Cuckoo Search Algorithm for Model Parameter Identification

Olympia Roeva[*], Vassia Atanassova

Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str.
Sofia 1113, Bulgaria
E-mails: *olympia@biomed.bas.bg,*
vassia.atanassova@gmail.com

[*]*Corresponding author*

Abstract: *In this paper, the metaheuristics algorithm Cuckoo Search (CS), is adapted and applied for a model parameter identification of an E. coli fed-batch cultivation process. The dynamics of bacteria growth and substrate (glucose) utilization is described by a system of ordinary nonlinear differential equations. Using real experimental data set from an E. coli MC4110 fed-batch cultivation process a parameter optimization is performed. 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 CS. The simulation results and comparison with genetic algorithm and ant colony optimization algorithm confirm the effectiveness of the applied CS algorithm in solving a cultivation model parameter identification problem.*

Keywords: *Metaheuristic algorithm, Cuckoo search, E. coli cultivation, Parameter identification.*

Introduction

More and more metaheuristic algorithms inspired by animal behavior phenomena has received considerable attention among researchers in case of solving complex optimization problems [2, 20, 24]. Algorithms like genetic algorithms (GA) and evolution strategies, ant colony optimization (ACO) [5], artificial bee colony (ABC) optimization [7], bat algorithm (BA) [27], Firefly algorithm (FA) [22], particle swarm optimization (PSO) [9], etc. are among a broad class of meta-heuristics that have been developed. The so-called *nature-inspired* metaheuristic algorithms have been used in a wide range of optimization problems [23].

Parameter identification of a nonlinear dynamic model of a cultivation process is more difficult than that of a linear one, as no general analytic results exist. Some of the difficulties that may arise include: 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 cultivation process systems, these problems are very often multimodal. Thus, traditional gradient-based methods may fail to identify the global solution. Despite the availability of multiple various global optimization methods, the efficacy of the optimization method is always problem-dependent. In this case, nature-inspired optimization algorithms have received the early attention.

There are already several applications of metaheuristic algorithms to cultivation process modelling and control – GA [4, 13, 19], FA [15], ACO [17] and BA [16]. The published

results confirm that the metaheuristic algorithms are a powerful and efficient tool for identification of the parameters in non-linear dynamic models of cultivation processes and their control. However, there is still room for finding new, more adequate modeling metaphors and concepts.

Another optimization algorithm inspired by animal behavior phenomena is Cuckoo Search (CS). It takes as a metaphor the reproduction strategy of cuckoo species in the nature. CS is based on the interesting breeding behavior called brood parasitism that certain cuckoo species exhibit. The CS algorithm was proposed by Yang and Deb in 2009 [25]. Up to now, CS has been applied to solving many optimization problems [3, 10, 11, 28]. According to obtained results, CS is very efficient and can outperform other meta-heuristics, such as genetic algorithms [1, 26].

In this paper, CS algorithm, is applied for the first time in the field of mathematical modeling of bioprocesses. An optimization algorithm based on CS is proposed here for parameter identification of an *E. coli* MC4110 fed-batch cultivation process.

The paper is organized as follows. The problem formulation is given in Section 2. The Cuckoo search algorithm for parameter identification of cultivation processes is proposed in Section 3. The numerical results and a discussion are presented in Section 4. Conclusion remarks are done in Section 5.

Problem formulation

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 increasingly accelerated 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 [6, 8, 12, 14].

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 [15]:

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

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

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

$$\mu = \mu_{max} \frac{S}{k_S + S} , \tag{4}$$

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

In the model development of the *E. coli* fed-batch cultivation, the following assumptions are made: (i) The bioreactor is completely mixed; (ii) The substrate glucose is mainly consumed oxidatively and its consumption can be described by Monod kinetics; (iii) Variation 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

Parameter estimation problem of the presented non-linear dynamic system (1)-(4) is stated as the minimization of the distance measure *J* between the experimental and the model predicted values of the considered state variables:

$$J = \sum_{i=1}^{n} \left(X_{\exp}(i) - X_{\mathrm{mod}}(i) \right)^2 + \left(S_{\exp}(i) - S_{\mathrm{mod}}(i) \right)^2 \to \min , \tag{5}$$

where *n* is the length of data vector for each state variable; $X_{\exp}$ and $S_{\exp}$ are known experimental data of biomass and substrate; X_{mod} and S_{mod} are model predictions with a given set of the parameters.

