BioSystems 162 (2017) 81–89

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

An improved hybrid of particle swarm optimization and the gravitational search algorithm to produce a kinetic parameter estimation of aspartate biochemical pathways

Ahmad Muhaimin Ismail [a], Mohd Saberi Mohamad [c,g,h,*], Hairudin Abdul Majid [a], Khairul Hamimah Abas [b], Safaai Deris [c,g,h], Nazar Zaki [d], Siti Zaiton Mohd Hashim [e], Zuwairie Ibrahim [f], Muhammad Akmal Remli [a]

[a] Artificial Intelligence and Bioinformatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
[b] Department of Control and Mechatronic Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia,
[c] Faculty of Creative Technology and Heritage, Universiti Malaysia Kelantan, Bachok, Kelantan, Malaysia,
[d] College of Information Technology, United Arab Emirate University, Al Ain, United Arab Emirates
[e] Soft Computing Research Group, Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
[f] Faculty of Electrical and Electronic Engineering, Universiti Malaysia Pahang, Pekan, Pahang, Malaysia,
[g] Center For Computing and Informatics, Universiti Malaysia Kelantan, City Campus, Pengkalan Chepa, 16100 Kota Bharu, Kelantan, Malaysia
[h] Institute For Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, City Campus, Pengkalan Chepa, 16100 Kota Bharu, Kelantan, Malaysia

ARTICLE INFO

Article history:
Received 6 October 2016
Received in revised form 23 June 2017
Accepted 21 September 2017
Available online 23 September 2017

Keyword:
Parameter estimation
Biochemical pathway
Particle swarm optimization
Gravitational search algorithm
Artificial intelligence
Bioinformatics
Metabolic engineering

ABSTRACT

Mathematical modelling is fundamental to understand the dynamic behavior and regulation of the biochemical metabolisms and pathways that are found in biological systems. Pathways are used to describe complex processes that involve many parameters. It is important to have an accurate and complete set of parameters that describe the characteristics of a given model. However, measuring these parameters is typically difficult and even impossible in some cases. Furthermore, the experimental data are often incomplete and also suffer from experimental noise. These shortcomings make it challenging to identify the best-fit parameters that can represent the actual biological processes involved in biological systems. Computational approaches are required to estimate these parameters. The estimation is converted into multimodal optimization problems that require a global optimization algorithm that can avoid local solutions. These local solutions can lead to a bad fit when calibrating with a model. Although the model itself can potentially match a set of experimental data, a high-performance estimation algorithm is required to improve the quality of the solutions.

This paper describes an improved hybrid of particle swarm optimization and the gravitational search algorithm (IPSOGSA) to improve the efficiency of a global optimum (the best set of kinetic parameter values) search. The findings suggest that the proposed algorithm is capable of narrowing down the search space by exploiting the feasible solution areas. Hence, the proposed algorithm is able to achieve a near-optimal set of parameters at a fast convergence speed. The proposed algorithm was tested and evaluated based on two aspartate pathways that were obtained from the BioModels Database. The results show that the proposed algorithm outperformed other standard optimization algorithms in terms of accuracy and near-optimal kinetic parameter estimation. Nevertheless, the proposed algorithm is only expected to work well in small scale systems. In addition, the results of this study can be used to estimate kinetic parameter values in the stage of model selection for different experimental conditions.

1. Introduction

Explaining the complex network biological processes that are characterized by dynamic behavior is one of the main issues in systems biology (Lillacci and Khammash, 2010; Raue et al., 2015). Pathways are used to describe the relationship between

* Corresponding author.
E-mail address: saberi@umk.edu.my (M.S. Mohamad).

http://dx.doi.org/10.1016/j.biosystems.2017.09.013
0303-2647/

parameters as a means of understanding the complex processes that are involved in biological systems, and mathematical models are commonly used to describe dynamic biological processes. Mathematical models can generate and predict the outcomes of the experimental hypotheses that can be employed to analyze the processes. The application of these models has enabled the construction of metabolic pathways. This phenomenon opens up opportunities for the optimization of metabolite productions in metabolic pathways (Ismail et al., 2015). Normally, these models are based on time-derivative expressions, especially ordinary differential equations (ODE), that describe a change in a state or a quantity of interest over time.

Generally, these models consist of a set of parameters that describe the physical properties of a dynamic system like the rate of reactions. Measuring these parameters is usually difficult and even impossible in some cases (Fernández Slezak et al., 2010). The parameters are often predicted based on fitting the estimated data of the model output with experimental time-series data. The goal of this fitting process is to minimize the errors between these two sets of data by adjusting the parameter values of the model (Rodriguez-Fernandez et al., 2006a). However, these experimental data are often incomplete and suffer from experimental noise (Villaverde et al., 2015). This drawback makes it challenging to find the best-fit parameters that adequately represent the actual biological processes involved. It is crucial that the best parameter values for the biochemical models are estimated and obtained by refining the model parameter values (Schilling et al., 2016). These parameter values are usually identified and measured through costly and time-consuming wet-lab experiments (Tashkova et al., 2011). Alternatively, these parameters can also be estimated using computational approaches. Thus, the estimation of the parameters can be converted into multimodal optimization problems, and global optimization algorithms are required to avoid local solutions (Banga, 2008; Sun et al., 2012). These local solutions can lead to poorly fitting data that the model itself can potentially match accurately with a set of experimental data.

