Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 407–414

ISBN 978-83-60810-51-4

# ACO for Parameter Settings of *E. coli* Fed-batch Cultivation Model

Stefka Fidanova and Olympia Roeva
Bulgarian Academy of Science,
Acad. G. Bonchev Str., bl. 25A,
1113 Sofia, Bulgaria
E-mails: stefka@parallel.bas.bg, olympia@biomed.bas.bg

Maria Ganzha
System Research Institute,
Polish Academy of Sciences,
Newelska Str. 6, 01-447 Warsaw, Poland
E-mail: maria.ganzha@ibspan.waw.pl

***Abstract*—*E. coli* plays significant role in modern biological engineering and industrial microbiology. In this paper the Ant Colony Optimization algorithm is proposed for parameter identification of an *E. coli* fed-batch cultivation process model. A system of nonlinear ordinary differential equations is used to model the biomass growth and the substrate utilization. We use real experimental data set from an *E. coli* MC4110 fed-batch cultivation process for performing parameter optimization. The objective function was formulated as a distance between the model predicted and the experimental data. Two different distances were used and compared – the Least Square Regression and the Hausdorff Distance. The Hausdorff Distance was used for the first time to solve the considered parameter optimization problem. The results showed that better results concerning model accuracy are obtained using the objective function with a Hausdorff Distance between the modeled and the measured data. Although the Hausdorff Distance is more time consuming than the Least Square Distance, this metric is more realistic for the considered problem.**

## I. Introduction

A lot of proteins are produced by the modified genetically microorganisms. One of the most used host organisms in the process is the *Escherichia coli* [33]. Furthermore, the *E. coli* is still the most important host organism for the recombinant protein production. In many cases, cultivation of recombinant micro-organisms e.g. the *E. coli*, is the only economical way to produce pharmaceutical biochemicals such as: interleukins, insulin, interferons, enzymes and growth factors, etc. Simple bacteria, like the *E. coli*, are manipulated to produce these chemicals so that they are easily harvested in vast quantities for use in medicine. Scientists may know more about the *E. coli* than they do know about any other species on earth. Research on the *E. coli* accelerated after 1997, after publication of its entire genome.The scientists were able to survey all 4,288 of its genes, discovering how groups of them worked together to break down food, make new copies of the DNA and do other tasks. However, despite decades of research, there rest a lot more to know about the *E. coli*. In 2002, they formed the *International E-coli Alliance*, for organization of projects that many laboratories could work together. As knowledge of the *E. coli* grows, scientists are starting to build models of the microbe that capture some of its behavior.It is important to be able to simulate how fast the microbe will grow on various sources of food, and how its growth changes if individual genes are knocked out. These questions are best answered by application of mathematical modeling.

Modeling of biotechnological processes is a common tool in process technology. It is obvious that the model is always a simplification of the reality. This is especially true when trying to model natural systems containing living organisms. However, for many industrial relevant processes, detailed models are not available due to the insufficient understanding of the underlying phenomena. These models can be too complicated and/or impossible to be solved. Therefore the specialists try to separate the most important components, and to create simplified models, which are as close as possible to the real processes. The mathematical models are very useful and effective tools in describing those effects. They are of great importance for control, optimization, or for understanding of the process. Thus the numerical solution of the models is fundamental for the development of powerful, though economical, methods in the fields of bioprocess design, plant design, scale-up, optimization and bioprocess control [26], [21]. Some of the recent researches and developed models of the *E. coli* were presented in [6], [15], [16], [20], [22], [28].

A common approach to model cellular dynamics is by systems of nonlinear differential equations. Obviously, parameter identification of a nonlinear dynamic model is more difficult than the linear one, as no general analytic results exist. The difficulties that may arise are such as: convergence to local solutions if standard local methods are used, over-determined models, badly scaled model function, etc. The problem is NP-hard and it is unpractical to be solved with exact or traditional numerical method. Therefore, existing research results indicate that the most useful solution method is by application of some metaheuristics. During the last decade metaheuristic techniques have been applied in a variety of areas. Heuristics can obtain suboptimal solution in ordinary situations and optimal solution in particular cases. Since the considered problem has been known to be NP-complete, using heuristic techniques can solve this problem more efficiently. Three best known (and most studied) heuristic approaches are: the iterative improvement algorithms, the probabilistic optimization algorithms, and the constructive heuristics. In this context, the evolutionary algorithms like: (a) Genetic Algorithms (GA) [13], [14], [18], (b) Evolution Strategies, (c)

Ant Colony Optimization (ACO) [7], [8], [9], [12], (d) Particle Swarm Optimization [32], (e) Tabu Search (TS) [35], (f) Simulated Annealing (SA) [17], (g) estimation of distribution algorithms, (h) scatter search, (i) path relinking, (j) greedy randomized adaptive search procedure, (k) multi-start and iterated local search, (l) guided local search, and (m) variable neighborhood search are - among others - often listed as examples of classical metaheuristics [3], [30], [31]. Obviously, they all have individual historical backgrounds and follow different paradigms and philosophies [4]. In this work the ACO is chosen as the most common direct method used for the global optimization.

