Parameter Estimation Using Improved Differential Evolution (IDE) and Bacterial Foraging Algorithm to Model Tyrosine Production in Mus Musculus (Mouse)

JIA XING YEOH[1], CHUII KHIM CHONG[1], YEE WEN CHOON[1], LIAN EN CHAI[1], SAFAAI DERIS[1], ROSLI M ILLIAS[2], and MOHD SABERI MOHAMAD[1, §]

[1]Artificial Intelligence and Bioinformatics Research Group, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia.
[2]Department of Biological Sciences, Faculty of Biosciences and Bioengineering, Universiti Teknologi Malaysia, Skudai 81310, Johor, Malaysia.

[§]Corresponding author

Email addresses:
jxyeoh3@live.utm.my, ckchong2@live.utm.my, ywchoon2@live.utm.my, lechai2@live.utm.my, safaai@utm.my, r-rosli@utm.my, saberi@utm.my

Keywords Parameter Estimation, Differential Evolution Algorithm, Bacterial Foraging Algorithm, Kalman Filtering Algorithm, Modeling.

Abstract

The hybrid of Differential Evolution algorithm with Kalman Filtering and Bacterial Foraging algorithm is a novel global optimization method that is implemented in this research to obtain the best kinetic parameter value. The proposed algorithm is then used to model tyrosine production in *mus musculus* (mouse) by using a dataset, JAK/STAT

(Janus Kinase Signal Transducer and Activator of Transcription) signal transduction pathway. Global optimization is a method to identify the optimal kinetic parameter using ordinary differential equation. From the ordinary parameter of biomathematical field, there are many unknown parameters and commonly the parameters are in nonlinear form. Global optimization method includes differential evolution algorithm which will be used in this research. Kalman Filter and Bacterial Foraging algorithm help in handling noise data and faster convergences respectively in the conventional Differential Evolution. The results from this experiment show estimatedly optimal kinetic parameters values, shorter computation time, and better accuracy of simulated results compared with other estimation algorithms.

Introduction

Metabolic pathway can be described by a combination of processes types including reversible reactions and in some with respects of multi-molecule reactions. Recently, many research have been done in the field of modelling in biology system where most of the pathways can be represented in the ordinary differential equation. Mathematical modelling of biological metabolic pathways is increasingly attracting attention and is a central theme in system biology to accomplish four goals which are system structure identification, system behaviour analysis system control and system design (Ko. *et.al*, 2006).

In designing the mathematical modelling of biological pathway, parameter estimation is the most challenging part to retrieve optimal parameter values that could obtain the best fit with the experimental data. Parameter estimation is a concept where a sample data is used to estimate the value of a population's parameter such as mean and variance. Usually, the ordinary differential equation is used in modelling biological data in analysis, prediction, and optimizing the biological system itself. For this research, Differential Evolution with the implementation of Bacterial Foraging algorithm has been design to conduct the parameter estimation on JAK/STAT signal transduction pathway to model the tyrosine production in mus musculus.

Modeling is a process of generating abstract, conceptual, graphical and mathematical models. There are several processes in the biology modeling. In the process of modeling the biological system, the most challenging part is the determination of the model parameter. Furthermore, biological processes and interaction are highly non-linear and complex, hence mathematical analysis is needed to capture the nonlinear of the data. Therefore, parameter estimation plays an important role in modeling the biological system, which however is very difficult. Parameter estimation is used to determine rate constants and kinetic orders so that the dynamic profiles satisfactorily fit the measured observation in the biology system. Basically biological processes are modeled using Ordinary Differential Equations (ODEs) to describe the evolution over time of certain quantities of interest(*Lillacci* and *Khammash*, 2010). Generally, equation depends on several parameters and usually the parameters are unknown.