Global optimization algorithms employ stochastic searching strategies to identify a set of possible solutions that are randomly selected based on the given search space. Furthermore, these algorithms are widely used to estimate parameters for various biological models (Chassagnole et al., 2001; Curien and Bastien, 2009; Galazzo and Bailey, 1990; Sun, 2012). Particle swarm optimization (PSO) (Ng et al., 2013; Shi and Eberhart, 1999), the Bee algorithm (BA) (Leong et al., 2013; Pham et al., 2006), the Firefly Algorithm (FA) (Yang, 2009), differential evolution (DE) (Chong et al., 2014), scatter search (SS)(Rodriguez-Fernandez et al., 2006b), simulated annealing (SA) (Villaverde et al., 2012), and others, have already been used to estimate the parameters involved in various biological system models. The main advantages of these models are that they offer researchers the ability to find the best and easiest ways to implement solutions for high-dimensional problems. Despite these advantages, these algorithms often suffer from high computational costs as they try to obtain a global optimum within the large search space (Baker et al., 2010; Fong, 2014; Sun, 2012). In addition, the generated solutions might not represent the actual near-optimal solutions.

In multimodal optimization problems, PSO is often stuck in local optimal, which is the result of a poor global search. The standard gravitational search algorithm (GSA) also has some drawbacks; for example, it has a poor convergence if the initial population is not well generated (Kumar and Sahoo, 2014). Moreover, it often incurs a large computational cost due to the large searching space. Thus, a hybrid of particle swarm optimization and the gravitational search algorithm (PSOGSA) is proposed, which combines the social thinking (*gbest*) ability of PSO with the exploration capability of GSA. This hybrid is able to perform well in optimization problems, especially in minimization problems (Mirjalili and Hashim, 2010). Nevertheless, PSOGSA often incurs a high computational cost when obtaining the global optimum solutions. Besides, the advantages of employing PSO capability in the hybrid in terms of its rapid convergence speed are also weakened (Shanhe and Zhicheng, 2014). Standard and previous algorithms of parameter estimation that have been employed to deal with noisy data often suffer from such poor solutions, and there are typically high errors between experimental and estimated outputs. Hence, a high-performance optimization algorithm is required to maintain fast convergence frequently and improve the quality of the solutions.

This paper proposes an improved hybrid of PSOGSA (IPSOGSA). This improvement has enhanced the search for a global optimum (the best set of kinetic parameter values) by reducing the searching space and focusing the search on the high possibility of feasible solution areas. Hence, there are upsurges in the performance of the proposed algorithm with the advantage of fast convergence in obtaining the global optimum and near-optimum solutions.

The paper is structured as followed. First, we present a problem formulation on the parameter estimation in kinetic model. Then, we present the description of the proposed algorithm phases accompanied with the details on each phase. Next, the experimental setup is explained and consists of the description of model case studies, parameter setting and performance evaluation for the estimation results. We then present the result and discussion section that discuss the results and findings from this study. Finally, the paper is summarized in the conclusion section.

2. Materials and algorithms

The problems that are inherent in biological system estimation will be briefly formulated and explained in this section before the IPSOGSA and experimental setups are explained.

2.1. Problem formulated

The aim of the parameter estimation problem is to attain the near-optimal set of parameters that can minimize the differences between the estimated model output and the experimental time series data. Usually, the nonlinear least squares error function is implemented to minimize differences. Parameter estimation for biological systems can be expressed as per Eqs. (1)–(3) (Lillacci and Khammash, 2010). Where s is the compound in the biochemical system model $s(x)$, which comprises a set of parameters $x = \{x_1, x_2, .. x_n\}$ where n is the number of parameters. The reaction rate of compound s can be represented as a series ODE in the following form:

$$\frac{ds}{dt} = g(s(u,x), t), \tag{1}$$

$$s(t_0) = s(0), \tag{2}$$

$$y = h(s(u,x), t) + e, \tag{3}$$

where g and h are the nonlinear functions, t is the sampling time, and e is the generated measurement noise by random Gaussian noise $N(0,1)$, while y is the rate of reaction and $s(x)$ is the biochemical compounds with set of parameter x. On the other hand, u is the input signal to the reaction of s process.