The ACO is a rapidly growing research area of population-based metaheuristics that can be used to find approximate solutions to difficult optimization problems. It is applicable for a broad range of optimization problems, can be used in dynamic applications (adapts to changes such as new distances, etc.) and in some complex biological problems [10], [11], [27]. Recall that the ACO can compete with other global optimization techniques like GAs and SA. Overall, the ACO algorithms have been inspired by the real-world ant behavior. In nature, ants usually wander randomly, and upon finding food return to their nest while laying down pheromone trails. If other ants find such a path, they are likely to not continue traveling at random, but to follow the trail instead, returning and reinforcing it (if they eventually find food). However, as time passes, the pheromone starts to evaporate. Therefore, the more time it takes for an ant to travel down the path and back again, the more time the pheromone has to evaporate and the path becomes less noticeable. A shorter path, in comparison, will be visited by more ants and thus the pheromone density remains high for a longer time. The ACO is usually implemented as a team of intelligent agents which simulate the ants behavior, walking around the graph representing the problem to solve using mechanisms of cooperation and adaptation.

In this paper the ACO is applied for parameter identification of a system of the *E. coli* fed-batch cultivation process, described in terms of a mathematical model. Specifically, a system of nonlinear ordinary differential equations is proposed to model the *E. coli* biomass growth and substrate (glucose) utilization. The parameter optimization is performed using real experimental data set from the *E. coli* MC4110 fed-batch cultivation process. The cultivation was performed in the *Institute of Technical Chemistry, of the University of Hannover, Germany* during the collaboration work with the *Institute of Biophysics and Biomedical Engineering, BAS, Bulgaria*, and was funded by a grant *DFG*. The experimental data set includes records for the substrate feeding rate, concentration of biomass and substrate (glucose), and the cultivation time. In the nonlinear mathematical model considered here, the parameters that should be estimated are the maximum specific growth rate ($\mu_{max}$), the saturation constant ($k_S$), and the yield coefficient ($Y_{S/X}$).

The parameter estimation is performed based upon the use of a modified Hausdorff Metric [25] and the most commonly used metric – the Least Square Regression. The Hausdorff Metrics are used in the geometric settings for measuring the distance between sets of points. They have been used extensively in areas such as computer vision, pattern recognition and computational chemistry [34], [29], [19], [5]. The modified Hausdorff Distance is proposed to evaluate the mismatch between the experimental and the model predicted data. The results from both metrics are compared and analyzed.

The rest of the paper is organized as follows. The optimal parameters setting problem is formulated in Section 2. The ACO algorithm for the considered problem is defined in Section 3. The numerical results and the discussion are presented in Section 4. Concluding remarks are introduced in Section 5.

## II. Problem Formulation

The costs of developing mathematical models for the bioprocess improvement are often too high and the benefits too low. The main reason for this is related to the intrinsic complexity and non-linearity of biological systems. In general, mathematical descriptions of growth kinetics assume extensive simplifications. These models are often not accurate enough to correctly describe the underlying mechanisms. Another critical issue is related to the nature of the bioprocess models. Quite often, the parameters involved are not identifiable. Additionally, from the practical point of view, such identification would require data from specific experiments, which are themselves difficult to design and to realize. However, the estimation of model parameters with high parameter accuracy is essential for successful model development.

The real parameter optimization of simulation models, has become a research field of great interest in recent years. Nevertheless, after all completed research (for example, see references quoted above), this task still represents a very difficult problem. This mathematical problem, the so-called inverse problem, is a big challenge for the traditional optimization methods. In this case only the direct optimization strategies can be applied, because they exclusively use information about values of the goal function. Additional information about the goal function, like its gradients, etc., which could be used to accelerate the optimization process, is not available. Since an evolution of a goal for one string is provided by one simulation run, completing of the optimization algorithm may require a lot of computation time. Therefore, various metaheuristics are used as an alternative to surmount the parameter estimation difficulties.

### A. Problem Model

The general state space dynamical model of the process of interest was described by Bastin and Dochain in [2]. It is accepted as representing the dynamics of an *n* components and $m$ reactions bioprocess:

$$\frac{dx}{dt} = K\varphi(x,t) - Dx + F - Q. \tag{1}$$

Here, *x* is a vector representing the state components; *K* is the yield coefficient matrix; $\varphi$ is the growth rates vector; the

vectors $F$ and $Q$ are the feed rates and the gaseous outflow rates. The scalar $D$ is the dilution rate, which will be the manipulated variable, and which is defined as follows:

$$D = \frac{F_{in}}{V} \tag{2}$$

where $F_{in}$ is the influent flow rate and $V$ is the bioreactor volume.