This research focuses on the optimization result of the kinetic parameter estimation. Yao and William (1994) used Genetic Algorithm (GA) to solve the parameter estimation for linear and nonlinear digital filters and were applied to both feed forward

and recurrent neural network. There was a problem in using the GA stem from its computational complexity and it was trapped in local minimal. Rodolfo *et al.* (2009) studied on the optimal tuning of the parameters of a fuzzy controller for a network based control system. In this research, the SA faced a problem of time consumption for the estimation of the parameter estimation. Moonchai *et al.* (2005) implemented DE as parameter estimation approach by enhancing the production of lactic acid production, glucose consumption, and bacteriocin production. Differential evolution algorithm is developed for the purpose to optimize real parameters and real valued functions. Although differential evolution is a good algorithm in estimating kinetic parameter, there are still challenges where the algorithm may be influenced by noisy data during parameter estimation. The problem of noisy data can be solved by using Kalman Filtering algorithm, where the Kalman Filter can filter the noisy data by updating the population and also improving the performance of parameter estimation. Besides that, the performance of parameter estimation can also be improved by implementing Bacterial Foraging in the algorithm with differential evolution and Kalman Filtering algorithm, where the Bacterial Foraging algorithm help in faster convergences by implementing the reproduction and chemostatic state into mutation and crossover of the Differential Evolution. The reproduction state of the Bacterial Foraging algorithm is implemented in mutation state of DE while the chemostatic state of the Bacterial Foraging Algorithm is modified in the crossover path of DE, where these will help in faster convergence and avoid from being trapped in local minima.

In order to get the best performances of the modeling of tyrosine production, the estimation of the best kinetic parameters should be perform. To get the best value of parameter estimation, differential evolution with Kalman Filter and Bacterial Foraging algorithm are used in this research, where the differential evolution is used to find the true global minimum regardless of the initial parameter values, fast convergence, and using few control parameters (Karaboga and Okdem, 2004). This algorithm has not been implemented in modeling the tyrosine production in mus musculus, and the performance of the implementation of this algorithm is believe in improve the performance in parameter estimation. This algorithm is able to produce best result with shortest computational time and improvement on the accuracy of the parameter estimation.

Material and Method

Based on the previous study, this study proposes the differential evolution algorithm and Kalman Filtering with Bacterial Foraging algorithm, which is a hybrid of IDE and BF. Table 1 shows the difference between the existing algorithm and IDEBF, where the existing algorithm comprises of only DE, whereas IDEBF is a hybrid of IDE and BF and IDE is a hybrid of DE and KF. Fixed control parameter values used in this study are as follow:

I. Population size, *NP*: 10
II. Mutation factor, *F*: 0.5
III. Crossover constant, *CR*: 0.9

Table 1. Difference between existing algorithm with DEBF, IDE and IDEBF

Existing Algorithm	DEBF	IDE	IDEBF
DE	DE+BF	DE+KF	IDE+BF

Note: Shaded column represents the hybrid algorithms proposed in this research

The conventional Differential Evolution algorithm is enhanced with Kalman Filtering algorithm and Bacterial Foraging algorithm. Kalman Filtering would help in updating the population where a new step has been added to the conventional Differential Evolution. In the initialization, the m x n population matrix is generated from the first generation till the maximum generation. m and n represent the number of identifiable parameter and number of generation respectively. Meanwhile in the evaluation process, the fitness function, J is represented as

$$J = \sum_{i=i}^{n} |f(X,u,\emptyset) - f(Y,u,\emptyset)|^2 \quad (1)$$

to evaluate the fitness of the individual. X represents the state vector for measurement system, Y represents the state vector for simulated system, $\emptyset$ represents the set of unknown parameters that are used for parameter estimation, whereas u represents the external force e.g. noisy data, N=the ending index, and i=the index variable.