2.2. An improved hybrid of particle swarm optimization and the gravitational search algorithm (IPSOGSA)

In PSOGSA, the standard PSO has been improved through modifying the process by which acceleration is calculated, before employing this to update the velocity and population process. As with PSO, this hybrid also carries other operations, such as initialization, update velocity, and position. This hybrid adopts the

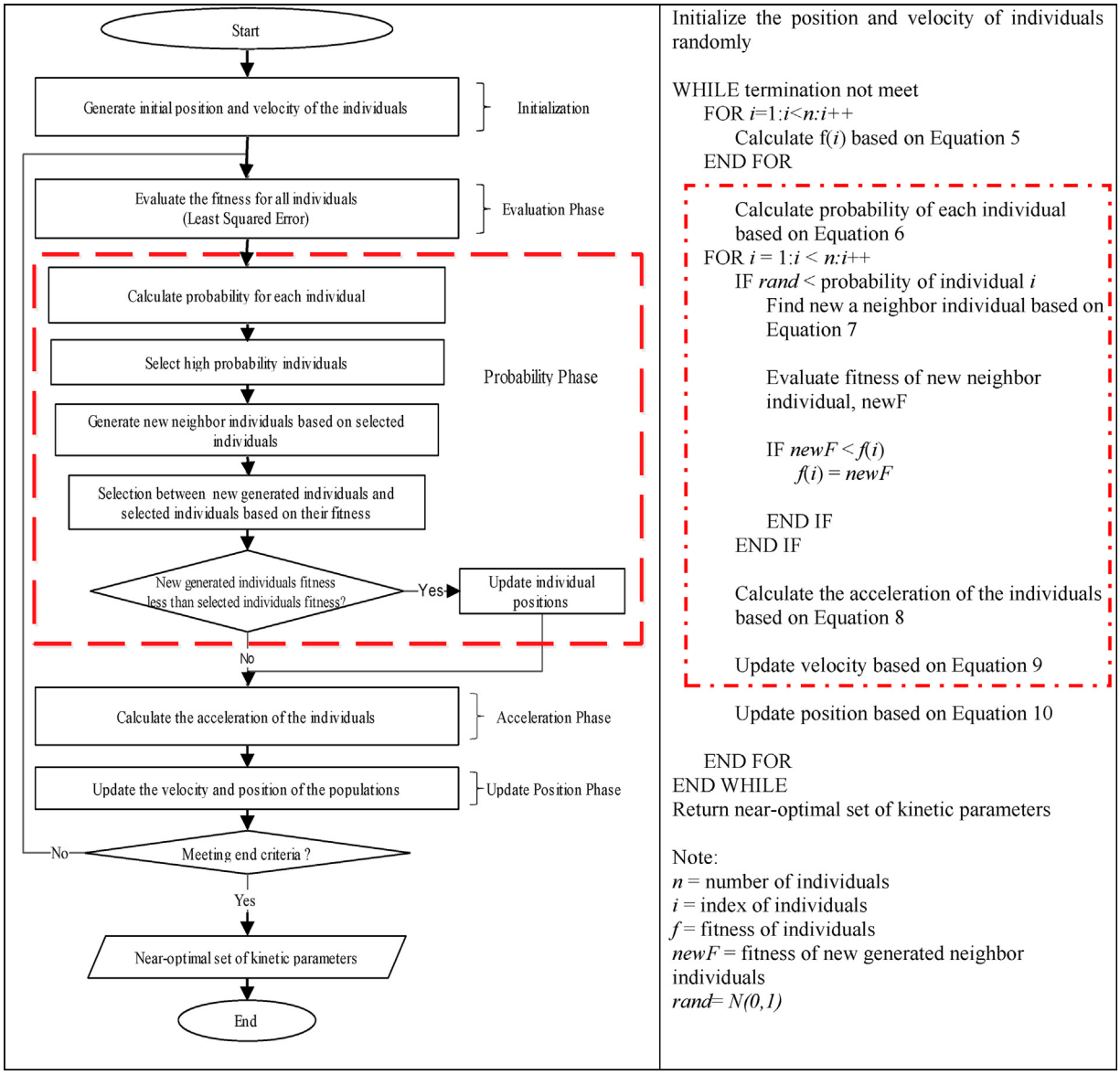

Fig. 1. The flowchart and pseudocode of IPSOGSA (Note: Dotted line describes the improved part).

process of acceleration in GSA by using gravitational theory to update its velocity.

In IPSOGSA, the onlooker phase in the artificial bee colony (ABC) algorithm is applied within the PSOGSA. The onlooker phase of ABC is a local search that utilizes high-possibility solutions to find the new neighboring best solutions (Karaboga and Akay, 2009). Thus, IPSOGSA is proposed as a means of searching for any solution in high-probability areas. Fig. 1 presents the detail of an improved version of the IPSOGSA.