Application of the general state space dynamical model [2] to the *E. coli* cultivation fed-batch process leads to the following nonlinear differential equation system [23]:

$$\frac{dX}{dt} = \mu_{max}\frac{S}{k_S + S}X - \frac{F_{in}}{V}X \tag{3}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\mu_{max}\frac{S}{k_S + S}X + \frac{F_{in}}{V}(S_{in} - S) \tag{4}$$

$$\frac{dV}{dt} = F_{in} \tag{5}$$

where:

| | | |
|---|---|---|
| $X$ | – | biomass concentration, [g/l]; |
| $S$ | – | substrate concentration, [g/l]; |
| $F_{in}$ | – | feeding rate, [l/h]; |
| $V$ | – | bioreactor volume, [l]; |
| $S_{in}$ | – | substrate concentration in the feeding solution, [g/l]; |
| $\mu_{max}$ | – | maximum value of the specific growth rate, $[h^{-1}]$; |
| $k_S$ | – | saturation constant, [g/l]; |
| $Y_{S/X}$ | – | yield coefficient, [-]. |

The mathematical formulation of the nonlinear dynamic model (Eqs. (3) - (5)) of the *E. coli* fed-batch cultivation process is described according to the mass balance and the model is based on the following a'priori assumptions:

- the bioreactor is completely mixed;
- the main products are biomass, water and, under some conditions, acetate;
- the substrate glucose is consumed mainly oxidatively and its consumption can be described by the Monod kinetics;
- the variation in the growth rate and the substrate consumption do not significantly change the elemental composition of the biomass, thus only balanced growth conditions are assumed;
- parameters, e.g. temperature, pH, or $pO_2$, are controlled at their individual constant set points.

For the parameter estimation problem the real experimental data of the *E. coli MC4110* fed-batch cultivation process is used. Off-line measurements of the biomass and on-line measurements of the glucose concentration are used in the identification procedure. The cultivation condition and the experimental data have been published in [24]. Here only the fermentation conditions described.

The fed-batch cultivation of the *E. coli* MC4110 is performed in a 2l bioreactor (Bioengineering, Switzerland), using a mineral medium [1], in the *Institute of Technical Chemistry, University of Hannover*. Before inoculation, a glucose concentration of 2.5 g/l is established in the medium. Glucose in the feeding solution is 100 g/l. The initial liquid volume is 1350 ml. The pH is controlled at 6.8 and the temperature is kept constant at 35°C. The aeration rate is kept at 275 l/h air, the stirrer speed at start 900 rpm, and after 11 hours the stirrer speed is increased in steps of 100 rpm. At end the stirrer sped reaches 1500 rpm. Oxygen is controlled at around 35%.

*Off-line analysis*

For the off-line glucose measurements, as well as the biomass and the acetate concentration determination, samples of about 10 ml are taken approximately at every hour. Off-line measurements are performed by using the Yellow Springs Analyser (Yellow Springs Instruments, USA).

*On-line analysis*

For the on-line glucose determination a flow injection analysis (FIA) system has been employed, using two pumps (ACCU FM40, SciLog, USA) for the continuous sample and the carrier flow rate. To reduce the measurement noise the continuous-discrete extended Kalman filter was used [1].

*Glucose measurement and control system*

For on-line glucose determination, the same FIA system has been employed for the continuous sample and the carrier flow rate at 0.5 ml/min and 1.7 ml/min respectively. A total of 24 ml of cells containing the culture broth were injected into the carrier stream and mixed with an enzyme solution of 350 000 U/l of glucose oxidase (Fluka, Germany) of a volume of 36 ml. After passing a reaction coil of 50 cm length, the oxygen uptake was measured using an oxygen electrode (ANASYSCON, Germany). To determine the oxygen consumed by cells only, no enzyme solution were injected. Calculating the difference of both dissolved oxygen peak heights, the glucose concentration can be determined. The time between sample taking and the measurement of the dissolved oxygen was $\Delta t$ = 45 s.

For the automation of the FIA system, as well as glucose concentration determination, the software CAFCA (ANASYSCON, Germany) was applied. To reduce the measurement noise the continuous-discrete extended Kalman filter was used. This program was running on a separate PC and got the measurement results via a serial connection. A PI controller was applied to adjust the glucose concentration to the desired set point of 0.1 g/l [1].

The initial process conditions were [1]:
$t_0$ = 6.68 h, $X(t_0)$ = 1.25 g/l, $S(t_0)$ = 0.8 g/l, $S_{in}$ = 100 g/l.

### B. Optimization Criterion

From the practical perspective, modelling studies are performed to identify simple and easy-to-use models that are suitable to support the engineering tasks of process optimization

and, especially, of control. The most appropriate model must satisfy the following conditions:

(i) the model structure should be able to represent the measured data in a proper manner;
(ii) the model structure should be as simple as possible, while remaining compatible with the first requirement.

On account of that, the cultivation process dynamic is described using a simple Monod-type model, the most common kinetics applied for modelling of the cultivation processes [2].

