



Simulation of Fermentation Pathway Using Bees Algorithm

Yi Ying Leong¹, Chuii Khim Chong¹, Yee Wen Choon¹, Lian En Chai¹, Safaai Deris¹, Rosli M Illias², Sigeru Omatu³, Mohd Saberi Mohamad^{1,*}

¹Artificial Intelligence and Bioinformatics Research Group, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia.

²Department of Bioprocess Engineering, Faculty of Chemical Engineering, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia.

³Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka 535-8585, Japan.

*Corresponding author

Email address:

dark_leaf06@hotmail.com, ckchong2@live.utm.my, ywchoon2@live.utm.my, lechai2@live.utm.my, safaai@utm.my, r-rosli@utm.my, omatu@rsh.oit.ac.jp, saberi@utm.my.

KEYWORD

*Parameter Estimation
Bees Algorithm
Simulation
Saccharomyces cerevisiae
Glucose
Glucose-6-Phosphate
Fermentation Pathway*

ABSTRACT

*In this paper, we propose Bees Algorithm (BA) to enhance the performance in estimating the parameters for metabolic pathway data to simulate fermentation pathway for *Saccharomyces cerevisiae*. However, the parameter estimation of biological processes has always been a challenging task due to the complexity and nonlinear equations. Therefore, we present this algorithm as a new approach for parameter estimation for biological interactions to obtain more accurate parameter values. The result shows that BA outperforms other estimation algorithms as it produces the most accurate kinetic parameters, which contributes to the precision of simulated kinetic model.*

1 Introduction

Developing a model that functions as biological function simulator is one of the goals in the study of system biology. It is a promising way to test whether combining the properties of the individual metabolites in isolation would yield a proper description of the fermentation pathway, without having to conduct the real experiment process [CHOU and EBERHARD, 2007; Albert, Curto, and Cascante, 1994]. Therefore, it is important to optimize the fermentation kinetic model to increase the efficiency of

the simulation process. Parameter estimation is the important key to generate optimal values for related parameters. Unlike the linear differential equation that is straight forward, the difficulties of parameter estimation increase with the size and complexity of the fermentation pathway, which contributes to the increasing numbers of equations and variables in the model. The complicated equations in this model also lead to high complexity of the model [LIU and KEVIN, 2000]. Fermentation simulation model has done a great favour in simulating the real process and therefore it needs to be accurate and precise [WANG and SHEU, 2000]. Previously, parameter estimation has been done by using other estimation algorithms



such as simulated annealing (SA) and simplex. Gonzalez *et al.*(2007) used SA to estimate the parameters in for S-System models of biochemical network. The parameter estimation of intracellular uptake and delivery for plasmid DNA by using simplex method has also been introduced by Zhou *et al.*(2007). However, both estimation algorithms have a common problem that they tend to trap at local minimum. As a result, the accuracy of the estimated parameter values will decrease as the algorithms are not able to look for better solutions in the search space. To solve this problem, we propose Bees Algorithm as a new method to compute the parameter values of glucose and glucose-6-phosphate in fermentation pathway. These metabolites were selected to be observed because they are the main starting materials which contribute to the yield of ethanol as the end product. The advantages of BA are its simplicity, ease of use and effectiveness as it always finds the global optimum and ignore local optimum. The outcomes from this study will promote the best value of kinetic parameters and increase the precision of simulated fermentation model. In Section 2, the method will be explained and in Section 3, the results will be shown in graph with the measurement data and simulated data generated for comparison. Section 4 will be the discussion. The last section of this paper is the conclusion and future work.

2 Methods

In previous works, SA and simplex have been used to estimate the parameter values. In this paper, we propose BA as a more efficient method for parameter estimation. There are 87 parameters in the fermentation pathway model for *Saccharomyces cerevisiae* but we only focused on the parameters that are relevant to glucose inside the cell and glucose-6-phosphate. BA will be implemented to estimate these identifiable parameters and the steps will be discussed in the following sections.