2.2.1. Initialization Phase

In the first step of the initialization phase, the initial position of the population on the problem boundary is randomly generated, as shown in Eq. (4). Each individual position (x) is a representation of the candidate set of kinetic parameter values of the target metabolites. The initialization of the population is illustrated in Fig. 2.

$$\text{position} = \text{R and } (n,\ m) \times (up - low) + low \tag{4}$$

where the position is the position of individuals in a population, Rand represents the n x m matrix with normally distributed random values between 0 and 1. The representations of n and m are the number of individuals and the number of kinetic parameters to be estimated respectively. Whereas *up* and *low* are the upper and lower boundaries of each search space respectively.

2.2.2. Evaluation phase

In the evaluation phase, each individual in a population is assessed based on their fitness. The fitness cost is the sum of squares error between the experimental and simulated data. Thus, the least squares error is used, as per Eq. (5). Fig. 3 shows the details of the evaluation process.

$$f(x) = min \sum_{i}^{n} \left(y_i^{exp} - y_i\right)^2 \tag{5}$$

where n is the total number of samples (maximum generation value) and i is the index variable. Whereas y^{exp} is experimental time series data and y is estimated time series data.

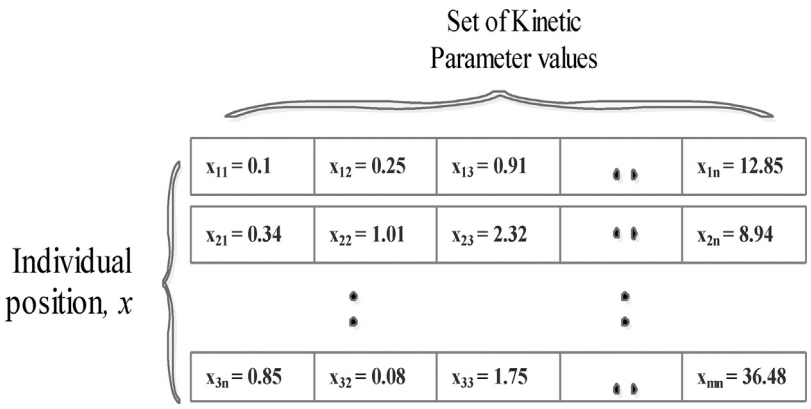

Fig. 2. Initialization of random population.

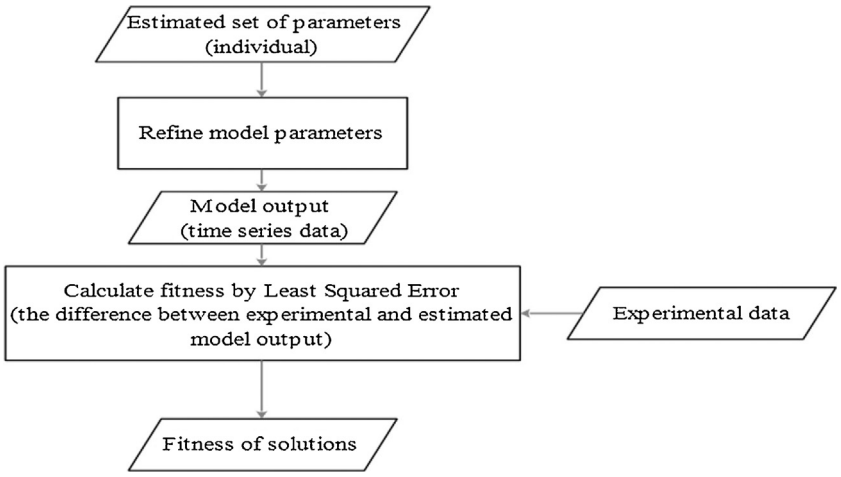

Fig. 3. Evaluation process calculation.

2.2.3. Probability Phase (Improved part)

In this phase, each probability value of the individual is based on their fitness. The probability values indicate the chance of the individual solutions within the feasible solution areas. The probability calculation of the individuals relies on Eq. (6). The individuals with a higher probability will have a higher chance of being selected. The selection criterion is based on a random number between 0 and 1 for each iteration. As the proposed algorithm is a stochastic method, the position of global solution (optimal) is unknown for each iteration. Hence, the random condition is used instead of prior defined condition as the selection criterion. Once an individual is selected, a new individual will be generated within the selected neighborhood using Eq. (7). If the fitness of the newly generated individual is better than the current one, it will replace the current one or vice versa. The selection of individual fitness depends on greedy selection. Then, the global best values are chosen based on the minimum fitness value that is generated within the population. The highlighted element of Fig. 1 shows the summarization in this phase.

$$p_i = 0.9 \times \frac{\min(fit)}{fit_i} + 0.1 \tag{6}$$

where p_i is the probability for individual i, and fit_i is the fitness value of that particular ith individual. Then, $\min(fit)$ is the minimum fitness value (current global near-optimal) in the population.

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right) \tag{7}$$

where k and i are the indexes of individuals. Whereas j is the index of solution dimensions and the value of ϕ_{ij} is a random number between [-1, 1], which controls the generated neighbor individual. Then, x_{ij} is the selected individual and x_{kj} is the randomly selected individual. In this equation, v_{ij} is the newly generated individual.