The optimization criterion is a certain factor, value of which defines the quality of an estimated set of parameters. To evaluate the mishmash between the experimental and the model predicted data, a modified Hausdorff Distance and the Least Square Regression are proposed.

In this work the Hausdorff Metric is used for the first time to solve the parameter optimization problem involving cultivation processes models.

*1) Hausdorff Distance:* When talking about distances, it usually means the shortest: for instance, if a point $X$ is said to be at distance $D$ of a polygon $P$, it is generally assumed that $D$ is the distance from $X$ to the nearest point of $P$. The same logic applies for polygons: if two polygons $A$ and $B$ are at some distance from each other, it id commonly understood that the distance is the shortest one between any point of $A$ and any point of $B$. That definition of distance between polygons can become quite unsatisfactory for some applications. However, it would be natural to expect that a small distance between two polygons means that no point of one polygon is far from the other polygon. Unfortunately, the shortest distance concept carries very low informative content.

In mathematics, the Hausdorff Distance, or the Hausdorff Metric (named after Felix Hausdorff), also called Pompeiu-Hausdorff Distance [25], measures how far two subsets of a metric space are from each other. It turns the set of non-empty compact subsets of a metric space into a metric space in its own right. Informally, two sets are close in the Hausdorff Distance if every point of either set is close to some point of the other set. In other words, the Hausdorff Distance is the longest distance you can be forced to travel by an adversary who chooses a point in one of the two sets, from where you then must travel to the other set. Thus, it is the farthest point of a set that you can be at, to the closest point of a different set. More formally, the Hausdorff Distance from set $A$ to set $B$ is a maxmin function defined as:

$$h(A,B) = \max_{a \in A} \left\{ \min_{b \in B} \{ d(a,b) \} \right\}, \tag{6}$$

where $a$ and $b$ are points of sets $A$ and $B$ respectively, and $d(a,b)$ is any metric between these points. For simplicity, in this work, the $d(a,b)$ as the Euclidean distance between $a$ and $b$ is taken. If sets $A$ and $B$ are made of lines or polygons instead of single points, then $h(A,B)$ applies to all defining points of these lines or polygons, and not only to their vertices. The Hausdorff Distance gives an interesting measure of mutual proximity, by indicating the maximal distance between any point of one set to the other set. IN this way it is better than the shortest distance, which applied only to one point of each set, irrespective of all other points of the sets.

*2) Least Squares Regression:* The objective of the modelling process consists of adjusting the parameters of a model function to best fit the data set. A simple data set consists of $n$ points (data pairs) $(x_i, y_i)$, $i = 1, 2, \ldots, n$, where $x_i$ is an independent variable and $y_i$ is a dependent variable value of which is found by observation. The model function has the form $f(x, \beta)$, where the $m$ adjustable parameters are held in the vector $\beta$. The goal is to find the parameter values for the model which "best" fits the data. The least squares method finds its optimum when the sum $S$ of squared residuals:

$$S = \sum_{i=1}^{n} r_i^2$$

is at a minimum. A residual is defined as the difference between the actual value of the dependent variable and the value predicted by the model. A data point may consist of more than one independent variable. For example, when fitting a plane to a set of height measurements, the plane is a function of two independent variables, $x$ and $z$. In the most general case there may be one or more independent variables and one or more dependent variables at each data point.

$$r_i = y_i - f(x_i, \beta).$$

## III. Ant Colony Optimization (ACO)

The ACO is a stochastic optimization method that mimics the social behavior of real ants colonies, which manage to establish the shortest rout to feeding sources and back. Real ants foraging for food lay down quantities of pheromone (chemical cues) marking the path that they follow. An isolated ant moves essentially at random but an ant encountering a previously laid pheromone will detect it and decide to follow it with high probability and thereby reinforce it with a further quantity of pheromone. The repetition of the above mechanism represents the auto-catalytic behavior of a real ant colony, where the more the ants follow a trail, the more attractive that trail becomes. The original idea comes from observing the exploitation of food resources among ants, in which ants' individually limited cognitive abilities have collectively been able to find the shortest path between a food source and the nest.

**Basic of Ant Algorithm**

The ACO is usually implemented as a team of intelligent agents, which simulate the ants behavior, walking around the graph representing the problem to solve, using mechanisms of cooperation and adaptation. The requirements of the ACO algorithm are as follows [3], [8]:

- The problem needs to be represented appropriately, which would allow the ants to incrementally update the solutions through the use of a probabilistic transition rules, based

on the amount of pheromone in the trail and other problem specific knowledge.
- A problem-dependent heuristic function, that measures the quality of components that can be added to the current partial solution.
- A rule set for pheromone updating, which specifies how to modify the pheromone value.
- A probabilistic transition rule based on the value of the heuristic function and the pheromone value, that is used to iteratively construct a solution.