2.1 Bee Algorithm

In 2005, BA was first developed and it appears to be a new population-based search algorithm that mimics the natural foraging behaviour of swarms honey bees to find the optimal solution. Pham, Eldukhri and Soroka (2007) stated that the basic version

of algorithm performs a kind of neighbourhood search combined with random search and can be used widely in different optimization problems.

BA has the ability to locate optimal parameter values efficiently. It uses a population of solutions for each iteration and the outcome of each iteration is also a population of solutions. Only the bee with the highest fitness will be selected to form the next bee population. At the final stage, the fittest site is determined, which points to the best optimal parameter values for the relevant parameters as the output of the program. Figure 1 shows the important steps of BA and basic algorithm, respectively.

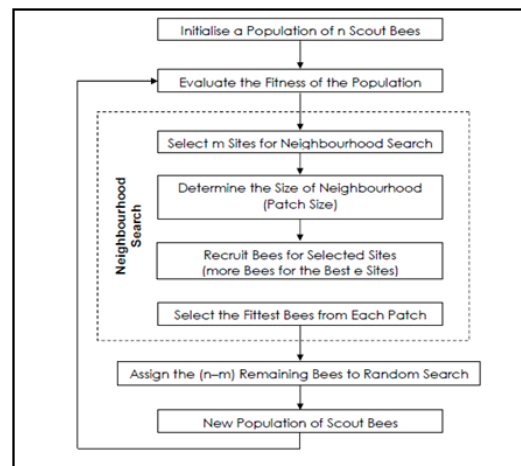


Fig. 1. Flowchart of BA.

2.2 Implementation of Bees Algorithm

First of all, the parameter values were initialized as follow: $n = 20$; $itr = 15$; $m = 10$; $e = 6$; $n1 = 4$; $n2 = 7$; $ng = 0.03$, where n is the number of scout bees; itr is the number of iterations; m is the number of best selected patches; e is the number of elite selected patches; $n1$ is the number of recruited bees around best selected patches ; $n2$ is the number of recruited bees around elite selected patches and ng is the patch radius for neighbourhood search. The initial scout bees were placed randomly in the search space and the fitnesses of the sites visited by scout bees were evaluated by using cost function. The formula of cost function is as follows:

$$\sum_{i=1}^N (y - y_i)^2 \quad (1)$$

The starting bees' population represents the guessed optimal point for parameter values. The iteration continues to form new populations of bees until it meets the stopping criterion. The neighbourhood search is chosen by looking at the sites visited by the bees that have the highest fitnesses. Best elite (e) sites and best neighbourhood (m) sites would be chosen and more bees are assigned to search near the best e sites which are made more detailed because they represent the most promising solutions. Next, there will be searching in the neighbourhood of the selected sites and the fitness of the population will be evaluated. For each patch, only the bees of highest fitness value are selected to form the next bee population. For the remaining bees in the population (n-m), they are placed randomly around the search space to search for new potential solution. Therefore, there are two parts of population formed in the new population, which are the selected bees for the best sites and the scout bees assigned to search for new random solutions. The ability of BA to search for new potential solution is the effective way to avoid local minimum because it always finds the global minimum as shown in the formation of the next new population [PHAM, ELDUKHRI and SORIKA, 2007; OZBAKIR, BAYKASOG and TAPKAN, 2010]. Figure 2 to Figure 5 show each step of the implementation of BA in parameter estimation for the fermentation pathway.

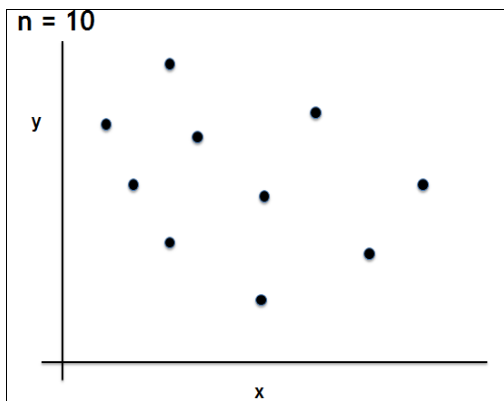


Fig. 2. Initialization of random population in the search space. In this figure, 10 bees are placed randomly as an example for illustration.

