Optimization of *E. coli* Cultivation Model Parameters Using Firefly Algorithm

Olympia Roeva

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

Received: December 21, 2011 **Accepted: April 6, 2012**

Published: April 30, 2012

Abstract: *In this paper, a novel meta-heuristics algorithm, namely the Firefly Algorithm (FA), is adapted and applied for a model parameter identification of an E. coli fed-batch cultivation process. A system of ordinary nonlinear differential equations is used to model the biomass growth and substrate utilization. Parameter optimization is performed using real experimental data set from an E. coli MC4110 fed-batch cultivation process. The FA adjustments are done based on several pre-tests according to the optimization problem considered here. The simulation results indicate that the applied algorithm is effective and efficient. As a result, a model with high degree of accuracy is obtained applying the FA.*

Keywords: *Optimization, Firefly algorithm, E. coli cultivation, Identification, Model parameters.*

Introduction

Modelling approaches are central in system biology and provide new ways towards the analysis and understanding of cells and organisms. A common approach to model cellular dynamics is by using sets of nonlinear differential equations. Real parameter optimization of cellular dynamics models has become a research field of particularly great interest. Such problems have widespread application.

Development of adequate models is an important step for process optimization and high-quality control. For many industrial relevant processes, however, detailed models are not available due to insufficient understanding of the underlying phenomena. The mathematical models, which could naturally be incomplete and inaccurate to a certain degree, can still be very useful and effective tools in describing those effects which are of great importance for control, optimization, or understanding of the process. Numerical solution of the models is the basis for the development of economic and powerful methods in the fields of bioprocess design, plant design, scale-up, optimization and bioprocess control [13].

Parameter identification of a nonlinear dynamic model is more difficult than that of a linear one, as no general analytic results exist. The difficulties that may arise are, for instance, convergence to local solutions if standard local methods are used, over-determined models, badly scaled model function, etc. Due to the nonlinearity and constrained nature of the considered systems, these problems are very often multimodal. Thus, traditional gradient-based methods may fail to identify the global solution. Although a lot of different global optimization methods exist, the efficacy of an optimization method is always problem-specific.

While searching for new, more adequate modeling metaphors and concepts, methods which draw their initial inspiration from nature have received the early attention. During the last decade, a broad class of meta-heuristics has been developed and applied to a variety of areas. The three most well-known heuristics are the iterative improvement algorithms, the probabilistic optimization algorithms, and the constructive heuristics. Recently, a new meta-heuristics called Firefly Algorithm (FA) algorithm has emerged. This algorithm was proposed by Xin-She Yang [15]. According to recent bibliography [15-18], the FA is very efficient and can outperform other meta-heuristics, such as genetic algorithms, for solving many optimization problems. Although the FA has many similarities with other swarm intelligence based algorithms, such as Particle Swarm Optimization, Artificial Bee Colony Algorithm, and Bacterial Foraging Algorithm, it is indeed much simpler both in concept and implementation [17, 18]. There are already several applications of FA for different optimization problems. Based on bibliography results, it is evident that the FA is a powerful novel population-based method for solving optimization problems. Authors in [8] present application of FA for moving peaks benchmark problem. In [1], FA is applied to solving the economic emissions load dispatch problem. A method based on FA for scheduling jobs on grid computing is presented in [19]. Authors in [3] adapt FA to find optimal solutions of noisy non-linear continuous mathematical model.

This paper aims to introduce, for the first time, application of the FA in the field of mathematical modeling of bioprocesses. An optimization algorithm based on FA is proposed for parameter identification of an *E. coli* fed-batch cultivation process.

Cultivation of recombinant micro-organisms, e.g. *E. coli*, in many cases is the only economical way to produce pharmaceutical biochemicals such as interleukins, insulin, interferons, enzymes and growth factors. Research on *E. coli* has accelerated even more since 1997, when its entire genome was published. As knowledge of *E. coli* grows, scientists are starting to build models of the microbe that captures some of its behavior. Some of recent researches and developed models are presented in [4-6, 9, 10, 14].