After that, the updating of the population based on Kalman gain value K is retrieved from Equation 3. Kalman Filtering helps in handing the noisy data and updates the population once again. This is done until the evaluation process meets the stopping criterion. The update population process is carried out by using the formulas below.

$$temp\_population = (temp\_population' + K)' \quad (2)$$

$$K = P * H' * inv(H * P * H' + R) \quad (3)$$

H=observation matrix
R=measurement noise covariance
P=covariance of the state vector estimate
H' =inverse of matrix H

The Bacterial Foraging algorithm is implemented in this mutation and crossover process of the conventional DE where the reproduction and chemostatic state of the Bacterial Foraging algorithm are implemented into the mutation and crossover of the Differential Evolution respectively. The Bacterial Foraging algorithm involved in the mutation step of DE is created by using the following equation:

$$y_j = \begin{cases} \tilde{y}_j + \Delta\left(k, y_j^{(U)} - \tilde{y}_j\right), \tau = 0 \\ \tilde{y}_j - \Delta\left(k, \tilde{y}_j - y_j^{(L)}\right), \tau = 1 \end{cases} \quad (4)$$

where the random constant τ becomes 0 or 1, $y_j^{(U)}$ and $y_j^{(L)}$ is the lower and upper range of y_j and $\Delta(k,w)$ is given as

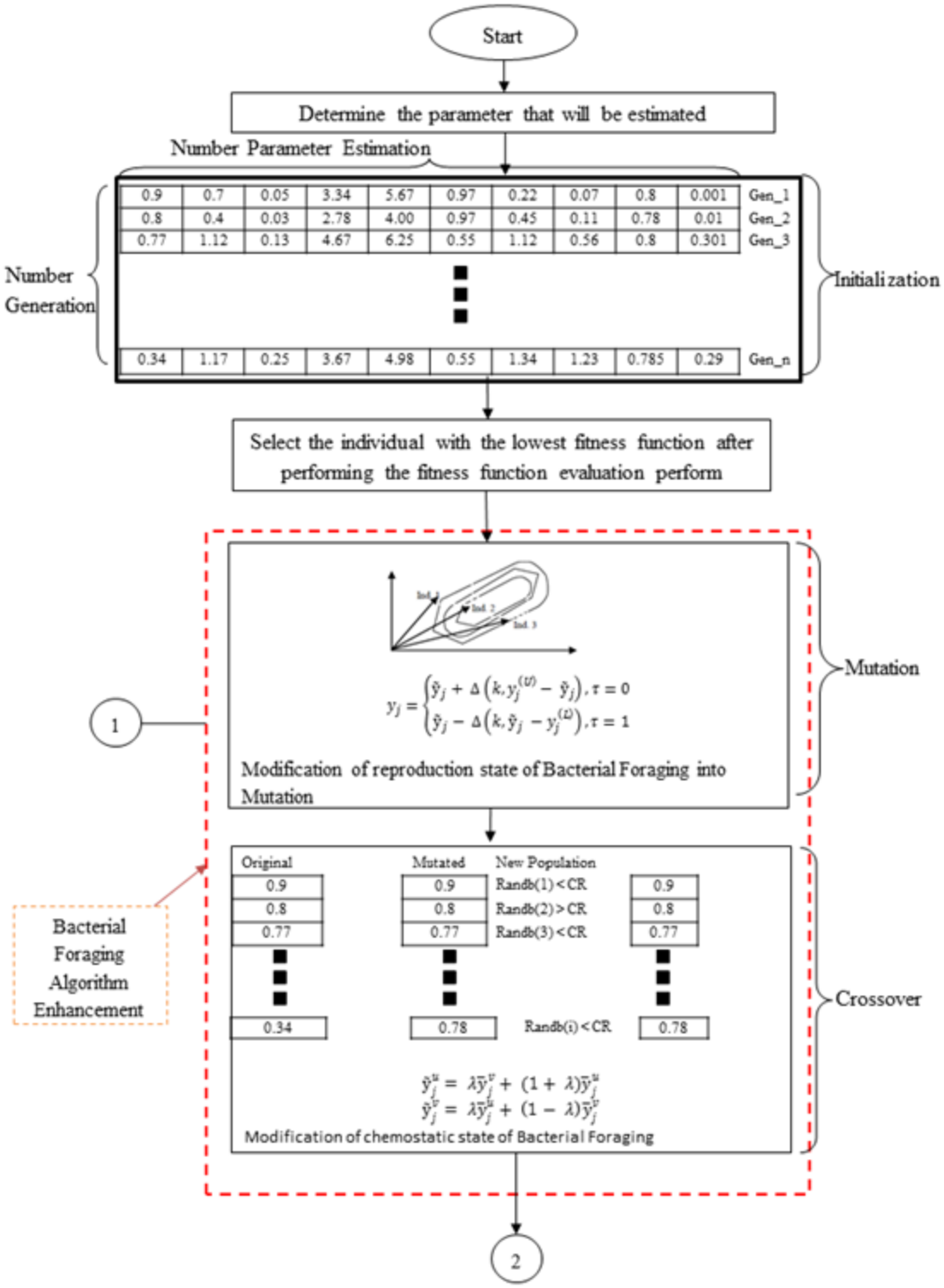
$$\Delta(k,w) = w.\eta.(1 - \frac{k}{z})^A \quad (5)$$