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 [18].

Cuckoo search algorithm

Cuckoos are fascinating birds, not only because of the beautiful sounds they make but also because of their aggressive reproduction strategy [21, 23, 25]. Some cuckoo species lay their eggs in communal nests, and may remove other birds' eggs to increase the hatching probability of their own. Quite a number of species engage in obligate brood parasitism by laying their eggs in the nests of other host birds (often other species) [21, 23, 25]. In this way cuckoos reduce the probability of their eggs being abandoned and thus increase their reproductivity.

For simplicity in describing the standard CS, the following three idealized rules are used [21, 23, 25]:
- Each cuckoo lays one egg at a time and dumps it in a randomly chosen nest;
- The best nests with high-quality eggs will be carried over to the next generations;
- The number of available host nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with a probability $p_a \in (0, 1)$. In this case, the host bird can either get rid of the egg or simply abandon the nest and build a completely new nest.

As a further approximation, this last assumption can be approximated by replacing a fraction p_a of the *n* host nests with new nests (with new random solutions).

The CS algorithm uses a balanced combination of a local random walk and the global explorative random walk, controlled by a switching parameter p_a. The local random walk can be written as:

$$x_i^{t+1} = x_i^t + \alpha s \otimes H(p_a - \varepsilon) \otimes (x_j^t - x_k^t), \tag{6}$$

where x_j^t and x_k^t are two different solutions selected randomly by random permutation, $H(u)$ is a Heaviside function, ε is a random number drawn from a uniform distribution, and s is the step size. Here, $\otimes$ means the entry-wise product of two vectors.

The global random walk is carried out using Lévy flights [21, 23, 25]:

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda), \tag{7}$$

where $\alpha > 0$ is the step size scaling factor and

$$L(s,\ \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi \lambda / 2)}{\pi} \frac{1}{s^{1+\lambda}},\ \left(s \gg s_0 > 0\right).$$

The initial solution is generated based on:

$$\boldsymbol{x} = Lb + (Ub - Lb)*rand(size(Lb)), \tag{8}$$

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

The described CS algorithm can be presented in pseudo code as shown in Fig. 1 [23].

```
Objective function F(x), x = (x₁, ..., x_d)ᵀ
Generate initial population of n host nests xᵢ
   while (t < MaxGeneration) or (stop criterion)
     Get a cuckoo randomly
     Generate a solution by Lévy flights [e.g., Eq. (7)]
     Evaluate its solution quality or objective value Fᵢ
     Choose a nest among n (say, j) randomly
     if (Fᵢ < Fⱼ),
        Replace j by the new solution i
     end
     A fraction (pₐ) of worse nests are abandoned
     New solutions (nests) are generated by Eq. (6)
     Keep best solutions
     Rank the solutions and find the current best
     Update t ← t + 1
   end while
Post-process results and visualization
```

Fig. 1 Pseudo code of Cuckoo search algorithm

Results and discussion
Numerical computations
For the parameter estimation of the considered model Eqs. (1)-(4), a set of identification procedures using CS, are performed in Matlab environment.

Computer specifications to run all optimization procedures are Intel® Core™i5-2320 CPU @ 3.00GHz, 8 GB Memory (RAM), Windows 7 (64bit) operating system.

The CS algorithm parameters, initially tuned on the basis of several pre-tests, taking into account the results in [1, 10, 11, 26], are as follows:

> MaxGeneration = 200;
> switching parameter $p_a = 0.25$;
> initial population of $n = 30$;
> Lévy exponent $\lambda = 1.5$.