Problem formulation

Mathematical model of E. coli fed-batch cultivation process

Application of the general state space dynamical model to the *E. coli* cultivation fed-batch process leads to the following nonlinear differential equation system [11]:

$$\frac{dX}{dt} = \mu_{max}\frac{S}{k_S + S}X - \frac{F}{V}X\,, \tag{1}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\mu_{max}\frac{S}{k_S + S}X + \frac{F}{V}\left(S_{in} - S\right), \tag{2}$$

$$\frac{dV}{dt} = F\,, \tag{3}$$

where: X is the biomass concentration, [g·l^{-1}]; S – substrate (glucose) concentration, [g·l^{-1}]; F – influent flow rate, [h^{-1}]; V – bioreactor volume, [l]; S_{in} – influent glucose concentration, [g·l^{-1}]; μ_{max} – maximum specific growth rate, [h^{-1}]; $Y_{S/X}$ – yield coefficient, [g·g^{-1}]; k_S – saturation constant, [g·l^{-1}].

The following assumptions are made in the model development of the *E. coli* fed-batch cultivation:

- The bioreactor is completely mixed.
- The substrate glucose is mainly consumed oxidatively and its consumption can be described by Monod kinetics.
- Variations in the growth rate and substrate consumption do not significantly change the elemental composition of biomass, thus balanced growth conditions are only assumed.

Objective function

The optimization criterion is a certain factor, whose value defines the quality of an estimated set of parameters. Parameter estimation problem of the presented nonlinear dynamic system is stated as the minimization of the distance measure *J* between the experimental and the model predicted values of the considered state variables:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{k} \{[\mathbf{y}_{exp}(i) - \mathbf{y}_{mod}(i)]_j\}^2 \to min\,, \tag{4}$$

where n is the length of data vector for each state variable k; $\mathbf{y}_{exp}$ are known experimental data; $\mathbf{y}_{mod}$ are model predictions.

For parameter identification procedure, real experimental data from an *E. coli* MC4110 fed-batch cultivation process are used. The cultivation experiments are performed in the Institute of Technical Chemistry, University of Hannover, Germany, during the collaboration work with the Institute of Biophysics and Biomedical Engineering, BAS, Bulgaria, funded by DFG. A detailed description of the cultivation conditions is presented in [2, 12].

Firefly algorithm

The Firefly Algorithm is a new meta-heuristic algorithm which is inspired from flashing light behaviour of fireflies in nature. The pattern of flashes is often unique for a particular species of fireflies. Two basic functions of such flashes are to attract mating partners or communicate with them, and to attract potential victim. Additionally, flashing may also serve as a protective warning mechanism.

The flashing light can be formulated in such a way that it is associated with the objective function to be optimized. This makes it possible to formulate new meta-heuristic algorithms idealizing some of the flashing characteristics of fireflies. According to [15], the FA uses three idealized rules for simplification the algorithm:

- All fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex.
- Attractiveness is proportional to their brightness, thus for any two flashing fireflies, the less bright one will move towards the brighter one. The attractiveness and brightness both decrease as the distance between them increases. If there is no brighter one than a particular firefly, it will move randomly.
- The brightness of a firefly is affected or determined by the landscape of the objective function. For a maximization or minimization problem, the brightness can simply be proportional to the value of the objective function.

Based on these three idealized rules [15], the basic steps of the FA can be summarized as the pseudo code, presented in Fig. 1.

```
begin
Define
  light absorption coefficient γ
  initial attractiveness β0
  randomization parameter α
  objective function f(x), where x = (x1, ..., xd)^T
Generate initial population of fireflies xi (i = 1, 2, ..., n)
Determine light intensity Ii at xi via f(xi)
   while (t < MaxGeneration) do
      for i = 1 : n all n fireflies do
         for j = 1 : i all n fireflies do
            if (Ij > Ii) then
            Move firefly i towards j based on Eq. (7)
            end if
            Attractiveness varies with distance r via exp[−γr^2]
            Evaluate new solutions and update light intensity
         end for j
      end for i
      Rank the fireflies and find the current best
   end while
   Postprocess results and visualization
end begin
```

Fig. 1 Pseudo code of FA

In the FA, there are two important issues of the variation of light intensity and the formulation of the attractiveness. For simplicity, it is assumed that the attractiveness of a firefly is determined by its brightness, which in turn is associated with the encoded objective function of the optimization problems.

