Jurnal Teknologi

Full paper

Parameter Estimation Using Improved Differential Evolution And Bacterial Foraging Algorithms To Model Tyrosine Production In *Mus Musculus*(Mouse)

Jia Xing Yeoh[a], Chuii Khim Chong[a], Mohd Saberi Mohamad[a,*], Yee Wen Choon[a], Lian En Chai[a], Safaai Deris[a], Zuwairie Ibrahim[b]

[a]*Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor Malaysia*
[b]Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, 26600 Pekan, Pahang

*Corresponding author: saberi@utm.my

Article history

Received: 6 June 2014
Received in revised form:
29 June 2014
Accepted: 1 December 2014

Graphical abstract

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

Abstract

The hybrid of Differential Evolution algorithm with Kalman Filtering and Bacterial Foraging algorithm is a novel global optimisation method implemented to obtain the best kinetic parameter value. The proposed algorithm is then used to model tyrosine production in *Musmusculus* (mouse) by using a dataset, the JAK/STAT(Janus Kinase Signal Transducer and Activator of Transcription) signal transduction pathway. Global optimisation is a method to identify the optimal kinetic parameter in ordinary differential equation. From the ordinary parameter of biomathematical field, there are many unknown parameters, and commonly, the parameter is in nonlinear form. Global optimisation method includes differential evolution algorithm, which will be used in this research. Kalman Filter and Bacterial Foraging algorithm helps in handling noise data and convergences faster respectively in the conventional Differential Evolution. The results from this experiment show estimated optimal kinetic parameters values, shorter computation time, and better accuracy of simulated results compared with other estimation algorithms.

Keywords: Parameter estimation, differential evolution algorithm, bacterial foraging algorithm, kalman filtering algorithm, modelling, metabolic engineering, bioinformatics, artificial intelligence

1.0 INTRODUCTION

Metabolic pathway can be described by a combination of process types including reversible reactions and, in some respects, multimolecule reactions. Recently, much research has 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: system structure identification, system behaviour analysis, system control, and system design (Koetal., 2006).

In designing the mathematical modelling of biological pathway, parameter estimation is the most challenging part estimate to retrieve optimal parameter values that obtain the best fit with the experimental data. Parameter estimation is a concept where sample data are used to estimate the value of a population's parameter such as mean and variance. Usually, an ordinary differential equation is used in modelling biological data in analysing, predicting, and optimising the biological system itself. For this research, Differential Evolution (DE) with implementation of Bacterial Foraging (BF) algorithm is being designed to conduct the parameter estimation on JAK/STAT signal transduction pathway to model the tyrosine production in Musmusculus.

Modelling is a process of generating abstract, conceptual, graphical, and mathematical models. There are several processes in the biology modelling. From the process of modelling the biological system, the most challenging part is the determination of the model parameter. Furthermore, biological processes and interaction are highly nonlinear and complex; hence, mathematical approach is needed to capture nonlinear data. Therefore, parameter estimation played an important role in the modelling of the biological system, but it was also very difficult. Parameter estimation determines rate constants and kinetic orders so that the dynamic profiles satisfactorily fit the measured observation in the biology system. Basically, biological processes are modelled using Ordinary Differential Equations (ODEs) to describe the evolution of certain quantities of interest over time (Lillacci and Khammash, 2010). In a general equation, there will be several parameters, and usually, the parameter is unknown.

The main problem on this research is focused on the optimisation result of the kinetic parameter estimation. Yao and Sethares (1994) used Genetic Algorithm (GA) to solve the parameter estimation for linear and nonlinear digital filters, which are applied to both feed forward and recurrent neural network. There is a problem in using GA stem from its computational

complexity and trap in local minimal. Rodolfo et al. (2009) introduced optimal tuning of the parameters of a fuzzy controller for a network-based control system. From this research, the Simulated Annealing (SA) is facing a time-consuming problem for parameter estimation. Sompop et al. implemented DE as a parameter estimation approach by enhancing lactic acid production, glucose consumption, and bacterioc in production. DE algorithm is developed for the purpose of optimising real parameters and real valued functions. Although DE 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 (KF) algorithm where Kalman Filter can filter noisy data by updating the population and improving the performance of parameter estimation. Besides that, the performance of parameter estimation can be improved by implementing BF in the algorithm with DE and KF algorithm, where BF algorithm helps in convergences faster by implementing the reproduction and chemostatic state into mutation and crossover of DE. The reproduction state of the BF algorithm was implemented in the mutation state of DE, while the chemostatic state of the BF algorithm was modified in the crossover path of DE; this helps in convergence faster and helps avoid from getting trapped in local minima.