The structure of the ACO algorithm is shown by the pseudocode below (Figure 1). The transition probability $p_{i,j}$, to choose the node $j$ when the current node is $i$, is based on the heuristic information $\eta_{i,j}$ and the pheromone trail level $\tau_{i,j}$ of the move, where $i, j = 1, \ldots, n$.

$$p_{i,j} = \frac{\tau_{i,j}^a \eta_{i,j}^b}{\sum\limits_{k \in Unused} \tau_{i,k}^a \eta_{i,k}^b}, \tag{7}$$

where $Unused$ is the set of unused nodes of the graph.

The higher the value of the pheromone and the heuristic information, the more profitable it is to select this move and resume the search. In the beginning, the initial pheromone level is set to a small positive constant value $\tau_0$; later, the ants update this value after completing the construction stage. The ACO algorithms adopt different criteria to update the pheromone level.

**Ant Colony Optimization**
Initialize number of ants;
Initialize the ACO parameters;
**while not** end-condition **do**
  **for** $k = 0$ **to** number of ants
    ant $k$ choses start node;
    **while** solution is not constructed **do**
      ant $k$ selects higher probability node;
    **end while**
  **end for**
  Update-pheromone-trails;
**end while**

Fig. 1. Pseudocode for ACO

The pheromone trail update rule is given by:

$$\tau_{i,j} \leftarrow \rho \tau_{i,j} + \Delta \tau_{i,j}, \tag{8}$$

where $\rho$ models evaporation in the nature and $\Delta \tau_{i,j}$ is new added pheromone which is proportional to the quality of the solution. Thus better solutions will receive more pheromone than others and will be more desirable in a next iteration.

## IV. Numerical Results and Discussion

In this section, a more precise description, concerning the application of the ACO for the parameter optimization of the *E. coli* cultivation process model, is presented. Here, the parameters $\mu_{max}$, $k_S$ and $Y_{S/X}$ have to be estimated. First, the problem is represented by a graph. It is needed to find th optimal values of three parameters which are interrelated. Therefore, the problem is represented with three-partitive graph. The graph consists of three levels. Every level represents a search area of one of the parameters that will be optimized. Every area is thus discretized, to consists of 1000 points (nodes), which are uniformly distributed in the search interval of every parameter. The first level of the graph represents the parameter $\mu_{max}$. The second level represents the parameter $k_S$. The third level represents the parameter $Y_{S/X}$. There are arcs between nodes from consecutive levels of the graph and there are no arcs between nodes from the same level. The pheromone is deposited on the arcs, to indicate how good is this parameter combination. Every level of the graph of the problem consists of 1000 points, thus the number of possible solutions is $10^9$, therefore is unpractical to apply the exact methods.

### A. ACO for Parameter Optimization

Here the proposed ACO approach is very close to real ant behavior. Starting to create a solution, the ants chose a node from the firs level in a random way. Next, for nodes from the second and the third level, they apply the probabilistic rule. The transition probability depends only on the pheromone level. The heuristic information is not used. Thus the transition probability is as follows:

$$p_{i,j} = \frac{\tau_{i,j}}{\sum\limits_{k \in Unused} \tau_{i,k}}, \tag{9}$$

The ants prefer the node with maximal probability, which is the node with maximal quantity of the pheromone on the arc (starting from the current node). If there is more than one candidate for next node, the ant chooses randomly between the candidates. The process is iterative. At the end of every iteration the pheromone on the arcs is updated. The quality of the solutions is represented by the value of the objective function. In this case the objective function is the mean distance between the simulated data and the experimental data, which are the concentration of the biomass and the concentration of the substrate. The aim of the process is to minimize it, therefore the new added pheromone by ant $i$ is:

$$\Delta \tau = (1 - \rho)/J(i) \tag{10}$$

where $J(i)$ is the value of the objective function according the solution constructed by ant $i$. Thus the arcs corresponding to solutions with the lesser value of the objective function will receive more pheromone and will be more desirable in the next iteration.

The values of the parameters of the ACO algorithms are very important, because they manage the search process. Therefore, it is necessary to find appropriate parameter settings, where the number of ants is the main parameter. In the ACO a small number of ants between 10 and 20 can

be used, without need to increase the number of iterations to achieve good solutions. The next parameter is the initial pheromone. Normally it has a small value. The last parameter is the evaporation rate, which shows the importance of the last found solution, as related to the previous ones. Parameters of the ACO were tuned based on several pre-tests according considered here optimization problem. After tuning procedures the main algorithm parameters are set to the optimal settings. The parameter settings for the ACO are shown in Table I.