In this algorithm, each firefly has a location $x = (x_1, ..., x_d)^T$ in a d-dimensional space and light intensity $I(x)$ or attractiveness $\beta(x)$, which are proportional to an objective function $f(x)$. Attractiveness $\beta(x)$ and light intensity $I(x)$ are relative and these should be judged by the rest fireflies. Thus, attractiveness will vary with the distance $r_{i,j}$ between firefly i and firefly j. So attractiveness β of a firefly can be defined by Eq. (5) [16-18]:

$$\beta(r) = \beta_0 e^{-\gamma r^2}, \tag{5}$$

where r or $r_{i,j}$ is the distance between the i-th and j-th of two fireflies. β_0 is the initial attractiveness at $r = 0$ and γ is a fixed light absorption coefficient that controls the decrease of the light intensity.

The initial solution is generated based on:

$$x_j = rand(Ub - Lb) + Lb, \tag{6}$$

where *rand* is a random number generator uniformly distributed in the space [0, 1]; *Ub* and *Lb* are the upper range and lower range of the *j*-th firefly, respectively.

When firefly *i* is attracted to another more attractive firefly *j*, its movement is determined by:

$$x_{i+1} = x_i + \beta_0 e^{-\gamma r_{i,j}^2}(x_i - x_j) + \alpha(rand - \frac{1}{2}), \tag{7}$$

where the first term is the current position of a firefly, the second term is used for considering a firefly attractiveness to light intensity seen by adjacent fireflies $\beta(r)$ (Eq. (5)), and the third term is used to describe the random movement of a firefly in case there are no brighter ones. The coefficient α is a randomization parameter determined by the problem of interest. The distance $r_{i,j}$ between any two fireflies *i* and *j* at $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively, is defined as a Cartesian or Euclidean distance, according to [16-18]:

$$r_{i,j} = \left\| x_i - x_j \right\| = \sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2}, \tag{8}$$

where $x_{i,k}$ is the *k*-th component of the spatial coordinate $\mathbf{x}_i$ of the *i*-th firefly.

Results and discussion

A series of parameter identification procedures for the considered model Eq. (1) - (3), using FA, are performed. Computer specifications to run all optimization procedures are Intel® Core™i5-2320 CPU @ 3.00GHz, 8 GB Memory (RAM), Windows 7 (64bit) operating system.

In order to increase the performance of the FA, it is necessary to provide the adjustments of the parameters depending on the problem domain. The parameters of the FA are tuned based on several pre-tests according to the parameter identification problem considered here. The following ranges of parameters variation are explored [15-18]:

$\beta_0 \in [0, 10]$, $\gamma \in [0.1, 100]$, $\alpha \in [0, 1]$.

After tuning procedures, the main FA parameters are set to the optimal settings:

$\beta_0 = 1$, $\gamma = 1$, $\alpha = 0.2$, number of fireflies = 10, number of iterations = 100.

Because of the stochastic characteristics of the applied algorithm, FA has been run at least 50 times so as to carry out meaningful statistical analysis. The best and the mean results of the parameters estimates, total time for the solver to run (CPU time) and objective function value *J* are observed. The obtained results are summarized in Table 1.

The parameter estimates presented in Table 1 are adequate and meet the physical sense they bear. The estimates of the yield of glucose per biomass, maximum specific growth rate and saturation constant using considered search methods are similar to the results reported in [7, 11-12, 20]. As the results indicate, the applied algorithm performs very well. The obtained parameters best results are very closer to the mean ones.

Table 1. Identified model parameters

Model parameters	Estimates	
	best	mean
μ_{max}	0.4864	0.4859
k_S	0.0170	0.0166
$1/Y_{S/X}$	2.0361	2.0363
J	6.0259	6.1572
CPU time	170.9927	201.1274

The obtained criteria values show that the developed model is adequate and has a high degree of accuracy. A graphical representation of the convergence of the objective function with time (in logarithmic scale) is shown on Fig. 2. As it can be seen, the FA shows a good performance in terms of speed of convergence. A stable convergence of objective function of the given problem is observed with time. From the simulation results, it seems that the proper selection of the number of fireflies, the number of iterations, and the absorption coefficient is of paramount importance for the convergence of the algorithm as this heavily depends on the nature of the applied problem.