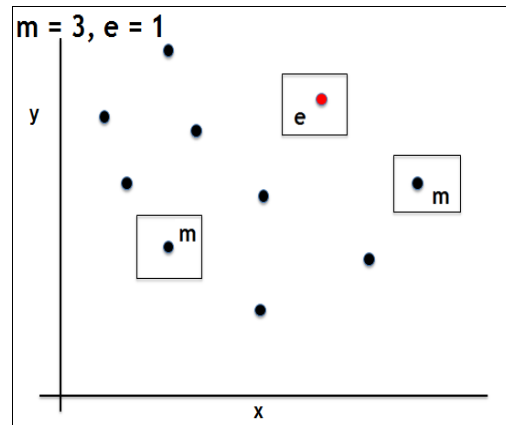


Fig. 3. Three m sites which have the best fitness values are chosen and e site is chosen as the best of the best site.

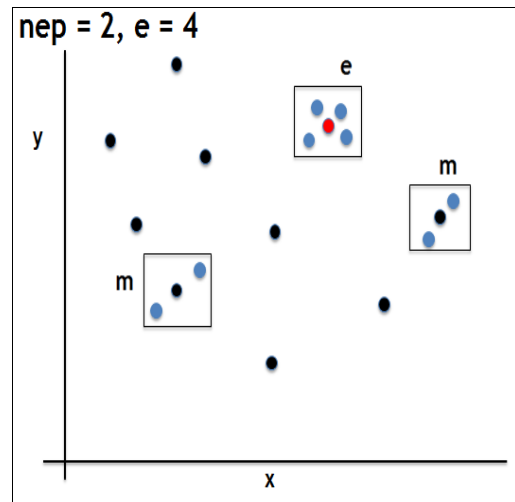


Fig. 4. Bees are recruited to the m sites. More bees are sent to e site while m sites have fewer bees based on their fitness value. At the end of the process, only one bee will be chosen from each patch.

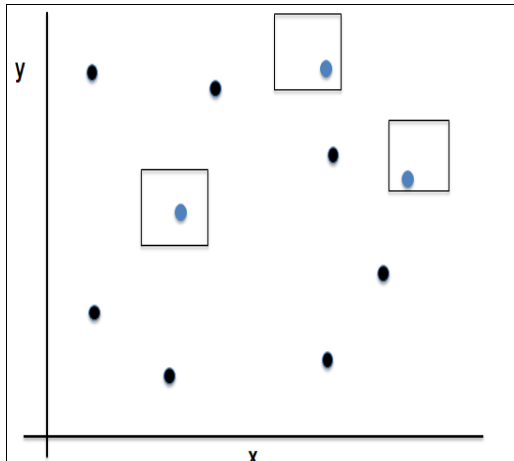


Fig. 5. At each of the selected patch, the bees that have the fittest function value are chosen. They are then used to form the next new population together with the new random bees.

2.3 Evaluation Method

After the determination of the suitable assessment technique for measuring the performance of BA in parameter estimation, the result is tested and the performance is evaluated by calculating the average of error rate (*A*), standard deviation (*STD*) value for 50 runs and constructing a production graph. The formula of error rate (*e*), *A* and *STD* values are as follow:

Error rate:

$$e = \sum_{i=1}^N (y - y_i)^2 \quad (2)$$

Average of error rate:

$$A = \frac{e}{N} \quad (3)$$

Standard deviation value:

$$STD = \sqrt{\frac{e}{N}} \quad (4)$$

The value of standard deviation value for each run will be recorded. An accurate simulated result would have a value close to zero.

3 Results

In this study, a comparison is made between three estimation algorithm, which are Simulated Annealing (SA), simplex and BA. The time series data for concentration of glucose inside the cell and glucose-6-phosphate are generated to evaluate the accuracy of each estimation algorithm. From the time series data, the average of error rate and standard deviation (*STD*) values will be calculated and the production graph with the comparison of three estimation algorithms will be produced. Table 1 shows the values of experimental parameter and simulated parameter of the three algorithms for the related kinetic parameters. The simulated results are used to calculate the average of error rate and *STD* values for metabolite glucose and glucose-6-phosphate.