The upper range *Ub* and lower range *Lb* are defined as follows:

> Lb = [0.30 0.0005 1.0],
> Ub = [0.6 0.05 3].

The results of model parameter identification − μ_{max}, $Y_{S/X}$ and k_S − with CS algorithm are obtained from 30 independent runs of the algorithm. The best, the worst and the mean results of the parameters estimates and the objective function value *J* are observed. The obtained results are summarized in Table 1.

Table 1. Results from model parameters identification

Results	Model parameters			J
	μ_{max}	$Y_{S/X}$	k_S	
best	0.4840	2.0194	0.0110	4.4440
worst	0.5003	2.0222	0.0138	4.6641
mean	0.4716	2.0205	0.0092	4.5662

The results from the application of CS are compared to the results from application of GA for parameter identification of the cultivation model Eqs. (1)-(4) obtained in [17]. According to [17] the main ACO and GA parameters and operators are as follows:

- Genetic algorithm

  fitness function – *linear ranking*;   generation gap = 0.97;
  selection function – *roulette wheel selection*;   crossover probability = 0.75;
  crossover function – *simple crossover*;   mutation probability = 0.01;
  mutation function – *binary mutation*;   number of generations = 200;
  reinsertion – *fitness-based*;   number of population = 30;

- Ant colony optimization algorithm

  evaporation parameter $\rho = 0.5$;   number of generations = 200;
  $a = b = 1$;   number of population = 30.

In Table 2, the results obtained from GA, ACO and CS for parameter identification of the cultivation model Eqs. (1)-(4) are summarized.

Table 2. Comparison of the parameter identification results from GA, ACO and CS

Model parameters	Estimates	Genetic algorithm	Ant colony optimization	Cuckoo search algorithm
μ_{max}	best	0.4909	0.4956	0.4840
	worst	0.5181	0.4641	0.5003
	mean	0.5039	0.5028	0.4716
$Y_{S/X}$	best	2.0226	2.0220	2.0194
	worst	2.0214	2.0240	2.0222
	mean	2.0223	2.0200	2.0205
k_S	best	0.0118	0.0146	0.0110
	worst	0.0182	0.0104	0.0138
	mean	0.0147	0.0151	0.0092
J	best	4.4816	4.7408	4.4440
	worst	5.0094	6.2202	4.6641
	mean	4.6519	5.2849	4.5662

In this paper, the results obtained from the three algorithms – GA, ACO and CS – under identical conditions, are compared. All algorithms have been run for 200 iterations with population size of 30 (respectively, chromosomes, ants, nests). As we can see from Table 2, the herewith presented CS algorithm yields better results than the ones produced by GA and ACO. It is noteworthy that the GA and ACO algorithms might have achieved better results if a different set of algorithm parameters have been applied. For example, GA with population size of 110 chromosomes, run for 200 iterations, achieves the best objective function result: $J = 4.4332$ [17]. Such algorithms required fine tuning of parameters for a specific problem, whereas CS algorithm is a more generic and robust one. In the CS algorithm, there are only two parameters − population size n, and switching parameter p_a − that determine the algorithm's efficiency. Once n is fixed, p_a essentially controls the elitism and the balance of the randomization and local search [25]. So, the explanation of the fact that CS has outperformed GA and ACO could be that there are fewer parameters to be fine-tuned in CS compared to ACO, and especially to GA. Thus, in CS a good balance of intensive local search and an efficient exploration of the search space is easily achieved. Finding such a balance, mainly controlled by the algorithm parameters, leads to a more efficient algorithm [25].