η = 0 or 1 randomly and z is the maximum number of the generations as defined by the user. kth is represented as reproduction state.

A modification in simple crossover is used in DE algorithm using

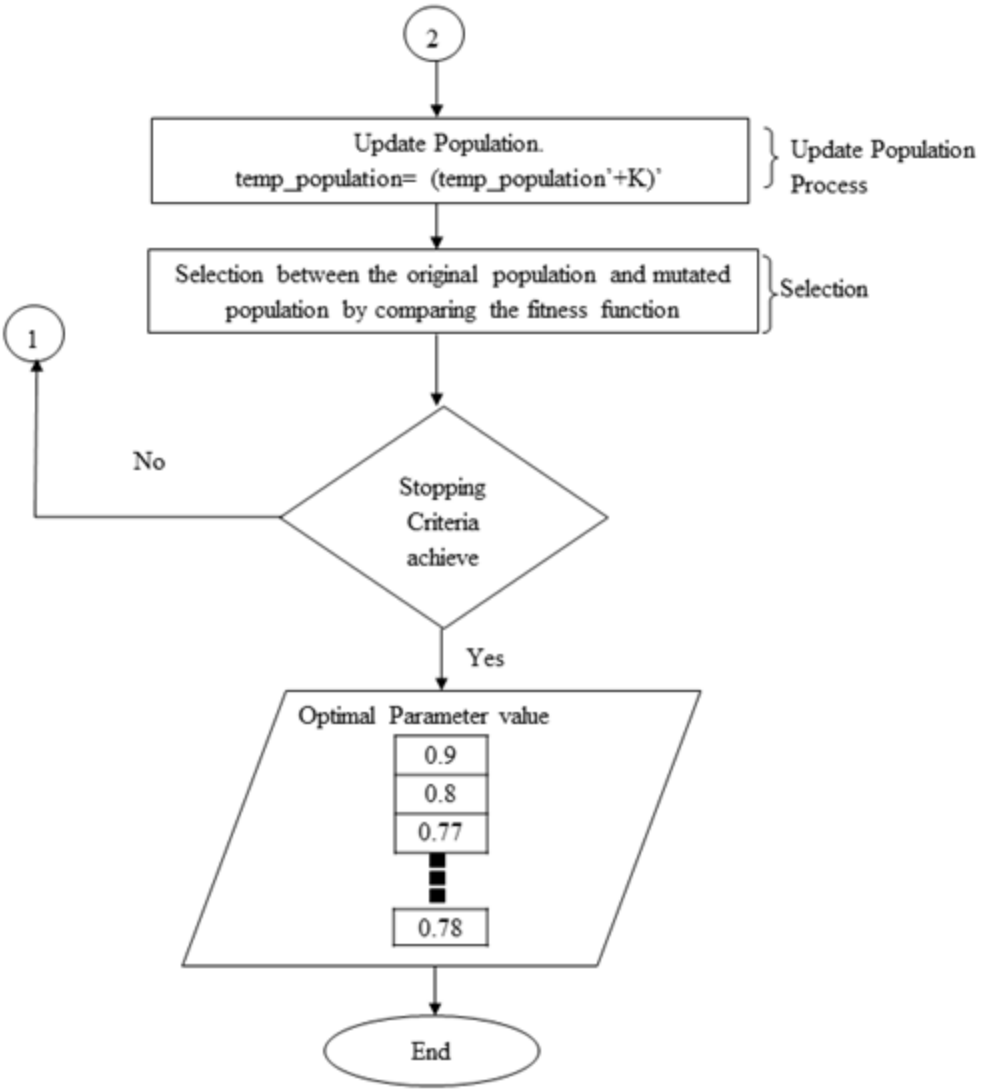
$$\tilde{y}_j^u = \lambda \bar{y}_j^v + (1 + \lambda)\bar{y}_j^u \quad (6)$$