In order to get the best performances of the modelling of tyrosine production, the estimation of the best kinetic parameters should be performed. To get the best value of parameter estimation, DE with Kalman filter and BF algorithm is used in this research, with DE finding the true global minimum regardless of the initial parameter values, fast convergence, and using few control parameters (Karaboga, 2004). These algorithms are not implemented in modelling the tyrosine production in Musmusculus, and the performance of the implementation of this algorithm is believed to improve the performance in parameter estimation. These algorithms are able to produce the best results with shorter computational time and improve the accuracy of the parameter estimation.

2.0 EXPERIMENTAL

From the previous study, this study proposes the DE algorithm and KF with the BF algorithm, which is a hybrid of improved DE (IDE) and BF. Table 1 shows the difference between existing algorithm and IDEBF, where existing algorithm comprises 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 follows:

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

The values are set by conducting a small number of trials within a specific range. The fixed values are the values that generate the best results in this paper.

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

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

Note: The shaded column represents the hybrid algorithm that was proposed in this research.

The conventional DE algorithm was enhanced with the KF algorithm and the BF algorithm. KF would help in updating the population where a new step is being added to the conventional DE. In the initialisation, the $m \times n$ population matrix is generated from the first generation until the maximum generation. The variable m and n represent the number of identifiable parameters and the number of generations, respectively, while in the evaluation process, the fitness function *Jis* represented as

$$J = \sum_{i=i}^{n} |f(X, u, \emptyset) - f(Y, u, \emptyset)|^2 \quad (1.1)$$

to evaluate the fitness of the individual. *X* represents state vector for measurement system, *Y* represents state vector for simulated system, *Ø* represents set of unknown parameters that is used for parameter estimation, 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 is based on Kalman gain value *K* retrieved from equation 1.3. KF helps in handing the noisy data and updating the population once again by undergoing the evaluation process until the stopping criterion is met. The update population process is carried out by the formulas below:

$$temp\_population = (temp\_population' + K)' \quad (1.2)$$

$$K = P * H' * inv(H * P * H' + R) \quad (1.3)$$

K=Kalman gain value
A=state transition matrix
B=input matrix
H=observation matrix
Q=process noise covariance
R=measurement noise covariance
P=covariance of the state vector estimate
H' =inverse of matrix H

The BF algorithm was implemented in this mutation and crossover process of the conventional DE where the reproduction and chemostatic state of the BF algorithm were implemented into the mutation and crossover of the DE, respectively. The BF algorithm is involved in the mutation step of DE by the equation

$$y_j = \begin{cases} \tilde{y}_j + \Delta\left(k, y_j^{(U)} - \tilde{y}_j\right), \tau = 0 \\ \tilde{y}_j - \Delta\left(k, \tilde{y}_j - y_j^{(L)}\right), \tau = 1 \end{cases} \quad (1.4)$$

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

$$\Delta(k, w) = w.\eta.(1 - \frac{k}{z})^A \quad (1.5)$$

η = 0 or 1 is random, and z is the maximum number of the generations as defined by the user. The function kth represents reproduction state.