Kinetic parameter	Experimental value	Simulated values		
		BA	Simplex	SA
Vin_Vm1	19.67	20.36	18.34	15.56
Vin_Ki1G6P	3.7	3.5	4.96	4.9
Vhk_Vm2	68.5	64.44	105.3	78.89
Vhk_Km2Glc	0.11	0.038	1.563	0.126
Vhk_Km2ATP	0.1	0.003	0.027	1.432
Vhk_Ks2Glc	0.006	0.0058	0.0043	2.81
Vpol_Vm3	14.31	27.44	10.65	4.44
Vpol_n3	8.25	10.34	4.28	16.42
Vpol_K3Gly	2	1.64	1.62	3.33
Vpol_Km30	1	1.5	10.63	6.82
Vpol_Km3G6P	1.1	0.353	1.24	5.82
Vpfk_Vm4	31.67	64.638	29.62	69.232
Vpfk_g4R	10	9.34	8.231	12.932
Vpfk_K4F6P	1	0.055	3.69	0.002
Vpfk_K4ATP	0.06	0.0045	0.1	1.52
Vpfk_L40	3342	2827	3432.34	1322.45
Vpfk_C4AMP	0.019	0.0002	0.0012	0.843
Vpfk_K4AMP	0.025	0.3232	7.384	4.11
Vpfk_C4F6P	0.0005	2.68	0.002	0.938
Vpfk_C4ATP	1	3.74	1.112	0.032
Vpfk_gT	1	1.043	7.398	0.2351

Table 1. Kinetic parameter values of BA compared with simplex and SA.

Error rate (*e*), Average of error rate (*A*), and *STD* value are then calculated and the results for glucose and glucose-6-phosphate are respectively shown in Table 2 and Table 3.



Evaluation Criteria	BA	Simplex	SA
Average of error rate, <i>A</i>	0.000368754	0.000648702	0.0012727
STD	0.000158632	0.00100982	0.00136283

Table. 2. Average of error rate and STD values for glucose. The shaded area shows the lowest value

Evaluation Criteria	BA	Simplex	SA
Average of error rate, <i>A</i>	3.67234E-06	4.76233E-06	2.12423E-05
STD	4.19832E-06	4.00213E-05	4.73538E-05

Table. 3. Average of error rate and STD values for glucose-6-phosphate. The shaded area shows the lowest value

Time series data for concentration of glucose and glucose-6-phosphate are then generated. The time series data consists of the experimental result and simulated results for BA, Simplex and SA.