$$\tilde{y}_j^v = \lambda \bar{y}_j^u + (1 - \lambda)\bar{y}_j^v \quad (7)$$



Note: Modification in mutation and crossover by using Bacterial Foraging Algorithm in DE to improve DE performance (highlighted with the dotted box).

Figure 1 (a): Schematic Overview of IDEBF.



Note: Modification in mutation and crossover by using Bacterial Foraging Algorithm in DE to improve DE performance (highlighted with the dotted box).

Figure 1 (b): Schematic Overview of IDEBF.

where $\bar{y}_j^u$ and $\bar{y}_j^v$ refer to parent's generations and $\tilde{y}_j^u$ and $\tilde{y}_j^v$ refer to the offspring's generations and and j is the chromosome of chemotactic step and $\lambda$is the multiplier (Dong *et al*, 2007).

After the improvement of the algorithms, the algorithms will be implemented in the SBToolBox in Matlab and run in the Matlab with the dataset to get the best kinetic parameter estimation. Figure 1 shows the overall process of IDEBF in the estimation of the kinetic parameter values.

Result and Discussion

Five algorithms have been compared in this journal which include Genetic Algorithm (GA), Differential Evolution (DE), Improve Differential Evolution (IDE) ,Differential Evolution and Bacterial Foraging Algorithm (DEBF) and Improved Differential Evolution and Bacterial Foraging Algorithm (IDEBF). To evaluate the accuracy for each estimation algorithm, the kinetic parameter values have been indicated. From the mechanism of JAK/STAT signal transduction pathway (Satoshi *et al*., 2002), SOCS1 is the activator for the tyrosine production, therefore the ordinary differential evolution (ODE) for estimating parameter value for tyrosine production is

$$\frac{d(SOCS1)}{dt} = \frac{v26 - v28 - v29 + v32 + v40 - v42 - v43 - v44}{cytoplasm} \quad (8)$$

where
$v26 = cytoplasm * v26\_kf * mRNAc$, $v28 = cytoplasm * v28\_kf * SOCS1$, $v29 = cytoplasm * (v29\_kf * SOCS1 * IFNRJ2\_star - v29\_kb * IFNRJ2\_star\_SOCS1$,
$v32 = cytoplasm * v32\_kf * IFNRJ2\_star\_SHP2\_SOCS1\_STAT1c$, $v40 = cytoplasm * v40\_kf * IFNRJ2\_star\_SHP2\_SOCS1$, $v42 = cytoplasm * (v42\_kf * SOCS1 * IFNRJ2\_star\_STAT1c - v42\_kb * IFNRJ2\_star\_SOCS1\_STAT1c)$, $v43 = cytoplasm * (v43\_kf * SOCS1 * IFNRJ2\_star\_SHP2 - v43\_kb * IFNRJ2\_star\_SHP2\_SOCS1)$, $v44 = cytoplasm * (v44\_kf * SOCS1 * IFNRJ2\_star\_SHP2\_STAT1c - v44\_kb * IFNRJ2\_star\_SHP2\_SOCS1\_STAT1c)$, cytoplasm = fixed value of 1. IFNRJ2_star, IFNRJ2_star_SHP2_SOCS1_STAT1c, IFNRJ2_star_SHP2_SOCS1, IFNRJ2_star_STAT1c, IFNRJ2_star_SOCS1_STAT1c, IFNRJ2_star_SHP2, IFNRJ2_star_SHP2_STAT1c show the concentration of different activators.