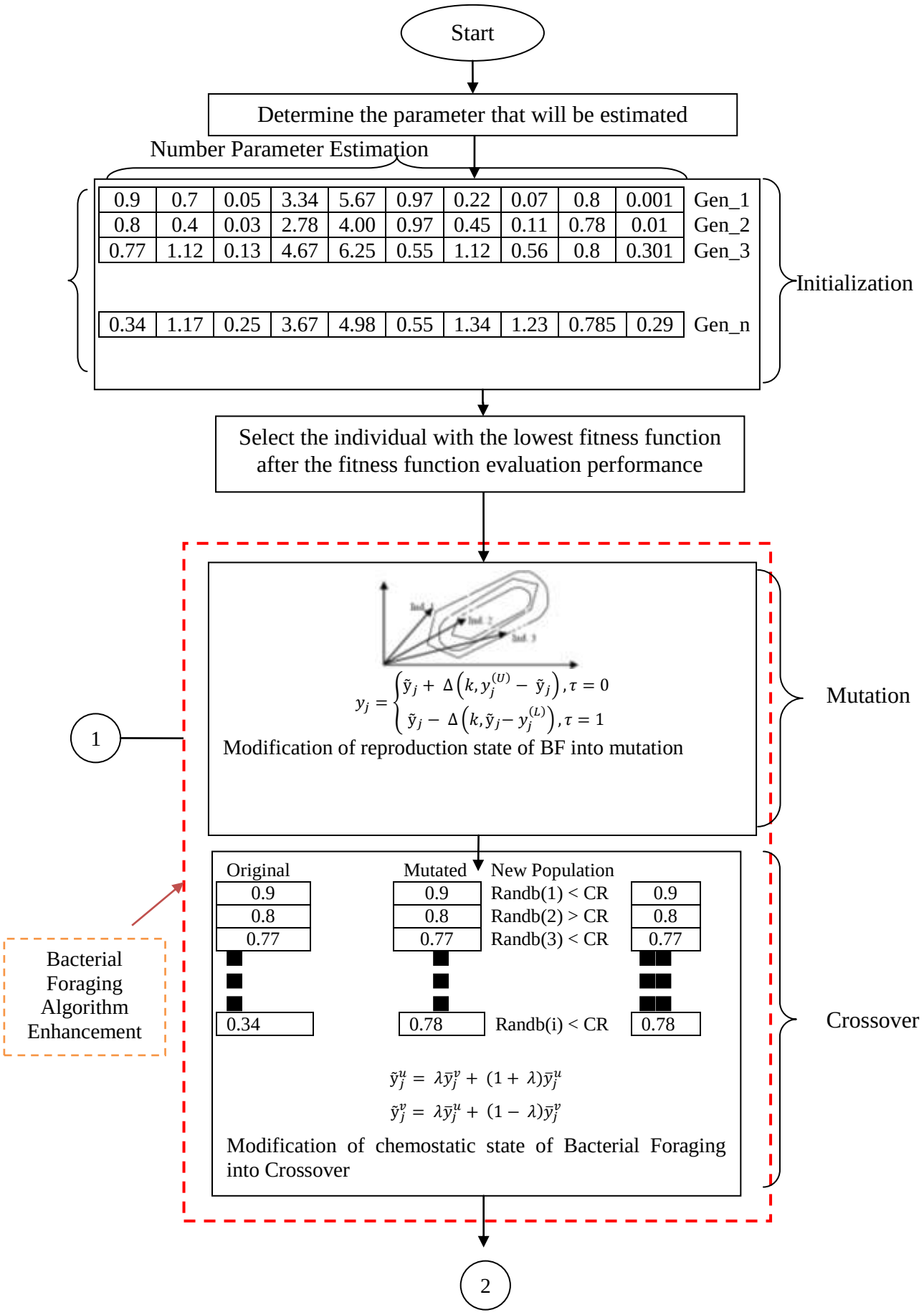
A modification in simple crossover is used in DE algorithm using