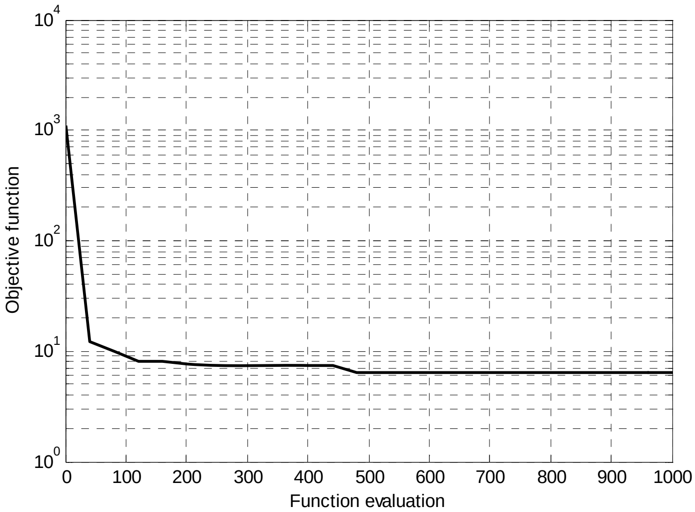

Fig. 2 Convergence of the objective function with function evaluations

In Fig. 3, the parameter estimations through algorithm iterations are presented. It is shown that the algorithm stable sets the model parameters values to the optimal ones in a half of function evaluation.

In most cases, graphical comparisons clearly show the presence or absence of systematic deviations between model predictions and measurements. It is evident that a quantitative measure of the differences between calculated and measured values is an important criterion for the adequacy of a model. The model predictions of the state variables, based on FA estimated set of model parameters, are compared to the experimental data points of the *E. coli* MC4110 cultivation in Fig. 4 and Fig. 5.