2.2.4. Acceleration phase

In this phase, the accelerations are calculated as per the calculation employed in standard GSA (Rashedi et al., 2009) and PSOGSA (Mirjalili and Hashim, 2010). The acceleration calculation is based on the law of motion, where the acceleration of an individual is proportional to the resultant force and is inverse to the mass. The resultant force and the mass of the individual are calculated as shown in Eqs. (8) and (13) respectively.

$$F_i = \sum_{j=1, j\neq i}^{n} rand_j F_{ij} \tag{8}$$

where F_i is the total of F_{ij} (resultant force) that acts on the individual i from individual j and n is the total number of individuals. The resultant force is calculated as shown in Eqs. (9) and (10).

$$G = G0\exp\left(-\alpha \times \frac{iter}{maxiter}\right), \tag{9}$$

$$F_{ij} = G\frac{M_{pi}M_{ai}}{R_{ij} + \varepsilon}\left(x_j - x_i\right) \tag{10}$$

where M_p and M_a are the passive and active gravitational mass related to individual i, G is the gravitational constant and R is the Euclidian distance between individuals i and j. Then, x is the position of an individual. Whereas $G0$ and α are respectively the initial value

of gravitational constant and a constant, *iter* is the current iteration and *maxiter* is the maximum number of iterations.

For each individual, the mass is calculated using Eqs. (11)–(13).

$$M_{ai} = M_{pi} = M_{ii} = M_i, \tag{11}$$

$$m_i = \frac{fit_i - worst}{best - worst}, \tag{12}$$

$$M_i = \frac{m_i}{\sum_{i=1}^{n} m_i} \tag{13}$$

where M_a is the active gravitational mass, M_p is the passive gravitational mass, M_i is the inertia mass, and M is the mass of i^{th} individual. Whereas *fit* is the fitness value of i^{th} individual, *worst* is the current worst fitness and *best* is the current best fitness in the current iteration.

Then, each individual acceleration is calculated using Eq. (14).

$$ac_i = \frac{F_i}{M_i} \tag{14}$$

where *ac*, *F*, and *M* are the acceleration, force, and mass of the individuals, respectively. Whereas *i* is the index of the individuals

2.2.5. Update position phase

In this phase, the individual positions based on the velocity of the individuals are updated using Eq. (15). Then, the position of the individuals is updated, as shown in Eq. (16). The updating process is repeated until the criteria are met.

$$V_i(t+1) = V_i(t) + c1\ x rand x ac_i(t) + c2\ x rand x (gbest - x_i(t)) \tag{15}$$

where $V_i(t)$ is the velocity of individual *i* at iteration *t*, *w* is a weighting function, *rand* is a random number between 0 and 1, $ac_i(t)$ is the acceleration of individual *i* at iteration *t*, and *gbest* is the best solution so far. The symbols *c1* and *c2* are weighting factors. The values of *i* and *t* are the index of individuals and iterations respectively, whereas $x_i(t)$ is the position of individual *i* at iteration *t*.

$$x_i(t+1) = x_i(t) + V_i(t+1) \tag{16}$$

where $x_i(t)$ represents the position of an individual *i* at iteration *t*, and $V_i(t+1)$ represents the velocity of the individuals that were derived from Eq. (15).

2.3. Model and experimental setup

We applied our estimation algorithm on two aspartate biochemical pathways (Christophe Chassagnole et al., 2002; Curien and Bastien, 2009). This section explains both the estimation and the simulation setups.

In this study, Copasi (Schaber, 2012) and SBtoolbox (Schmidt, 2007) for MATLAB were used as the main software. The aspartate biochemical pathways were retrieved from an online database known as the BioModel. The proposed algorithm for parameter estimation in biological models was evaluated using the aspartate biosynthesis pathway of the *A.thaliana* model (Curien and Bastien, 2009) and threonine biosynthesis pathway of the *E. coli* model (Christophe Chassagnole et al., 2002). The target metabolites, the Isoleucine (Ile) and homoserine phosphate (HSP) metabolite, were focused on these pathways. Copasi software was used to analyze the pathways and the kinetic parameter involved in target metabolites. Then, the proposed algorithm was implemented in SBtoolbox as an optimization algorithm in estimating the kinetic parameter values that previously derived. Next, the near-optimal set of kinetic parameter values was estimated for both metabolites. These kinetic parameter values were substituted into the ODE of the metabolites.