The estimation of the kinetic parameter values is estimated by implementing the estimation algorithm in the SBToolBox of Matlab. The parameter values that are retrieved from Matlab will be substituted in the Copasi with the simulated kinetic parameter values to evaluate the average error rate and standard deviation for estimating the accuracy of the estimation algorithm. Table 2 shows the parameter estimation values for the estimation algorithms.

Table 2 Kinetic parameter values of DEBF compared with GA and DE.

Kinetic parameters	Measurement kinetic parameter values	Simulated kinetic parameter values				
		GA	DE	DEBF	IDE	IDEBF
v26kf	0.0100	0.2884	0.0073	0.0055	0.0044	0.0046
v28kf	0.0005	0.0007	0.0017	0.0001	0.0007	0.0004
v29kf	0.0200	0.0478	0.0216	0.0084	0.0241	0.3436
v29kb	0.1000	0.0975	0.0839	0.6912	0.102	0.5888
v32kf	0.0030	0.0006	0.006	0.0016	0.0015	0.0025

v40kf	0.0030	0.0347	0.0074	0.0014	0.0025	0.0061
v42kf	0.0200	0.0536	0.0979	0.0321	0.1654	0.0045
v42kb	0.1000	0.1859	0.1112	0.0639	0.1091	0.6518
v43kf	0.0200	0.0149	0.0195	0.0428	0.0194	0.0119
v43kb	0.1000	0.0424	0.3522	0.0994	0.0816	0.1151
v44kf	0.0200	0.0054	0.235	0.0199	0.0145	0.0428
v44kb	0.1000	0.1701	0.0883	0.4386	0.1368	0.0424

The time series data for the concentration of SOCS1 is generated from Equation 8. Measurement result, *y*, and simulated result *yi*, are in the time series data for GA, DE, IDE and IDEBF respectively. Equation 9, Equation 10 and Equation 11 show the formula to obtain the Error rate (*e*), Average error rate (*A*), and Standard Deviation (STD) values respectively.

$$e = \sum_{i=0}^{n} (y - yi)^2 \quad (9)$$

$$A = \frac{e}{N} \quad (10)$$

$$STD = \sqrt{\frac{e}{N}} \quad (11)$$

Table 3 displays the average error rate and the standard deviation for five estimation algorithms for the tyrosine production in JAK/STAT signal transduction pathway.

Table 3: Average of error rate and STD values for SOCS1.

Evaluation criteria	GA	DE	DEBF	IDE	IDEBF
Average of error rate, A	2.201E-07	2.687E-07	1.786E-07	1.763E-07	1.682E-07
Standard Deviation, STD	3.640E-07	4.489E-07	3.155E-07	2.882E-07	2.872E-07

Note: Shaded column represents the best results.

Each algorithm was compared through 50 runs for the JAK/STAT signal transduction pathway dataset to retrieve the standard deviation and the average error rate for the SOCS1. From the result displayed in Table 3, IDEBF shows the lowest average of error rate and standard deviation with values of 1.6820E-07and 2.8718E-07 respectively. DE shows the worst performance of the average error rate and the standard deviation among the three estimation algorithms with values of 2.6867E-07 and 4.4891E-07 respectively. Meanwhile, IDE shows the second lowest average of error rate and standard deviation with values of 1.7627E-07and 2.8816E-07 respectively, followed by DEBF with the average of error rate value of 1.7860E-07 and standard deviation value is 3.1548E-07, while GA has values of 2.2095E-07 and 3.6401E-07 for average error rate and standard deviation respectively. The average error rate and the standard deviation values of IDEBF are close to 0. This shows that the result is more consistent and IDEBF shows the best accuracy compared to the other methods where it has the lowest average error rate and standard deviation among all the comparison methods. The hybrid of Kalman Filtering algorithm and Bacterial Foraging algorithm into conventional Differential Evolution algorithm helps in updating the population and faster convergence to retrieve the best kinetic parameter values.