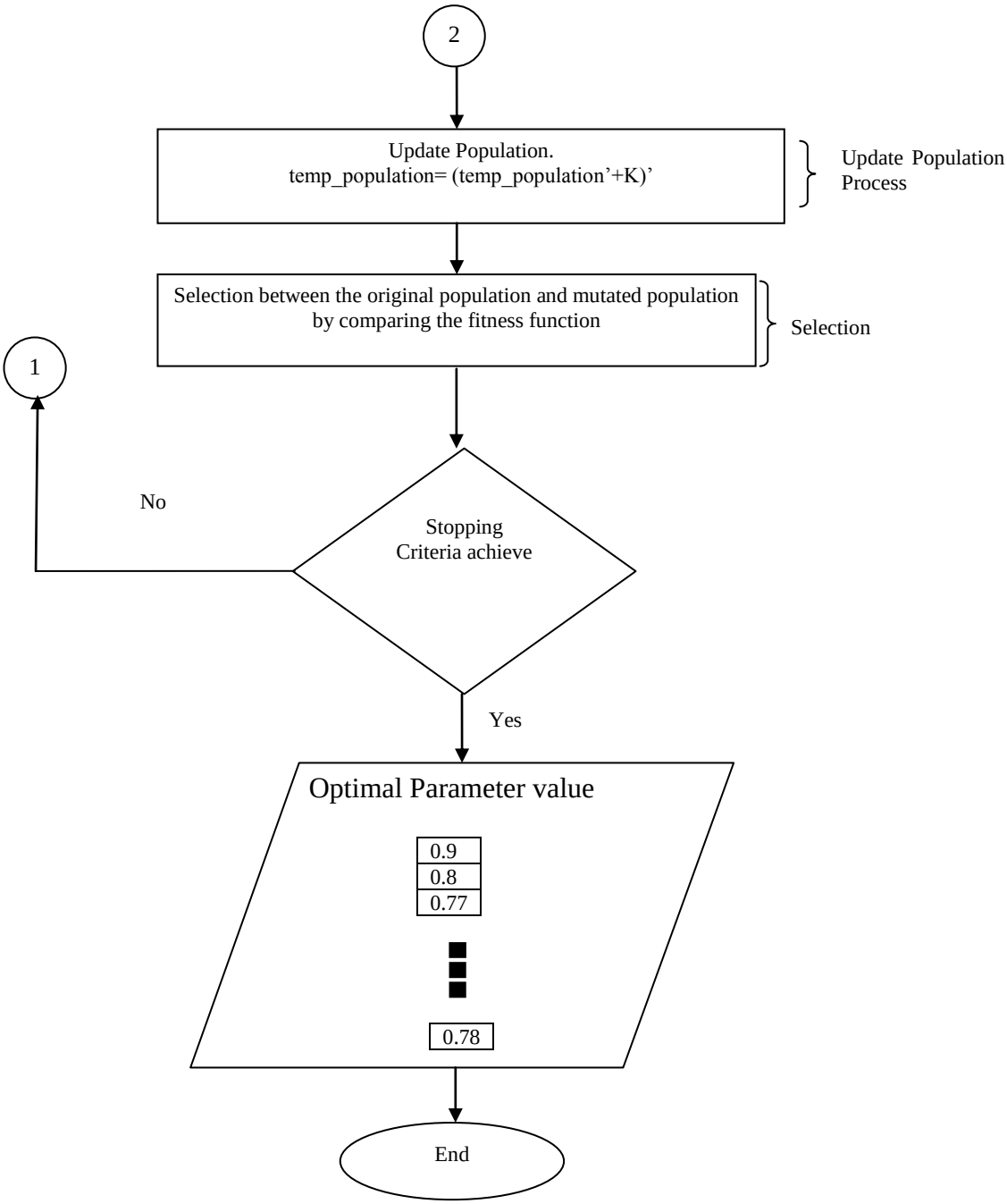
$$\tilde{y}_j^u = \lambda \bar{y}_j^v + (1 + \lambda)\bar{y}_j^u \quad (1.6)$$

$$\tilde{y}_j^v = \lambda \bar{y}_j^u + (1 - \lambda)\bar{y}_j^v \quad (1.7)$$

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 j is the chromosome of chemotactic step and λ is the multiplier (Dong et al., 2007).

After the improvement of the algorithm, 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 estimating the kinetic parameter values.