The ODE is important as these equations are solved to generate the time series data. Then, evaluating and comparing the performances with the output of time series data among estimation algorithms facilitated the identification of the near-optimal set of parameter values. The estimation results of the kinetic parameter values were based on 30 runs and 100 iterations per run for each of the algorithms to obtain the best set of kinetic parameter values. The values for the control parameters in algorithms used were $np = 10\ x\ D$ (Chong et al., 2014; Ng et al., 2013), $G0 = 1$, $\alpha = 23$ (Mirjalili and Hashim, 2010; Rashedi et al., 2009), $c1 = 0.5$, and $c2 = 1.5$ (Mirjalili and Hashim, 2010). Additionally, *np* was the number of populations, and *D* was the dimension of the problems that described the number of kinetic parameters to be estimated. *G0* and α were descending coefficients used in acceleration calculations. The values *c1* and *c2* were the weighting factors used in update velocity for IPSGOSA, PSOGSA, and PSO. For the initial guess, the assigned values were based on existing literature data (Chong et al., 2014; Ng et al., 2013).

In the aspartate metabolism of *A.thaliana* pathway model (Curien and Bastien, 2009), the target metabolite focused is Isoleucine (Ile). Ile is involved in three reactions. First, it acts as a reactant in Ile aminoacyl-tRNA synthetase enzyme reaction (*vileTRNA*); second, Ile acts as a modifier in threonine deaminase; and the third reaction is the product of the threonine deaminase enzyme reaction (*vtd*). On top of that, there is also a parameter involved in the reaction of *vileTRNA*, namely *V_Ile_RS*, with a value of 0.43. In this model, the value of *V_Ile_RS* is based on *V_Aa_RS* (global parameter) that is involved in the aminoacyl-tRNA synthetase reaction for Isoleucine, Lysine, and Threonine in the model (Curien and Bastien, 2009). Eqs. (17)–(19) show the ODE, which describe the changes in the concentration of isoleucine over time in the model.

$$\frac{d[Ile]}{dt} = vtd - vileTRNA \tag{17}$$

where

$$vtd = c1 \times TD \times Thr \times \frac{TD\_k\_app\_exp}{1 + \left(\frac{Ile}{TD\_Ile\_Ki\_no\_Val\_app\_exp + \frac{Vtd\_TD\_Val\_Ka1\_app\_exp \times Val}{Vtd\_TD\_Val\_Ka2\_app\_exp \times Val}} \right)^{TD\_k\_app\_exp}} \tag{18}$$

$$vileTRNA = \frac{c1 \times V\_Ile\_RS \times Ile}{Ile\_tRNA\_Ile\_Km + Ile} \tag{19}$$

*c*1=1

Ile = Concentration of isoleucine
V_Ile_RS = 0.43 (global parameter)
TD = Concentration of threonine deaminase
Thr = Concentration of threonine
Val =Concentration of valine

For the aspartate metabolism of *E. coli* pathways model (Chassagnole et al., 2001), the target metabolite focused on HSP. There two reactions are derived from the involvement of the HSP metabolite. Both operate as a reactant in threonine synthase (*vtsy*), and a product in homoserine kinase enzyme reaction (*vhk*). Eqs. (20)–(22) highlight how the ODE describes the changes in the concentration of HSP over time in the model. Table 1 shows the list of kinetic parameters to be estimated for Isoleucine and HSP metabolites.

$$\frac{d[HSP]}{dt} = vtsy - vhk \tag{20}$$

where

$$vtsy = \frac{compartment \times vm5 \times hsp}{hsp + k5hsp} \tag{21}$$

Table 1
Kinetic parameters to be estimated.

Metabolite	Isoleucine	HSP
List of kinetic parameter involved	Vtd_TD_k_app_exp	Vtsy_vm5
	Vtd_TD_Ile_Ki_no_Val_app_exp	Vtsy_k5hsp
	Vtd_TD_Val_Ka1_app_exp	Vhk_vm4f
	Vtd_TD_Val_Ka2_app_exp	Vhk_lys
	Vtd_TD_nH_app_exp	Vhk_k4lys
	VileTRNA_Ile_tRNAS_Ile_Km	Vhk_k4atp
		Vhk_k4ihs
		Vhk_k4hs
		Vhk_k4thr
		Vhk_k4iatp

$$vhk = \frac{vm4f \times hs \times atp}{(1+\frac{lys}{k4lys}) \times (atp + k4atp \times (1+\frac{hs}{k4ihs})) \times (hs + k4hs \times (1+\frac{thr}{k4thr}) \times (1+\frac{atp}{k4iatp}))} \quad (22)$$

compartment = 1

hsp = Concentration of homoserine phosphate

lys = Concentration of lysine

atp = Concentration of ATP

hs = Concentration of homoserine

thr = Concentration of threonine

Once the set of kinetic parameter values are estimated, the time series data is generated based on the solving ODE equation for each metabolite. These generated data are used as the performance of the estimated result for each algorithm. The performance of each estimation result is evaluated using sample standard deviation and root mean squared error (RMSE). The estimation results are based on the fitted model with the experimental data, where the STD and RMSE values are calculated according to Eqs. (23)–(26).

$$\text{Error rate}, \quad e = \sum_{i=1}^{n} \left(y_i^{exp} - y_i\right)^2 \quad (23)$$

$$\text{Mean} = \quad A = \frac{\sum_{i=1}^{n} e}{n} \quad (24)$$

$$STD = \frac{\sum_{i=1}^{n} \left((y_i^{exp} - y_i)^2 - \mathrm{A}\right)^2}{n-1} \quad (25)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_i^{exp} - y_i\right)^2} \quad (26)$$

where n is the number of rows of experimental time series data values, i is the index variable, and y_i^{exp} and y_i are experimental and estimated time series data respectively.