Graphical comparisons usually clearly show the presence or absence of systematic deviations between model predictions and measurements. Obviously, one of the important criteria for the adequacy of a model is the quantitative measure of the differences between calculated and measured values. Based on CS, ACO and GA best estimated set of model parameters, the model predictions of the cultivation process variables, namely biomass *X* and substrate *S*, are compared to the experimental data points of the *E. coli MC4110* cultivation. The graphical results are presented in Fig. 3 and Fig. 4, respectively.
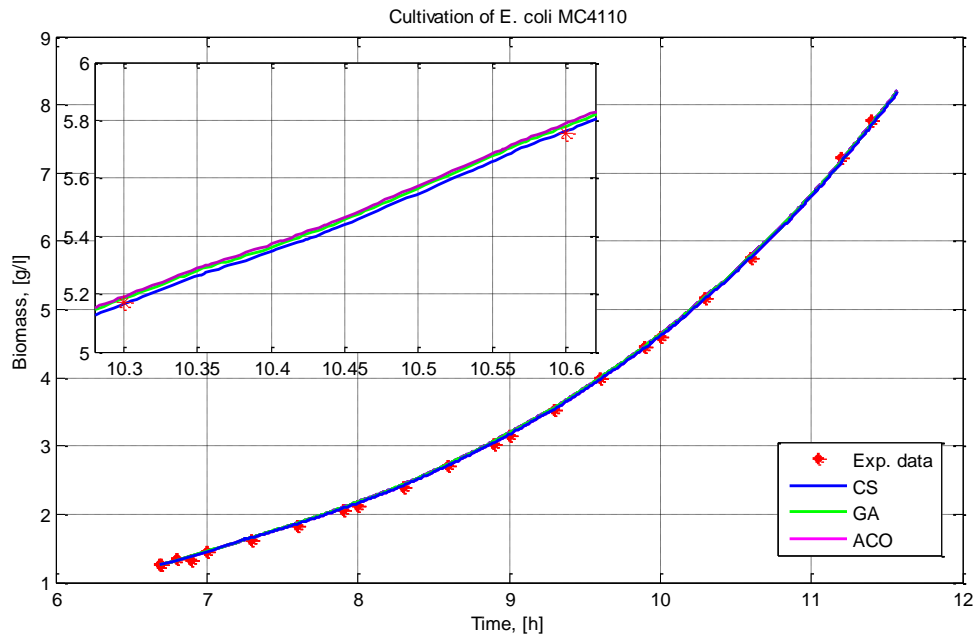
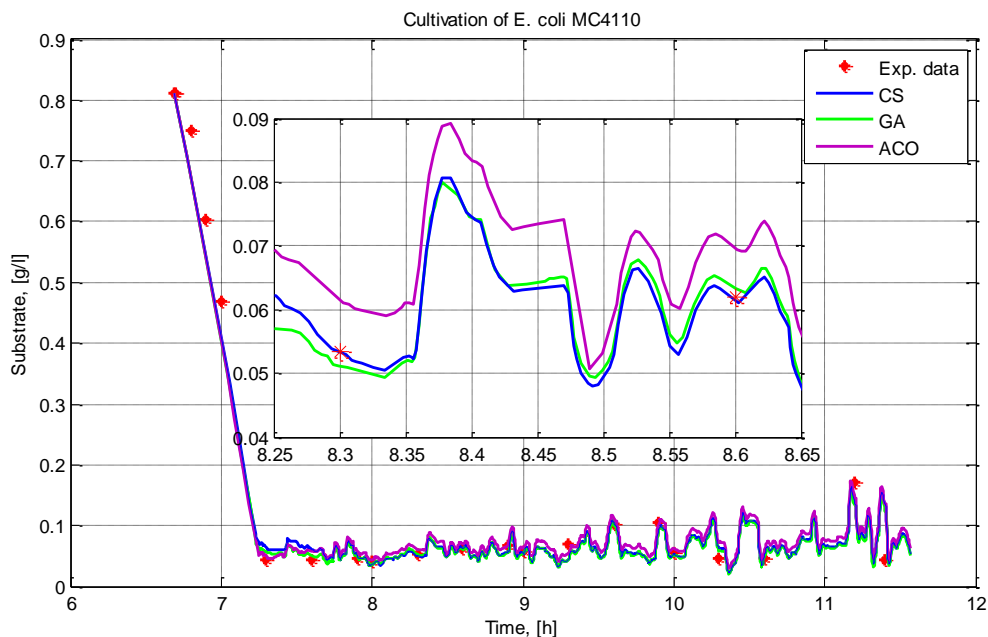Fig. 3 Experimental and model predicted data for biomass dynamics