Start
Determine the parameter that will be estimated
Number Parameter Estimation
0.9 0.7 0.05 3.34 5.67 0.97 0.22 0.07 0.8 0.001 Gen_1
0.8 0.4 0.03 2.78 4.00 0.97 0.45 0.11 0.78 0.01 Gen_2
0.77 1.12 0.13 4.67 6.25 0.55 1.12 0.56 0.8 0.301 Gen_3
0.34 1.17 0.25 3.67 4.98 0.55 1.34 1.23 0.785 0.29 Gen_n
Initialization
Select the individual with the lowest fitness function after the fitness function evaluation performance
$y_j = \begin{cases} \tilde{y}_j + \Delta\left(k, y_j^{(U)} - \tilde{y}_j\right), \tau = 0 \\ \tilde{y}_j - \Delta\left(k, \tilde{y}_j - y_j^{(L)}\right), \tau = 1 \end{cases}$
Modification of reproduction state of BF into mutation
Mutation
1
Original
Mutated
New Population
0.9 0.9 Randb(1) < CR 0.9
0.8 0.8 Randb(2) > CR 0.8
0.77 0.77 Randb(3) < CR 0.77
0.34 0.78 Randb(i) < CR 0.78
$\tilde{y}_j^u = \lambda \bar{y}_j^v + (1 + \lambda)\bar{y}_j^u$
$\tilde{y}_j^v = \lambda \bar{y}_j^u + (1 - \lambda)\bar{y}_j^v$
Modification of chemostatic state of Bacterial Foraging into Crossover
Crossover
Bacterial Foraging Algorithm Enhancement
2



Note: Modification in mutation and crossover by the BF algorithm in DE to improve DE performance, and it is highlighted with the dotted box.

Figure 1 Schematic Overview of IDEBF

3.0 RESULTS AND DISCUSSION

Five algorithms are compared in this journal: GA, DE, IDE,DE and BF algorithm (DEBF), and IDE and BF algorithm(IDEBF). To evaluate the accuracy for each of the estimation algorithm, the kinetic parameter values have been indicated. From the mechanism of JAK/STAT signal transduction pathway, SOCS1 is the activator for the tyrosine production; therefore, the ODE for estimating parameter value for tyrosine production is

$$\frac{d(SOCS1)}{dt} = \frac{v26 - v28 - v29 + v32 + v40 - v42 - v43 - v44}{cytoplasm} \quad (2.1)$$

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 showed the concentration of different activator.

The kinetic parameter values are being estimated by the implementation of the estimation algorithm in the SBToolBox of Matlab. The parameter values 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 1.1 shows the parameter estimation values for the estimation algorithms.

Table 1.1 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 were generated from equation 6. The measurement result, *y*, and the simulated result, *yi*, were in the time series data for GA, DE, IDE, and IDEBF, respectively. Equations 2.2, 2.3, and 2.4 show the formula to get the error rate (*e*), average error rate (*A*), and standard deviation (STD) value, respectively.

$$e = \sum_{i=0}^{n} (y - yi)^2 \quad (2.2)$$

$$A = \frac{e}{N} \quad (2.3)$$

$$STD = \sqrt{\frac{e}{N}} \quad (2.4)$$

Table 1.2 displays the average error rate and the standard deviation for five estimation algorithms for the tyrosine production in JAK/STAT signal transduction pathway.

Table 1.2 Average of error rate and STD values for SOCS1

Evaluation criteria	GA	DE	DEBF	IDE	IDEBF
Average of error rate, *A*	2.2095E-07	2.6867E-07	1.7860E-07	1.7627E-07	1.6820E-07
Standard deviation, STD	3.6401E-07	4.4891E-07	3.1548E-07	2.8816E-07	2.8718E-07

Note: Shaded column represents the best results.

Each of the algorithms was compared in 50 runs for the JAK/STAT signal transduction pathway dataset to retrieve the standard deviation and the average error rate for SOCS1. From the result display in Table 1.2, IDEBF showed the lowest average of error rate and standard deviation with the value of 1.6820E-07and 2.8718E-07, respectively. DE showed the worst performance of the average error rate and the standard deviation among the three estimation algorithms with a value of 2.6867E-07 and 4.4891E-07, respectively. Besides that, IDE showed the second lowest average of error rate and standard deviation with a value of 1.7627E-07and 2.8816E-07, respectively, followed by DEBF with an average error rate value of 1.7860E-07 and standard deviation value of 3.1548E-07, while GA has a value 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 are nearly 0; this showed that the result is more consistent, and IDEBF shows the best accuracy compared with the other methods where it has the lowest average error rate and standard deviation among all the comparison methods. The hybrid of the KF algorithm and the BF algorithm into the conventional DE algorithm helps in updating the population and convergence faster to retrieve the best kinetic parameter values.