TABLE I
PARAMETERS OF ACO ALGORITHM

| Parameter | Value |
|---|---|
| number of ants | 20 |
| initial pheromone | 0.5 |
| evaporation | 0.1 |

In this paper two different measures – the Least Square Regression and the modified Hausdorff Distance are compared. The modified Hausdorff Distance, which is conformable to the considered problem is applied. There are two sets of points, the simulated (model predicted) and the measured (experimental) data, which form two lines. The Euclidean distance $d(t)$ between points from the two lines, corresponding to the same time moment $t$, is calculated. After that, the Euclidean distance from the point of one of the lines in time $t$ to the points from other line in the time interval $(t - d(t), t + d(t))$ is calculated, and the minimal of these distances is taken. This is the distance between the two lines in the time moment $t$. Thus, the number of calculations, compared with the traditional Hausdorff Distance, is decreased, due to the fact that the distance to the points out of the interval $(t-d(t), t+d(t))$ will be large. At the end all distances between the points and the lines are combined. Thereby, eventual larger distance in some time moment, due to the measurement noise, is eliminated.

Thus, the objective function is presented as a minimization of the modified Hausdorff Distance measure $J_1$ between experimental and model predicted values of the state variables, represented by the vector $\mathbf{y}$:

$$J_1 = \sum_{i=1}^{m} h\left(\mathbf{y}_{\exp}(i), \mathbf{y}_{\mod}(i)\right)^2 \rightarrow \min \qquad (11)$$

where $m$ is the number of state variables (biomass and glucose concentrations); $\mathbf{y}_{\exp}$ is the known experimental data; while $\mathbf{y}_{\mod}$ model predictions with a given set of the parameters.

In the case of the Least Square Regression the objective function is:

$$J_2 = \sum_{i=1}^{m} \left(\mathbf{y}_{\exp}(i) - \mathbf{y}_{\mod}(i)\right)^2 \rightarrow \min \qquad (12)$$

### B. Numerical Results

All experiments have been conduced on a PC with Intel Core 2 2.8 GHz, 3.5 GB Memory, Linux operating system and using the Matlab 7.5 environment. Because of the stochastic characteristics of the applied ACO algorithm, a series of 30 runs for each algorithm was performed.

TABLE II
ACO WITH LEAST SQUARE REGRESSION AND HAUSDORFF DISTANCE

| | Method | Average | Worst | Best |
|---|---|---|---|---|
| ACO Least Square | – | 4.8866 | 6.7700 | 3.3280 |
| | Hausdorff | 2.3875 | 4.1290 | 1.7218 |
| ACO Hausdorff | – | 1.8744 | 2.5322 | 1.6425 |
| | Least Square | 3.9706 | 4.4283 | 3.4276 |

TABLE III
BEST PARAMETER VALUES OF THE MODEL

| Parameter | Value |
|---|---|
| $\mu_{max}$ | 0.5283 |
| $k_S$ | 0.0174 |
| $Y_{S/X}$ | 2.0300 |

To study the algorithm performance, the worst the best and the average results of the 30 runs, for the objective function values of the two variants are studied. For a realistic comparison, the number of iterations is fixed to be 100. The average, worst and best values of the objective functions are shown on Table II. In the first line of the second row, the average, worst and best value of the objective function are shown when Least Square Regression is used. The second line in the second row depicts the calculated Hausdorff Distance between the same solutions achieved when the objective function is the Least Square Regression. The first line of the third row shows the average, best and worst values of the objective function when it is the Hausdorff Distance. The second line of the third row represents the Least Square Distance of the same solutions achieved when the objective function is the Hausdorff Distance. Comparing the two rows it can be observed that the average and the worst achieved results are much better using the Hausdorff Distance than the Least Square Regression. The best achieved solutions are similar. In the best achieved solutions it can be seen that the Hausdorff Distance between achieved solutions is smaller when the objective function is Hausdorff, but the Least Square Distance is smaller when the objective function is the Least Squares Regression. During a number of runs of the algorithm the same phenomenon was observed – a small Hausdorff Distance between modeled and measured data and at the same time a big Least Square Distance between the data. When the Least Squares Regression is applied as the metric, the distance between the two lines can be very big, and in the same time it is seen that they are geometrically close to each other. It can happen especially in the steep parts of the lines. Applying the Hausdorff Metrics it can not happen, because it measures the geometrical similarity. Overall, the Hausdorff Distance is more time consuming than Least Square Distance, but much more realistic for the type of problems considered here. It can be concluded that algorithm proposed in this paper performs better when the objective function is the Hausdorff Distance.