Table 4 below shows the computational time execution for the estimation algorithms using a Core 2 PC with 2GB main memory. According to the result in Table 4, DE shows the worst execution time for the parameter estimation compared to GA, DEBF, IDE and IDEBF algorithms which used 14 minutes and 30 seconds to evaluate the kinetic parameter values. On the hand, IDEBF shows the shortest execution time for the estimation of the kinetic parameter values which only used 6 minutes and 1 second to complete the execution, followed by IDE with execution time of 7 minutes and DEBF with 8 minutes and 13 seconds. The hybrid of Kalman Filtering and Bacterial Foraging algorithm helps in shortening the computational time of the parameter estimation for the JAK/STAT signal transduction pathway dataset.

Table 4: Execution time of DEBF compared with GA and DE

Computation Usage	GA	DE	DEBF	IDE	IDEBF
Execution time (hh.mm.ss)	00:011:20	00:14:30	00:08:13	00:07:00	00:06:01

Note: Shaded column represents the best results.

Figure 2 above shows that the line of the simulated IDEBF is the closest to the experimental result, therefore is the most consistent compared to the other methods. The line of IDE is second closest to the experimental result, followed by DEBF and GA where DEBF and GA are less consistent compared to IDEBF. Meanwhile, the line of the simulated DE is farther apart from the estimation parameter values. Therefore, DE is the least inconsistent compared to the other methods. Kalman Filter helps in handling the noise data by updating the population and the Bacterial Foraging algorithm updates the mutation and crossover of the DE via implemented reproduction, *k*th and chemostatic, *j*th state where it helps in the convergence where modelling creates the tendency for genetic characteristics of populations to stabilize over time. Besides that, the local minima also can be avoided by modifying DE with bacterial foraging algorithm.