Fig. 4 Experimental and model predicted data for substrate dynamics

The graphical comparison shows that there is a good coincidence between model and experimental data for biomass and substrate concentrations for the three algorithms – CS, GA and ACO. However, the model obtained from CS predicts with a high degree of accuracy the dynamics of the process variables during the fed-batch cultivation of *E. coli MC4110*. As it can be seen from the detailed subwindow views of the time periods in Fig. 3 and Fig. 4, the obtained model applying CS perfectly predicts the dynamics of biomass and substrate. In the identification procedure, raw measurement data are used without any preprocessing. Despite the rather noisy experimental data for substrate concentration, the model successfully follows the substrate data trend. Thus, it can be concluded that the CS obtained model adequately predicts the dynamics of the biomass and glucose throughout the process.

Conclusion

In this paper, the Cuckoo Search metaheuristic algorithm has been adapted and applied to the parameter identification of a non-linear dynamical model of *E. coli* MC4110 fed-batch cultivation process. The mathematical model of a cultivation process is presented by a system of ordinary differential equations, describing the main process variables – biomass and substrate. Numerical and simulation results reveal that correct and consistent results can be obtained using the CS algorithm. As a result, an adequate high-quality mathematical model of *E. coli MC4110* cultivation process is obtained. Further the performance of CS algorithm is compared to the GA and ACO performance. Analysis shows that CS algorithm obtains better results than the ones yielded with GA and ACO, i.e. CS has outperformed both GA and ACO.

The CS algorithm is more generic and robust for many optimization problems, compared to other metaheuristic algorithms. This fact confirms that the CS algorithm could be used as a 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 hybridize the CS algorithm with other metaheuristic based methods in order to achieve further improvement of effectiveness in solving bioprocess model parameter optimization problem.

Acknowledgements

References

1. Abdel-Baset M., I. Hezam (2016). Cuckoo Search and Genetic Algorithm Hybrid Schemes for Optimization Problems, Applied Mathematics & Information Sciences, 10(3), 1185-1192.
2. Ali A. F., A.-E. Hassanien (2016). A Survey of Metaheuristics Methods for Bioinformatics Applications, In: Applications of Intelligent Optimization in Biology and Medicine, Hassanien A.-E. et al. (Eds.), Springer International Publishing Switzerland, 23-46.
3. Basu M. A. (2013). Chowdhury Cuckoo Search Algorithm for Economic Dispatch, Energy, 60(13), 99-108.
4. Benjamin K. K., A. N. Ammanuel, A. David, Y. K. Benjamin (2008). Genetic Algorithm Using for a Batch Fermentation Process Identification, Journal of Applied Sciences, 8(12), 2272-2278.
5. Dorigo M., T. Stutzle (2004). Ant Colony Optimization, MIT Press.
6. Jiang L., Q. Ouyang, Y. Tu (2010). Quantitative Modeling of *Escherichia coli* Chemotactic Motion in Environments Varying in Space and Time, PLoS Computational Biology, 6(4), e1000735, doi:10.1371/journal.pcbi.1000735.
7. Karaboga D. (2005). An Idea Based on Honey Bee Swarm for Numerical Optimization, Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department.
8. 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.

9. Kennedy J., R. Eberhart (1995). Particle Swarm Optimization, Proceedings of IEEE International Conference on Neural Networks, IV, 1942-1948, doi:10.1109/ICNN.1995. 488968.

10. Nguyen T. T., D. N. Vo (2015). Modified Cuckoo Search Algorithm for Short-term Hydrothermal Scheduling, Electrical Power and Energy Systems, 65, 271-281.