In addition, the evaluation is also done by comparing the predictive ability of each estimation result. In assessing the predictive ability of the estimated model results, Prediction Sum of Squares (PRESS) evaluation is used to cross-validate each of the model estimated (Bartoli, 2009; Tarpey, 2000 Tarpey, 2000). PRESS is a measurement based on the leave one out technique (Allen, 1974). The PRESS value is calculated based on Eq. (27).

$$PRESS = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}}\right)^2 \quad (27)$$

where n is the number of rows of time series values, i is the index variable, and y_i and $\hat{y}_i$ are observation and estimated time series (model) data for point i respectively. Whereas h_{ii} is the diagonal element of the hat matrix.

3. Results and discussion

Once the estimated kinetic parameter values were collected, as shown in Tables 2 and 4, they were substituted into the ODE in the pathway models as explained in the previous section. Subsequently, the model outputs that can be generated will be further evaluated for the performance of each estimation result, as shown in Tables 3 and 5.

Tables 2 and 4 show the kinetic parameters for both Ile and HSP metabolites, which were estimated by IPSOGSA, PSOGSA, PSO, and GSA. The estimated kinetic parameters that were determined by PSO for Ile were obtained from prior parameter estimation studies that employed the same model (Ng et al., 2013). Whereas, the estimated kinetic parameters for HSP by Improved Bee Memory Differential Evolution (IBMDE) were also obtained from prior studies that implemented the same model (Chong et al., 2013). The performance evaluation that used STD, RMSE, and PRESS for each algorithm is shown in Table 3, and the results revealed that IPSOGSA outperformed the other algorithms, obtaining the lowest STD value of 0.072117863. Thus, this result indicates the extent by which the IPSOGSA performs consistently compared to the other algorithms for Ile. For the RMSE results, IPSOGSA again managed to obtain a good result compared to the other algorithms by scoring the lowest RMSE, 12.2125, followed by PSOGSA, PSO, and GSA, which scored 15.5251, 16.6551, and 67.48635144 respectively. Hence, this result highlights the closeness of the error between the generated values with the noisy experimental values (model fitting), with IPSOGSA obtaining the lowest value of RMSE for the Ile metabolite. Furthermore, the cross validation on the predictive ability for Ile also shows that IPSOGSA is able to obtain a good score of 2.0375E+03 for PRESS value, with IPSOGSA score being the best PRESS value among algorithms in this study.

The results presented in Table 5 show that IBMDE is the score of the smallest STD by scoring 2.37871E-05 compared to other algorithms for HSP. In contrast, IPSOGSA manages to obtain a good RMSE score by obtaining the lowest value of 0.030397571 followed by PSOGSA, GSA, PSO, and IBMDE with the scores of 0.031181553, 0.032663132, 0.0350988, and 0.053665631 respectively. Hence, this provides the measurement of the error of the model fitting between estimated time series data with the experimental data. For the predictive ability of the estimated results, PRESS score for IBMDE is the best score with 0.0121 and IPSOGSA only score 0.0294 for HSP estimated results.

Table 2
Kinetic parameters estimated for isoleucine.

Algorithm/Kinetic Parameter	IPSOGSA	PSOGSA	PSO (Ng et al., 2013)	GSA
Vtd_TD_k_app_exp	0.017135	0.015099	0.0123	0.016484
Vtd_TD_Ile_Ki_no_Val_app_exp	18.744	19.277	75.5376	269.83
Vtd_TD_Val_Ka1_app_exp	55.662	85.764	460.8398	432.67
Vtd_TD_Val_Ka2_app_exp	595.02	626.28	352.7619	666.19
Vtd_TD_nH_app_exp	0.96763	10.783	11.0296	10.913
VileTRNA_Ile_tRNAS_Ile_Km	11.099	17.271	19.998	16.764

Table 3
Performance of estimation for isoleucine.

Evaluation	IPSOGSA	PSOGSA	PSO	GSA
STD	0.072117863	0.199461611	0.244264	0.48292
RMSE	12.2125	15.5251	16.6551	67.48635144
PRESS	2.0375E+03	2.8701E+03	4.3004E+03	9.5218E+03

Note: Shaded cells represent the best result.

Table 4
Kinetic parameter estimated for HSP.