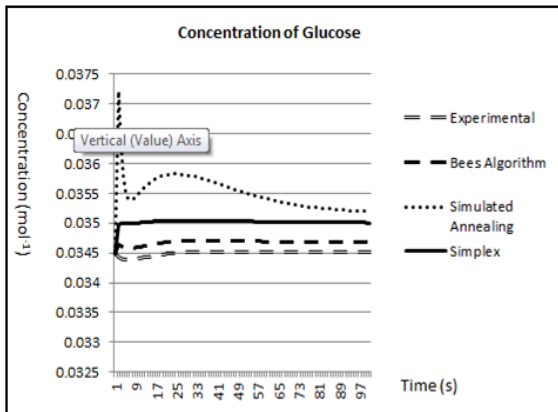


Fig. 6: Concentration of Glucose inside the Cell against time in second

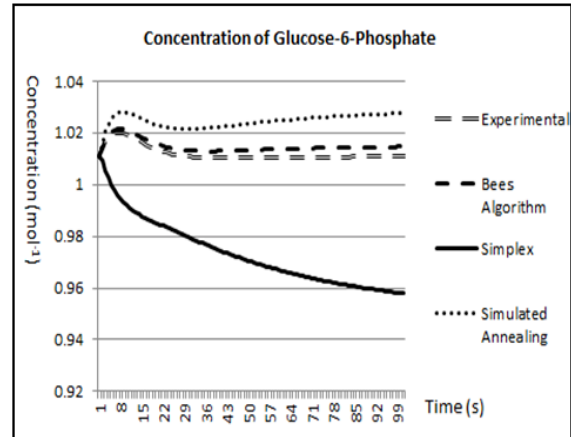


Fig. 7: Concentration of Glucose-6-Phosphate against time in second

4 Discussion

The experimental kinetic parameter values and simulated kinetic parameter values are substituted into the ordinary differential equations (ODEs) of glucose and glucose-6-phosphate as shown below.

$$d/dt(\text{Glc}) = (+\text{Vin} - \text{Vhk})/\text{cytoplasm} \quad (5)$$

$$d/dt(\text{G6P}) = (+\text{Vhk} - \text{Vpol} - \text{Vpfk})/\text{cytoplasm} \quad (6)$$

Vin represents the glucose in the cell which has two parameters properties, Vm1 and Ki1G6p. Vhk is hexokinase which is involved in the conversion of glucose-6-phosphate to glucose with the addition of ATP. Vpol's name is trehalose and glycogen formation, which involves in the conversion of carbo into glucose-6-phosphate and ATP. Vpfk represents phosphofructokinase and this is a reversible reaction of ATP and glucose-6-phosphate forming fructose 1,6-phosphate. The reactions are shown in Figure 8.



```

Vin = cytoplasm * (Vin_Vm1 - Vin_Ki1G6P * G6P)

Vhk = cytoplasm * Vhk_Vm2 / (1 + Vhk_Km2Glc / Glci + Vhk_Km2ATP / ATP + Vhk_Ks2Glc * Vhk_Km2ATP / (Glci * ATP))

Vpol = cytoplasm * 1.1 * Vpol_Vm3 * power(G6P, Vpol_n3) / (power(Vpol_K3Gly, Vpol_n3) + power(G6P, Vpol_n3)) / (1 + Vpol_Km30 / 0.7 * (1 + Vpol_Km3G6P / G6P))

Vpfk = cytoplasm * Vpfk_Vm4 * Vpfk_g4R * 0.3 * G6P / Vpfk_K4F6P * ATP / Vpfk_K4ATP * (1 + 0.3 * G6P / Vpfk_K4F6P + ATP / Vpfk_K4ATP + Vpfk_g4R * 0.3 * G6P / Vpfk_K4F6P * ATP / Vpfk_K4ATP) / (power(1 + 0.3 * G6P / Vpfk_K4F6P + ATP / Vpfk_K4ATP + Vpfk_g4R * 0.3 * G6P / Vpfk_K4F6P * ATP / Vpfk_K4ATP, 2) + Vpfk_L40 * power((1 + Vpfk_c4AMP * (3 - ATP - 0.5 * (-ATP + power(12 * ATP - 3 * power(ATP, 2), 0.5))) / Vpfk_K4AMP) / (1 + (3 - ATP - 0.5 * (-ATP + power(12 * ATP - 3 * power(ATP, 2), 0.5))) / Vpfk_K4AMP), 2) * power(1 + Vpfk_c4F6P * 0.3 * G6P / Vpfk_K4F6P + Vpfk_c4ATP * ATP / Vpfk_K4ATP + Vpfk_gT * Vpfk_c4F6P * 0.3 * G6P / Vpfk_K4F6P * Vpfk_c4ATP * ATP / Vpfk_K4ATP, 2))
    
```

Fig. 8: Reactions involving selected kinetic parameters

In BA, every formation of new population is made up of new possible random solution. The solutions are evaluated to obtain their fitness function values and the best solution will be chosen from the new population. The new best solution will be compared to the best solution from the previous population and the better ones will be kept while the worst ones would be eliminated. In this way, BA shows the ability to keep searching for new potential solutions instead of being trapped in local minimum. The global optimum search has made BA an efficient estimation algorithm.