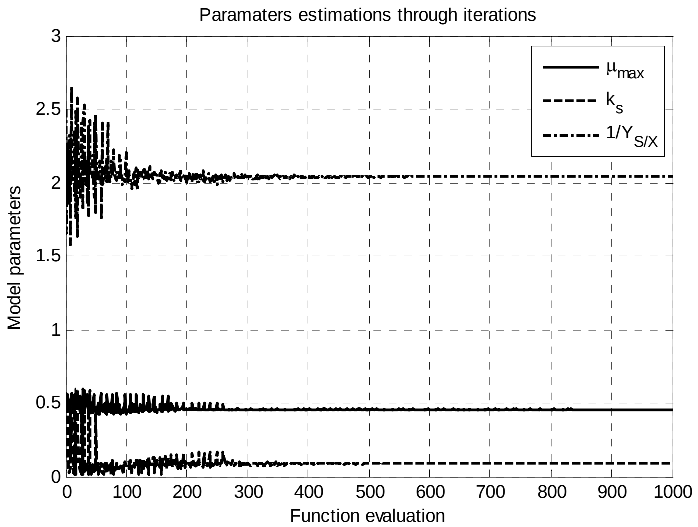

Fig. 3 FA parameter estimations with function evaluations

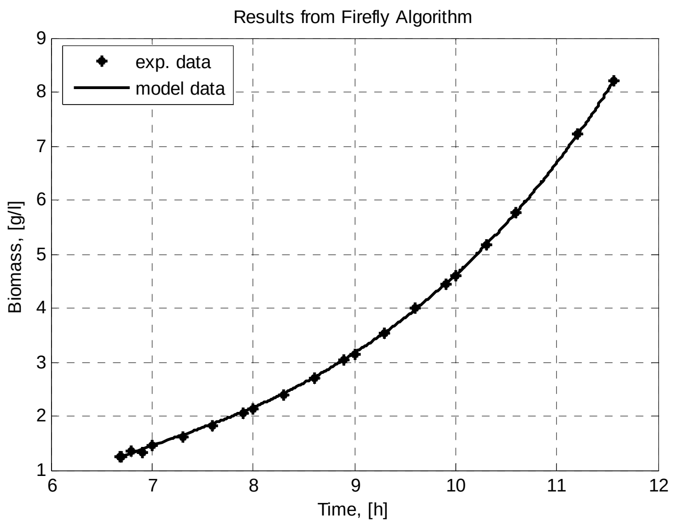

Fig. 4 Experimental and model predicted data for biomass dynamics

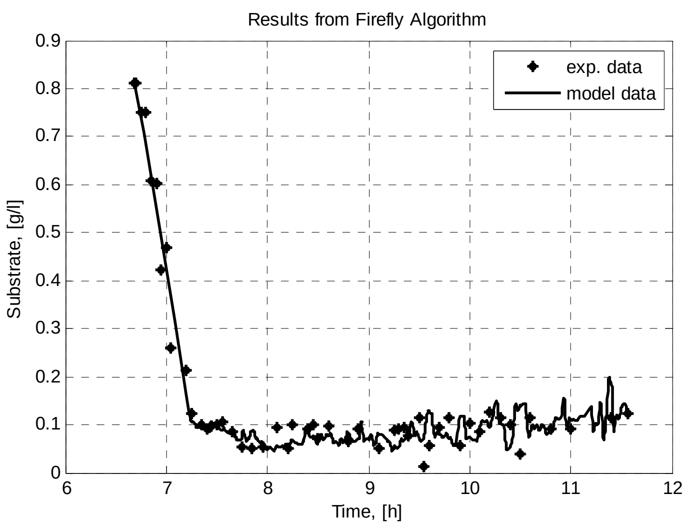

Fig. 5 Experimental and model predicted data for substrate dynamics

It is observed that there is coincidence between modelled and measured data for biomass and substrate concentrations. The model successfully predicts the dynamics of these process variables during the fed-batch cultivation of *E. coli* MC4110. As it can be seen from Fig. 4, the obtained model perfectly predicts the dynamics of biomass during the fed-batch process. In the identification procedure, raw experimental data are used without any preprocessing. Due to some specific peculiarities of the on-line measurement system during the cultivation [2], the data for the substrate concentration are very noisy (Fig. 5). Nevertheless, the model successfully follows the substrate data trend. Thus, it can be concluded that the obtained model predicts adequate the dynamics of glucose during the process.

Conclusion

In this paper, the recently developed Firefly Algorithm is proposed and tested for application to the parameter identification of a nonlinear dynamical model of *E. coli* MC4110 cultivation process. The identification procedure is formulated as an optimization problem. The mathematical model is presented by a system of ordinary differential equations, describing the considered process variables – biomass and substrate. Numerical and simulation results reveal that correct and consistent results can be obtained using the Firefly Algorithm. As a result, an adequate, high-quality mathematical model of *E. coli* MC4110 fed-batch cultivation process is obtained. The results confirm that the Firefly Algorithm is powerful and efficient tool for identification of the parameters in the non-linear dynamic model of cultivation processes.

In future works, it is intended to compare Firefly Algorithm with other population based methods, so as to identify the strengths and weaknesses of the current meta-heuristic algorithm and prove its powerfulness and effectiveness in solving bioprocess model parameter optimization problem.

Acknowledgements

The investigations are partially supported by the Bulgarian National Science Fund, Grants DID 02/29 “Modeling Processes with Fixed Development Rules (ModProFix)” and DMU 02/4 “High Quality Control of Biotechnological Processes with Application of Modified Conventional and Metaheuristics Methods”.

References

1. Apostolopoulos T., A. Vlachos (2011). Application of the Firefly Algorithm for Solving the Economic Emissions Load Dispatch Problem, International Journal of Combinatorics, Article ID 523806.
2. Arndt M., B. Hitzmann (2001). Feed Forward/feedback Control of Glucose Concentration during Cultivation of *Escherichia coli*, 8th IFAC Int Conf on Comp Appl in Biotechn, Canada, 425-429.
3. Chai-ead N., P. Aungkulanon, P. Luangpaiboon (2011). Bees and Firefly Algorithms for Noisy Non-linear Optimisation Problems, Proceedings of International MultiConference of Engineers and Computer Scientists – IMECS, 2, 1449-1454.
4. Covert M. W., N. Xiao, T. J. Chen, J. R. Karr (2008). Integrating Metabolic, Transcriptional Regulatory, and Signal Transduction Models in *Escherichia coli*, Bioinformatics, 24(18), 2044-2050.
5. Jiang L., Q. Ouyang, Y. Tu (2010). Quantitative Modeling of *Escherichia coli* Chemotactic Motion in Environments Varying in Space and Time, PLoS Comput Biol, 6(4), e1000735, doi:10.1371/journal.pcbi.1000735.