Table 1.3 below shows the computational time execution for the estimation algorithms on a Core 2 PC with 2GB main memory. According to the result in Table 1.3, DE showed the worst execution time for the parameter estimation compared with GA, DEBF, IDE, and IDEBF algorithms, where it used 14 minutes and 30 seconds to evaluate the kinetic parameter values. On the other hand, IDEBF showed 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 an execution time of 7 minutes and DEBF with 8 minutes and 13 seconds. The hybrids of the KF and BF algorithm help in shortening the computational time of the parameter estimation for the JAK/STAT signal transduction pathway dataset.

Table 1.3 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.

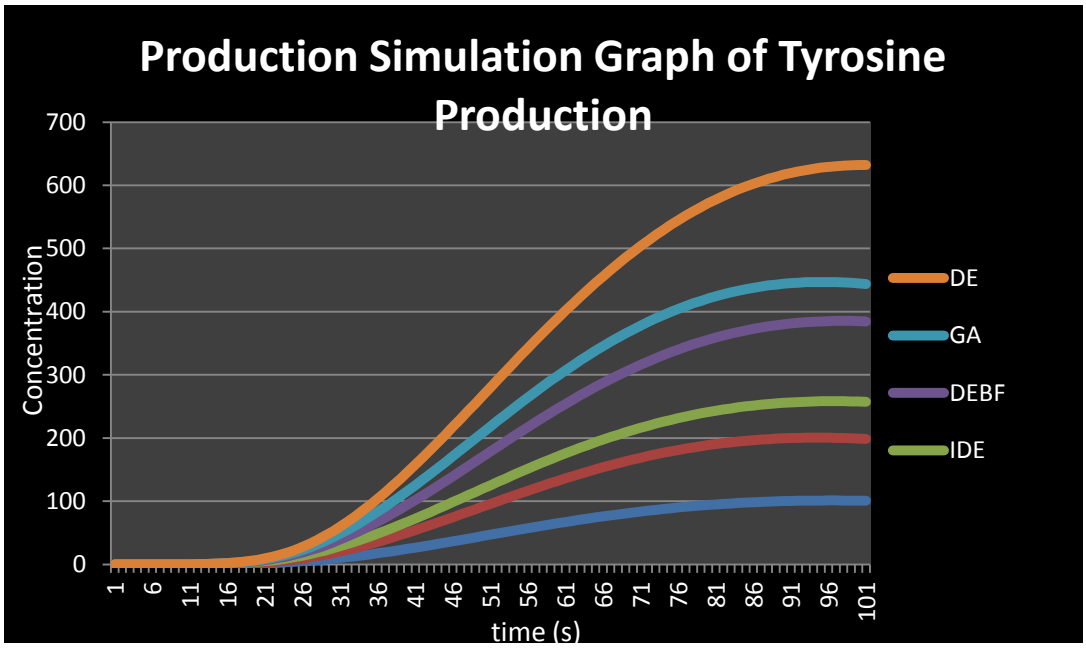


Figure 2 Comparison of the simulated result with the measurement result of kinetic parameter values.

Figure 2 above shows that the line of the simulated IDEBF is closest to the experimental result; therefore, it is more consistent compared with 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 than IDEBF, while the line of the simulated DE is farther apart from the estimation parameter values. Therefore, DE inconsistencies are compared with the other methods. Kalman filter helps in handling the noise data by updating the population, and the BF algorithm updated the mutation and crossover of DE by implementing reproduction, *k*th and chemostatic, *j*th state, where it helps in convergence in modelling the tendency for genetic characteristics of populations to stabilise over time. Besides that, the local minima can also be avoided through the modification of DE with the BF algorithm.