From Table 2, BA is proven to be the algorithm which has the lowest average of error rate with 0.000368754 and lowest STD value with 0.000158632. On the other hand, SA shows the worst performance with 0.0012727 and 0.00136283 for the average of error rate and STD value, respectively. Simplex shows a more moderate performance with average error rate of 0.000648702 and STD value of 0.00100982. From Table 3, for glucose-6-phosphate, BA once again outperforms the other estimation algo-

rithms with the lowest average of error rate that is 3.67234E-06 and 4.76233E-06 for STD value. The average of error rate and STD value for Simplex and SA are 4.76233E-06, 4.00213E-05 and 2.12423E-05, 4.73538E-05 respectively.

In the time series data for the concentration of glucose and glucose-6-phosphate, y-axis represents the concentration and x-axis represents the time. From Figure 6, the line for BA is the closest to the experimental line. The rank is followed by Simplex. SA has a quite big distance line from the experimental line and this shows that it is quite unstable in simulating the result. The same case happens in Figure 7. The line for BA is the closest to the experimental result. SA also displays a line that has a big distance from the experimental line. BA has been proven to be the most stable and accurate as this algorithm can produce a simulated result that is closest to the experimental results.

5 Conclusions and Future Work

In this paper, BA is proven to have the best performance among three optimization algorithms. BA appears to be more trustworthy due to its ability to find the global optimum and avoid local optimum. It has increased the accuracy of simulated pathway model for *Saccharomyces Cerevisiae* and made an improvement in finding the most optimum parameter values to perform better simulation. Further improvements would be done by adding in more biological pathways to allow users to have more choices, perform other evaluation method and adding more functions for the algorithms to handle noisy data more efficiently.

6 Acknowledgment

We also would like to thank Universiti Teknologi Malaysia for supporting this research by the UTM GUP research grants (Vot number: QJ130000.7107.01H29 and QJ130000.7123.00H67).



7 References

- [ALBERT, CURTO and CASCANTE, 1994] Albert, S.R., Curto, R., Cascante, M., Comparative Characterization of the Fermentation Pathway of *Saccharomyces cerevisiae* Using Biochemical Systems Theory and Metabolic Control Analysis: Model Validation and Dynamic Behavior, *Math Biosci* 130(1) (1994) 25-50.
- [CHOU and EBERHARD, 2007] Chou, I.C., Eberhard, O.V., Recent developments in parameter estimation and structure identification of biochemical and genomic systems, *Mathematical Biosciences* 219 (2007) 57-83.
- [GONZÁLEZ *et al.*, 2007] Gonzalez, O.R., Küper C., Kirsten, J., Naval, P.C., Mendoza, E., Parameter estimation using Simulated Annealing for S-system models of biochemical networks, *Bioinformatics* 23(4) (2007) 480-486.
- [LIU and KEVIN, 2000] Liu, Y., Kevin, M.P., Computer simulation of biological systems, *Molecular and Cellular Biochemistry* 73 (2000) 91-98.
- [OZBAKIR, BAYKASOG and TAPKAN, 2010] Ozbakir, L., Baykasog, A. Tapkan, P., Bees algorithm for generalized assignment problem, *Applied Mathematics and Computation* 215(11) (2010) 3782-3795.
- [PHAM, ELDUKHRI and SORIKA, 2007] Pham, D. T., Eldukhri, E. E., Soroka, A. J., Intelligent production machines and systems, in *Intelligent Production Machines and Systems*, Elsevier 2007.
- [WANG and SHEU, 2000] Wang, F.S., Sheu, J.W., Multiobjective parameter estimation problems of fermentation processes using a high ethanol tolerance yeast', *Chemical Engineering Science*, 55(18) (2000) 3685-3695.
- [ZHOU *et al.*, 2007] Zhou, J.Y., Yockman, J.W., Sung, W.K., Steven E.K., Intracellular Kinetics of Non-Viral Gene Delivery Using Polyethylenimine Carriers, *Pharmaceutical Research* 24(6) (2007) 1079-1087.