6. Karelina T. A., H. Ma, I. Goryanin, O. V. Demin (2011). EI of the Phosphotransferase System of *Escherichia coli*: Mathematical Modeling Approach to Analysis of Its Kinetic Properties, Journal of Biophysics, 2011, Article ID 579402, doi:10.1155/2011/579402.
7. Levisauskas D., V. Galvanauskas, S. Henrich, K. Wilhelm, N. Volk, A. Lubbert (2003). Model-based Optimization of Viral Capsid Protein Production in Fed-batch Culture of Recombinant *Escherichia coli*, Bioprocess and Biosystems Engineering, 25, 255-262.
8. Nasiri B., M. R. Meybodi (2012). Speciation-based Firefly Algorithm for Optimization in Dynamic Environments, Int J Artificial Intelligence, 8(S12), 118-132.
9. Opalka N., J. Brown, W. J. Lane, K.-A. F. Twist, R. Landick, F. J. Asturias, S. A. Darst (2010). Complete Structural Model of *Escherichia coli* RNA Polymerase from a Hybrid Approach, PLoS Biol, 8(9), e1000483, doi:10.1371/journal.pbio.1000483.
10. Petersen C. M., H. S. Rifai, G. C. Villarreal, R. Stein (2011). Modeling *Escherichia coli* and Its Sources in an Urban Bayou with Hydrologic Simulation Program – FORTRAN, Journal of Environmental Engineering, 137(6), 487-503.
11. Roeva O. (2008). Parameter Estimation of a Monod-type Model based on Genetic Algorithms and Sensitivity Analysis. Lecture Notes on Computer Science, Springer-Verlag Berlin Heidelberg, 4818, 601-608.
12. Roeva O., T. Pencheva, B. Hitzmann, St. Tzonkov (2004). A Genetic Algorithms Based Approach for Identification of *Escherichia coli* Fed-batch Fermentation, Int J Bioautomation, 1, 30-41.
13. Schuegerl, K., Bellgardt, K.-H. (2000). Bioreaction Engineering: Modeling and Control, Berlin Heidelberg, Springer-Verlag.
14. Skandamis P. N., G. E. Nychas (2000). Development and Evaluation of a Model Predicting the Survival of *Escherichia coli* O157:H7 NCTC 12900 in Homemade Eggplant Salad at Various Temperatures, pHs, and Oregano Essential Oil Concentrations, Applied and Environmental Microbiology, 66(4), 1646-1653.
15. Yang X. S. (2008). Nature-inspired Meta-heuristic Algorithms, Luniver Press, Beckington, UK.
16. Yang X. S. (2009). Firefly Algorithm for Multimodal Optimization, Lecture Notes in Computing Sciences, 5792, 169-178.
17. Yang X. S. (2010a). Firefly Algorithm, Levy Flights and Global Optimization, Research and Development in Intelligent Systems XXVI, Springer, London, UK, 209-218.
18. Yang X. S. (2010b). Firefly Algorithm, Stochastic Test Functions and Design Optimisation, International Journal of Bio-inspired Computation, 2(2), 78-84.
19. Yousif A., A. H. Abdullah, S. M. Nor, A. A. Abdelaziz (2011). Scheduling Jobs on Grid Computing using Firefly Algorithm, Journal of Theoretical and Applied Information Technology, 33(2), 155-164.
20. Zelic B., D. Vasic-Racki, C. Wandrey, R. Takors (2004). Modeling of the Pyruvate Production with *Escherichia coli* in a Fed-batch Bioreactor, Bioprocess and Biosystems Engineering, 26, 249-258.

Assoc. Prof. Olympia Roeva, Ph.D.
E-mail: olympia@biomed.bas.bg

Olympia Roeva received M.Sc. degree (1998) and Ph.D. degree (2007) from the Technical University – Sofia. At present she is an associate professor at the Institute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences. She has more than 80 publications, among those 10 books and book chapters, with more than 100 citations. Her current scientific interests are in the fields of modelling, optimization and control of biotechnological processes, meta-heuristic algorithms and generalized nets.