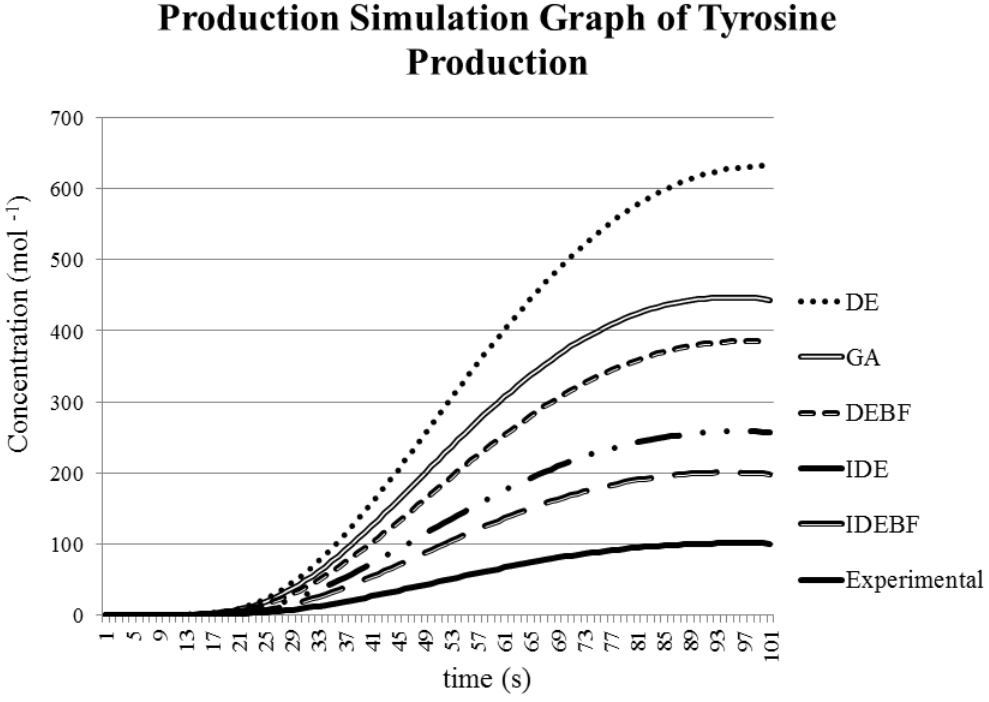


Figure 2 - Comparison of the simulated result with the measurement result of kinetic parameter values.

Conclusion and Future Work

This research has proven that Bacterial Foraging Algorithm help in faster convergences in the conventional differential evolution and helps in to shorten the computational time and good accuracy of the kinetic parameters values where the average error rate and standard deviation value are close to 0. Kalman Filter helps in handling the noise data by using the Kalman gain method while the Bacterial Foraging algorithm helps in faster convergences and avoids being trapped in the local minima in the reproduction state and chemostatic state in the mutation and crossover of differential evolution. Therefore, the hybrid of the Kalman Filtering and Bacterial Foraging algorithm in the Differential Evolution has improved the accuracy of the parameter estimation where the hybrid method has the lowest average error rate and standard deviation and IDEBF has been proven to shorten the computational time as well. In future work, the dataset can be pre-processed before utilizing the kinetic parameter estimation where it helps in shortening the computational time. However, there is only one dataset that has been conducted in this study. For future research, other datasets can be experimented to retrieve the optimal parameter values for the biological pathway.

Acknowledgments

We would like to thank Ministry of Higher Education of Malaysia and Universiti Teknologi Malaysia which sponsored the research grant with Vot number Q.J130000.2507.04H16.

Competing Interests

The authors declare that they have no competing interests.

References

Dong, H.K., Ajith, A., and Jae, H.C. (2007). A hybrid genetic algorithm and Bacterial Foraging approach for global optimization. *Information Science.***177**(18), 3918-3937.

Karaboga , D., and Okdem, S. (2004). A simple and global optimization algorithm for engineering problems: Differential evolution algorithm. *Turkish Journal of Electrical Engineering and Computer Sciences.* **12**, 53-60.

Ko, C.L., Wang, F.S., Chao, Y.P., and Chen, T.W. (2006). S-system Approach to Modeling Recombinant Escherichia coli Growth by Hybrid Differential Evolution with Data Collection. *Biochemical Engineering Journal.* **3**, 174-180.

Lillacci, G., and Khammash, M. (2010). Parameter Estimation and Model Selection in Computational Biology. *PloS Computational Biology.* **6**(3), 1-17.

Moonchai, S., Madlhoo, W., Jariyachavalit, K., Shimizu, H., Shioya, S., and Chauvatcharin, S. (2005) Application of a mathematical model and Differential Evolution

algorithm approach to optimization of bacteriocin production by Lactococcus lactis C7. *Bioprocess Biosyst Eng.* **28**, 1-17.

Rodolfo, E.H., Agustı´n, J., and Ramo´n, G. (2009). An optimal fuzzy control system in a network environment based on simulated annealing. An application to a drilling process. *Applied Soft Computing. Science Direct.* **9**(3), 889-895.

Satoshi, Y., Satoru, S., Akiko, J., and Akihiko, Y. (2002). Control Mechanism of JAK/ STAT signal Transduction Pathway. *Federation of European Biochemical Societies*. *Elsevier Science*. **534**(1-3), 190-196.

Yao L., and William A.S. (1994). Nonlinear Parameter Estimation via the Genetic Algorithm. *IEEE.* **42**(4), 927-935.