■4.0 CONCLUSION AND FUTURE WORK

This has proven that the BF algorithm helps in convergences faster in the conventional DE and helps in shortening the computational time and good accuracy of the kinetic parameter values where the average error rate and standard deviation value is close to 0. Kalman filter helps in handling the noise data by using Kalman gain method while the BF algorithm helps in faster convergences and avoids getting trapped in local minima where the reproduction state and the chemostatic state are in the mutation and crossover of DE. Therefore, the hybrid of the KF and BF algorithm in DE improves the accuracy of the parameter estimation where the hybrid method has the lowest average error rate and standard deviation and IDEBF is proven to shorten the computational time as well. In future works, the dataset can be preprocessed before undergoing the kinetic parameter estimation, where it helps in shortening the computational time. Besides that, there is only one dataset that is being conducted in this study; in the future, other datasets can be used to experiment to retrieve the optimal parameter value for the biological pathway.

Acknowledgements

We would like to thank the Universiti Teknologi Malaysia for supporting this research through a GUP research grant (Grant number: Q.J130000.2507.05H50) and Matching grant (Grant number: Q.J130000.3007.00M27).

References

[1] Al-Hamadi, H.M., and Soliman, S.A. 2003. Short-Term Electric Load Forecasting Based On Kalman Filtering Algorithm With Moving Window Weather And Load Model. Power System Research Group, Electrical Engineering Department, College of Engineering, University Qatay. *Science Direct, Elsevier*. 47–59.

[2] Chiou, J.P., and Wang, F.S. 2001. Estimation of Monod model parameter by hybrid differential evolution. *Bioprocess and Biosystem Engineering*. 24.109–113.

[3] Dong, H.K., Ajith, A., and Jae, H.C. 2007. A hybrid genetic algorithm and bacterial foraging approach for global optimization. *Science Direct, Elsevier*. 3918–3937.

[4] Gabriele, L., and Mustafa K. 2010. Parameter estimation and model selection in computational biology. *PloS Computational Biology*. 6(3): 17.

[5] Garrett, M.M., David, S.G., Robert S.H., Ruth, H., William, E.H., Richard, K.B., and Arthur, J.O. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*.19: 1639–1662.

[6] Gultyaev, A.P., van Batenburg, F.H., and Pleij, C.W. (1995). *The computer simulation of RNA folding pathways using a genetic algorithm. Journal of Molecular Biology*. 250(1):37–51.

[7] I-Chun, C., and Voit, E.O. 2009. Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Elsevier*. 57–83.

[8] 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.

[9] Ko, C.L., Wang, F.S., Chao, Y.P., and Chen, T.W. 2006. S-system approach to modeling recombinant Escherichia coligrowth by hybrid differential evolution with data collection. *Biochemical Engineering Journal*. 3:174–180.

[10] Pedro, M., and Dauglas, B.K. 1998. *Non-Linear Optimization Of Biochemical Pathways: Applications To Metabolic Engineering And Parameter Estimation.* Oxford University Press. 14(10): 869–883.

[11] Rodolfo, E.H., Agustín, J., and Ramón, 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.* 889–895.

[12] 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.* 190–196.

[13] Schmidt, H., and Jirstand, M. 2005. *Systems Biology Toolbox For MATLAB: A Computational Platform For Research In Systems Biology*. Oxford University Press. 22(4): 514–515.

[14] Sompop, M., Weeranuch, M., Kanidtha, J., Hiroshi, S., Suteaki, S., and Somchai C. 2005. Application of a mathematical model and differential evolution algorithm approach to optimization of bacteriocin production by Lactococcuslactis C7. *Springer-Verlag.* 15–26.

[15] Sun, X., Li, J., and Xiong, M. (2008*).* Extended Kalman filtering for estimation of parameters in nonlinear space models of biochemical networks. *PLoS ONE*. 3(11).

[16] Svergun, D. I. (1999). *Restoring Low Resolution Structure Of Biological Macromolecules From Solution Scattering Using Simulated Annealing*. Biophysical Society. 2879–2886.

[17] Wang, F.S., and Chiou, J.P. 2005. Differential evolution for dynamic optimization of differential-algebraic systems. *IEEE.* 531–536.

[18] Yao, L., and Sethares, W.A. 1994. Nonlinear parameter estimation via the genetic algorithm. *IEEE.* 927–935
.