11. Ong P. (2014). Adaptive Cuckoo Search Algorithm for Unconstrained Optimization, The Scientific World Journal, Volume 2014, Article ID 943403, http://dx.doi.org/ 10.1155/2014/943403.

12. 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 Biology, 8(9), e1000483, doi:10.1371/journal.pbio.1000483.

13. Pencheva T., M. Angelova, Modified Multi-population Genetic Algorithms for Parameter Identification of Yeast Fed-batch Cultivation, Bulgarian Chemical Communications, 2016, 48(4), in press.

14. 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.

15. Roeva O. (2012). Optimization of *E. coli* Cultivation Model Parameters Using Firefly Algorithm, International Journal Bioautomation, 16(1), 23-32.

16. Roeva O., S. Fidanova (2013). Hybrid Bat Algorithm for Parameter Identification of an *E. coli* Cultivation Process Model, Biotechnology and Biotechnological Equipment, 27(6), 4323-4326.

17. Roeva O., S. Fidanova, M. Paprzycki (2015). Population Size Influence on the Genetic and Ant Algorithms Performance in Case of Cultivation Process Modeling, Recent Advances in Computational Optimization, Studies in Computational Intelligence, 580, 107-120.

18. Roeva O., T. Pencheva, B. Hitzmann, St. Tzonkov (2004). A Genetic Algorithms Based Approach for Identification of *Escherichia coli* Fed-batch Fermentation, International Journal Bioautomation, 1, 30-41.

19. Slavov Ts., O. Roeva (2011). Genetic Algorithm Tuning of PID Controller in Smith Predictor for Glucose Concentration Control, International Journal Bioautomation, 15(2), 101-114.

20. Talbi El.-G. (2009). Metaheuristics: From Design to Implementation, John Wiley & Sons, Inc., Hoboken, New Jersey.

21. Yang X. S. (2008). Nature-inspired Meta-heuristic Algorithms, Luniver Press, Beckington, UK.

22. Yang X. S. (2009). Firefly Algorithm for Multimodal Optimization, Lecture Notes in Computing Sciences, 5792, 169-178.

23. Yang X.-S. (2014). Nature-inspired Optimization Algorithms, London, Elsevier Inc.

24. Yang X.-S., J. P. Papa (Eds.) (2016). Bio-inspired Computation and Applications in Image Processing, London, Elsevier Ltd.

25. Yang X.-S., S. Deb (2009). Cuckoo Search via Lévy Flights, Proceedings of World Congress on Nature & Biologically Inspired Computing (NaBIC), USA: IEEE Publications, 210-214.

26. Yang X.-S., S. Deb (2013). Multiobjective Cuckoo Search for Design Optimization, Computers & Operations Research, 40, 1616-1624.

27. Yang X.-S. (2010). A New Metaheuristic Bat-inspired Algorithm, In: NISCO 2010, SCI, Gonzalez J. R. et al. (Eds.), 284, 65-74.

28. Yildiz A. R. (2013). Cuckoo Search Algorithm for the Selection of Optimal Machine Parameters in Milling Operations, International Journal of Advanced Manufacturing Technology, 64(1-4), 55-61.

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 100 publications, among those 10 books and book chapters, with more than 400 citations. Her current scientific interests are in the fields of modelling, optimization and control of biotechnological processes, metaheuristic algorithms, intuitionistic fuzzy sets and generalized nets.

Senior Assist. Prof. Vassia Atanassova, Ph.D.
E-mail: vassia.atanassova@gmail.com

Vassia Atanassova completed Ph.D. in Informatics in the Institute of Information and Communication Technologies in 2013. Since 2014, she has been working in the Bioinformatics and Mathematical Modelling Department of the Institute of Biophysics and Biomedical Engineering at the Bulgarian Academy of Sciences. Her research interests are in modeling with generalized nets, theory and applications of intuitionistic fuzzy sets, as well as collaborative software and wiki technologies.