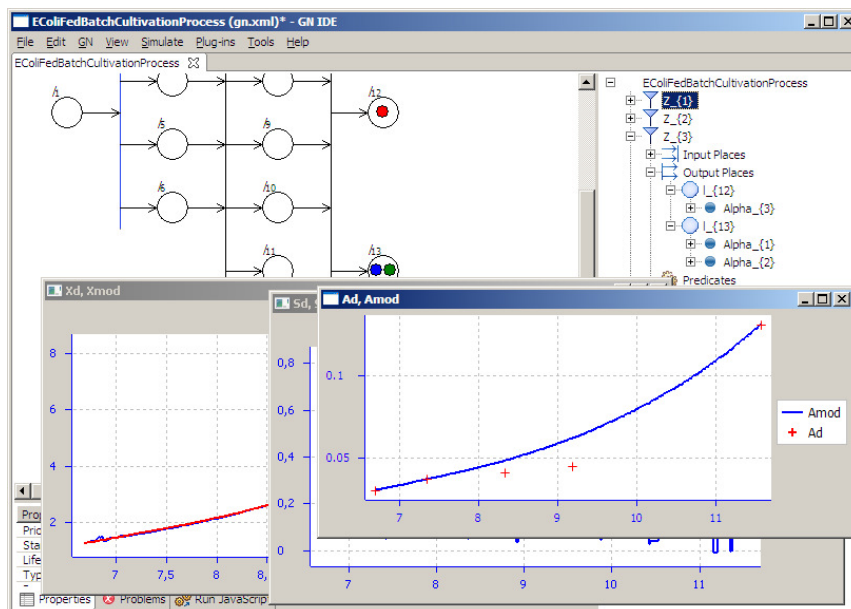
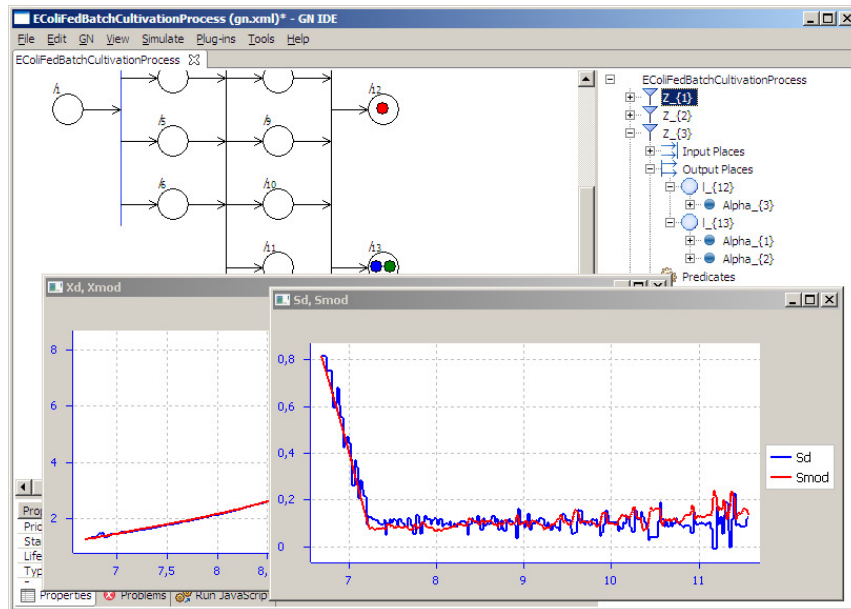
A token α enters the GN model, having as an initial characteristic the four mathematical models, presented in Section 3. In the first transition, token α splits into four tokens corresponding to each model. In the

simulation, each new token is given a distinctive color. The main initial characteristics of the *E. coli* fed-batch cultivation process are defined in token β that enters the net via the input place l_2 . In the next transition, Z_2 , the four models are checked for correctness according to the parameter intervals defined. Models that have passed the check are moved to the corresponding output place (l_7, l_8, l_9, l_{10}). The rest tokens leave the net through place l_{11} . In this case, the first three models (see Section 3) have passed the check and the fourth model leaves the net through place l_{11} .