In Table III the best parameter values ($\mu_{max}$, $k_S$ and $Y_{S/X}$), obtained using the ACO with the objective function based on

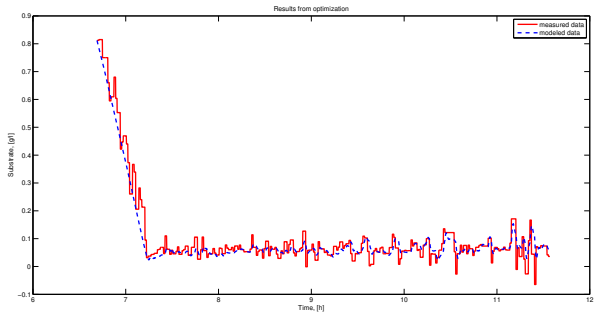


Fig. 2. Time profiles of the substrate: experimental data and models predicted data

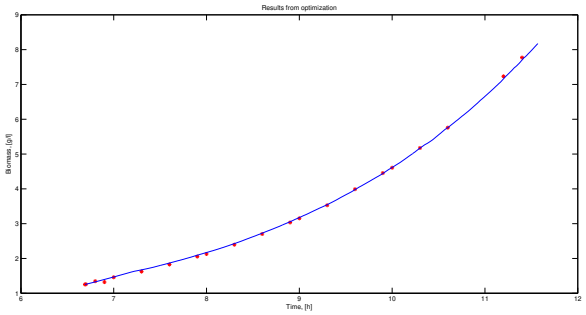


Fig. 3. Time profiles of the biomass: experimental data and models predicted data

the Hausdorff Distance, are presented.

The obtained model dynamics compared to the real experimental data is presented in Fig. 2 and Fig. 3.

In Figure 2, the modelled substrate is represented by the dash line, while by solid line the measured substrate is depicted. In Figure 3, line represents values of the modelled biomass, while stars represent values of the measured biomass.

The presented figures show a very good correlation between the experimental and model predicted data and confirm the obtained results.

## V. Concluding remarks

In this work the ACO algorithm for a parameter setting of the *E. coli* fed-batch cultivation process model was proposed. The ACO is chosen as the most common direct method used for global optimization. The process model is presented as a system of nonlinear ordinary differential equation describing the biomass and the substrate dynamics. In the identification procedures, the real experimental data was used. The objective function was formulated as the difference between the modeled and the measured data. When solving the optimization problem, two different measures were used – the commonly used Least Square Regression and, for the first time applied to this type of problem, the Hausdorff Distance. To adapt the Hausdorff Distance to the considered problem a modification of this metric was proposed. Comparison of the results shows that the Hausdorff Distance is more time consuming than the Least Square Distance. However, at the same time, the highest parameter accuracy is achieved when the objective function is measured as the Hausdorff Distance between the model predicted and the real experimental data.

## Acknowledgment

This work has been partially supported by the Bulgarian National Scientific Fund under the 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". Work presented here is a part of the Poland-Bulgarian collaborative Grant "Parallel and distributed computing practices".

## References

[1] M. Arndt and B. Hitzmann, "Feed Forward/feedback Control of Glucose Concentration during Cultivation of *Escherichia coli*", *8th IFAC Int. Conf. on Comp. Appl. in Biotechn*, Canada, 2001, pp. 425–429.

[2] G. Bastin and D. Dochain, *On-line Estimation and Adaptive Control of Bioreactors*. Els. Sc. Publ, 1991.

[3] E. Bonabeau, M. Dorigo and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, New York,Oxford University Press, 1999.

[4] J. Brownlee, *Clever Algorithms. Nature-Inspired Programming Recipes*. LuLu, p. 436, 978-1-4467-8506-5, 2011.

[5] Sh. Chen and B.C. Lovell, "Feature space Hausdorff distance for face recognition", In: *Proc. of 20th International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010, pp. 1465–1468.

[6] M.W. Covert, N. Xiao, T.J. Chen and J.R. Karr, "Integrating Metabolic, Transcriptional Regulatory, and Signal Transduction Models in *Escherichia coli*", *Bioinformatics*, Vol. 24, No. 18, 2008, pp. 2044–2050.

[7] M. Dorigo and G. Di Caro, "The Ant Colony Optimization Metaheuristic" *In: Corne, D, Dorigo, M., Glover, F. (eds).: New Idea in Optimization*, McGrow-Hill, 1999, pp. 11–32.

[8] M. Dorigo and T. Stutzle, *Ant Colony Optimization*, MIT Press, 2004.

[9] S. Fidanova, "ACO Algorithm with Additional Reinforcement", *Ant Colonies to Artificial Ants*, Lecture Notes in Computer Science No 2463, Springer, Germany, 2002, pp. 292–293

[10] S. Fidanova and I. Lirkov, "3D Protein Structure Prediction", *J. Analele Universitatii de Vest Timisoara, Seria Matematica-Informatica*, Vol XLVII(2),ISSN 1224-970X,2009, pp. 33–46.

[11] S. Fidanova, "An Improvement of the Grid-based Hydrophobic-hydrophilic Model", *Int. J. Bioautomation*, ISSN 1312-451X, Vol. 14, No. 2, 2010, pp. 147–156.

[12] S. Fidanova, E. Alba and G. Molina, "Hybrid ACO Algorithm for the GPS Surveying Problem", *Large Scale Scientfic Computing*, Springer, Berlin, Vol. 5910, 2010, pp. 318–325.

[13] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley Longman, London, 2006.

[14] J.H. Holland, *Adaptation in Natural and Artificial Systems*. 2nd Edn. Cambridge, MIT Press, 1992.