Token β splits for each output place corresponding to a valid model. After transition Z_2 , the tokens in places l_7 to l_{10} merge together. For each token in $l_i, 7 \leq i \leq 10$, the characteristic function Φ calls function *simul* from the Matlab environment, that solves the defined nonlinear system of differential equations for each set of model parameters, i.e. $x^{\alpha_{i-2}} = pr_{i-2} x^\alpha$. The function returns four values: $J, X_{\text{mod}}, S_{\text{mod}}$ and A_{mod} , which are added to the current characteristics of the corresponding tokens. Depending on the values of the parameter J , in transition Z_3 is chosen the token with minimal error, and it is transferred to place l_{12} , while the rest tokens proceed to place l_{13} . The token in l_{12} contains the mathematical model of the considered process that has exhibited the highest degree of accuracy. In this case, it is the first model (obtained using GA):

$$\mu_{\max} = 0.52 \text{ h}^{-1}, k_S = 0.023 \text{ g/l}, k_A = 0.59 \text{ g/l}, 1/Y_{S/X} = 0.5, 1/Y_{A/X} = 0.013.$$

GN IDE gives possibility for plotting chart diagrams, showing the values of a selected set of characteristics, as illustrated on Figs. 2-3, where the dynamics of the main process variables are shown. Data from real experiments is compared to the data, as predicted by the model. The presented graphical results show a very good correlation between the real and modelled process.



4 Conclusion

In the present research, a generalized net model has been developed for comparing different optimization models of the *E. coli* fed-batch cultivation process. For comparison of the models, a set of resulting mathematical models parameters, applying different optimization techniques, is used. The presented GN model compares the models obtained by a GA, TS algorithm and FA, and chooses the best process model on the basis of the defined criterion. For the purpose of model simulation of the proposed GN model, we use the developed software package GNLite, which implements the current state-of-the-art of GN theory and its graphical environment has user-friendly interface which assists the user through the whole modelling process. During the simulation, the GN model calls the Matlab software environment and in the end, the dynamics of the main process variables of the model are plotted with the highest degree of accuracy.

References

This work has been supported by the National Science Fund under Grant Ref. No. DN-02-10/2-16 “New Instruments for Knowledge Discovery from Data, and Their Modelling”.

References

- [1] Atanassov, K., *On Generalized Nets Theory*, Prof. M. Drinov Academic Publishing House, Sofia, 2007.
- [2] Dimitrov, D. G., Integration of the Generalized Nets Simulator GNTicker with MATLAB, *Annual of "Informatics" Section*, Union of Scientists in Bulgaria, Vol. 4, 2011, 49-52, (In Bulgarian).
- [3] Dimitrov, D. G., & Roeva, O. Comparison of Different Mathematical Models of an *E. coli* Fed-batch Cultivation Process Using Generalized Net Model, *Proc. of the 13th Int. Workshop on Generalized Nets*, 29 October 2012, London, UK, 2012, 15-23.
- [4] Dimitrov, D. G., Software Products Implementing Generalized Nets, *Annual of "Informatics" Section*, Union of Scientists in Bulgaria, Vol. 3, 2010, 37–50, (In Bulgarian).

- [5] Kosev, K., Trenkova, T., & Roeva, O. Tabu Search for Parameter Identification of an Fermentation Process Model, *J. of Int. Scientific Publication: Materials, Methods & Technologies*, Vol. 6, No 2, 2012, 457-464.
- [6] Krawczak, M., A Novel Modeling Methodology: Generalized Nets, In: *Artificial Intelligence and Soft Computing*, A. Cader, L. Rutkowski, R. Tadeusiewicz and J. Zurada, Eds., Publishing House Exit, Warsaw, 2006, 1160-1168.
- [7] Krawczak, M., Generalized net models of MLNN learning algorithms, *Lecture Notes in Computer Science*, Vol. 2, No 3697, 2005, 25-30.
- [8] Roeva, O., Optimization of *E. coli* Cultivation Model Parameters using Firefly Algorithm, *Int. J. Bioautomation*, Vol. 16, No 1, 2012, 23-32.
- [9] Roeva, O., Parameter Estimation of a Monod-type Model based on Genetic Algorithms and Sensitivity Analysis, *Lecture Notes on Computer Science*, Springer-Verlag Berlin Heidelberg, Vol. 4818, 2008, 601-608.
- [10] Shannon, A., Roeva, O., Pencheva, T., & Atanassov, K. *Generalized Nets Modelling of Biotechnological Processes*, Prof. M. Drinov Academic Publishing House, Sofia, 2004.
- [11] Trifonov, T., Georgiev, K., & Atanassov, K. Software for modelling with Generalized Nets, *Issues in Intuitionistic Fuzzy Sets and Generalized Nets*, Wydawnictwo Wsisiz, Warsaw, Vol. 6, 2008, 96-110.