[15] L. Jiang, Q. Ouyang and Y. Tu, . "Quantitative Modeling of *Escherichia coli* Chemotactic Motion in Environments Varying in Space and Time", *PLoS Comput Biol*, Vol. 6, No. 4, e1000735. doi:10.1371/journal.pcbi.1000735, 2010.

[16] T.A. Karelina, H. Ma, I. Goryanin and O.V. Demin, "EI of the Phosphotransferase System of *Escherichia coli*: Mathematical Modeling Approach to Analysis of Its Kinetic Properties", *Journal of Biophysics*, Vol. 2011, Article ID 579402, doi:10.1155/2011/579402, 2011.

[17] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, "Optimization by Simulated Annealing", *Science, New Series*, Vol. 220, No. 4598, 1983, pp. 671–680.

[18] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*. Second, Exended Edition, Springer-Verlag, Berlin, Heidelberg, 1994.

[19] S. Nutanong, E.H. Jacox and H. Samet, "An Incremental Hausdorff Distance Calculation Algorithm", In: *Proc. of the VLDB Endowment*, Vol. 4, No. 8, 2011, pp. 506–517.

[20] N. Opalka, J. Brown, W.J. Lane, K.-A.F. Twist, R. Landick, F.J. Asturias, and S.A. Darst, "Complete Structural Model of *Escherichia coli* RNA Polymerase from a Hybrid Approach", *PLoS Biol*, Vol. 8, No. 9, e1000483. doi:10.1371/journal.pbio.1000483, 2010.

[21] P.M. Pardalos and M.G.C. Resende, "Handbook of Applied Optimization", Oxford University Press 2002.

[22] C.M. Petersen, H.S. Rifai, G.C.Villarreal and R. Stein, "Modeling *Escherichia coli* and Its Sources in an Urban Bayou with Hydrologic Simulation Program – FORTRAN", *Journal of Environmental Engineering*. Vol. 137, No. 6, 2011, pp. 487–503.

[23] O. Roeva, (2008). "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, pp. 601–608.

[24] O. Roeva, T. Pencheva, B. Hitzmann and St. Tzonkov, "A Genetic Algorithms Based Approach for Identification of *Escherichia coli* Fed-batch Fermentation", *Int. J. Bioautomation*, Vol. 1, 2004, pp. 30–41.

[25] G. Rote, "Computing the minimum Hausdorff distance between two point sets on a line under translation", *Information Processing Letters*, Vol. 38, 1991, pp. 123–127.

[26] K. Schuegerl and K.-H. Bellgardt, *Bioreaction Engineering: Modeling and Control*. Berlin Heidelberg, Springer-Verlag, 2000.

[27] A. Shmygelska and H.H. Hoos, "An ant colony optimization algorithm for the 2D and 3D hydrophobic polar protein folding problem", *BMC Bioinformatics*, Vol. 6, No. 30, doi:10.1186/1471-2105-6-30, 2005.

[28] P.N. Skandamis and G.E. Nychas, "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*, Vol. 66, No. 4, 2000, pp. 1646–1653.

[29] M. Sugiyama, E. Hirowatari, H. Tsuiki and A. Yamamoto, "Learning figures with the hausdorff metric by fractals", In: *Proc. of the 21st international conference on Algorithmic learning theory*, Springer-Verlag Berlin, Heidelberg, 2010, pp. 315–329.

[30] W.P. Syam and I.M. Al-Harkan, "Comparison of Three Meta Heuristics to Optimize Hybrid Flow Shop Scheduling Problem with Parallel Machines", *World Academy of Science, Engineering and Technology*, Vol. 62, 2010, pp. 271–278.

[31] N. Tahouni, R. Smith and M.H. Panjeshahi, "Comparison of Stochastic Methods with Respect to Performance and Reliability of Low-temperature Gas Separation Processes", *The Canadian Journal of Chemical Engineering*, Vol. 88, No. 2, 2010, pp. 256–267.

[32] R. Umarani and V. Selvi, "Particle Swarm Optimization: Evolution, Overview and Applications", *Int J of Engineering Science and Technology*, Vol. 2, No. 7, 2010, pp. 2802–2806.

[33] U. Viesturs, D. Karklina and I. Ciprovica, *Bioprocess and Bioengineering*, Jeglava, 2004

[34] H. Yedjour, B. Meftah, D. Yedjour and A. Benyettou, "Combining Spiking Neural Network with Hausdorff Distance Matching for Object Tracking", *Asian Journal of Applied Sciences*, Vol. 4, 2011, pp. 63–71.

[35] M.K. Yusof and M.A. Stapa, "Achieving of Tabu Search Algorithm for Scheduling Technique in Grid Computing using GridSim Simulation Tool: Multiple Jobs on Limited Resource", *Int J of Grid and Distributed Computing*, Vol. 3, No. 4, 2010, pp. 19–31.