Algorithm/Kinetic Parameter	IPSOGSA	PSOGSA	PSO	GSA	IBMDE (Chong et al., 2013)
vtsy_vm5	0.0125	0.0206	0.0639	0.0121	0.181
vtsy_k5hsp	0.031	0.1601	0.8931	0.031	2.183
vhk_vm4f	0.3826	0.1402	0.2945	1	62.174
vhk_lys	2.5988	3.6474	4.6	4.2613	1.750
vhk_k4lys	17.4012	19.2087	56.5455	45.0903	109.980
vhk_k4atp	0.2239	0.3403	0.0118	0.238	0.110
vhk_k4ihs	28.6036	27.9923	14.3461	2.5428	2.327
vhk_k4hs	0.011	0.094	0.91	0.5468	51.068
vhk_k4thr	6.3562	0.9907	2.0398	10.8505	4.164
vhk_k4iatp	12.2367	15.5189	41.0991	30.4361	248.351

Table 5
Performance of estimation for HSP.

Evaluation	IPSOGSA	PSOGSA	PSO	GSA	IBMDE
STD	4.04684E-06	4.16389E-06	4.16656E-06	4.9379E-06	2.37871E-05
RMSE	0.030397571	0.031181553	0.0350988	0.032663132	0.053665631
PRESS	0.0294	0.0289	0.0282	0.0291	0.0121

Note: Shaded cells represent the best result.

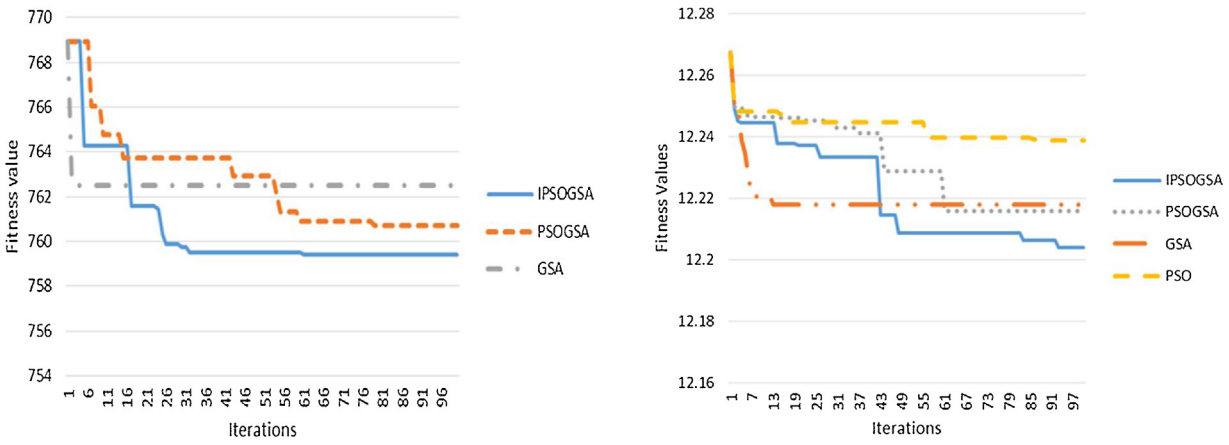

Fig. 4. The convergence graphs by estimation algorithms (Note: The left panel shows a convergence graph for the Ile metabolite whereas the right panel shows a convergence graph for the HSP metabolite).

Fig. 4 shows the convergence graph for both metabolites that were plotted based on the fitness values generated by the IPSOGSA, PSOGSA, GSA, and PSO estimation algorithms. For both Ile and HSP metabolites, the results show that the convergence speed of IPSOGSA is faster, and estimated a better solution than other algorithms, especially for PSO and GSA for both pathway cases. The convergence graph for HSP estimation shows that PSO and GSA are unable to improve their solution after some time, particularly around sixty iterations. Regardless, IPSOGSA and PSOGSA still can improve the solution in the same iteration. These demonstrate the factor that both PSO and GSA are stuck at one of the local optimal in finding the global solution. Moreover, this result is also mainly due to the fact that the IPSOGSA and standard PSOGSA can avoid the local optimal problem (Mirjalili and Hashim, 2010). The ability to avoid local optimal especially IPSOGSA is assisted by the probability phase, where the chances of obtaining the improved solution are increased. In addition, the result also shows that IPSOGSA managed to identify the better solution than PSOGSA in less time (iterations). This is because IPSOGSA has the ability to exploit the high-possibility individuals. Fig. 5 shows the time-series concentration graph based on the kinetic parameters that were obtained by estimating the results of IPSOGSA. It also shows the dispersion of the simulated time-series concentration on the experimental data. Broadly speaking, the results for both the Ile and HSP metabolite show that IPSOGSA was able to obtain a good result for model fitting especially in terms of the RMSE score. This indicates that IPSOGSA can generate more stable results in the model fitting process than PSOGSA, GSA, and PSO. Although the cross-validation on the predictive ability of the estimated model by IPSOGSA is only good in the Ile metabolite, the overall performance of IPSOGSA outperforms

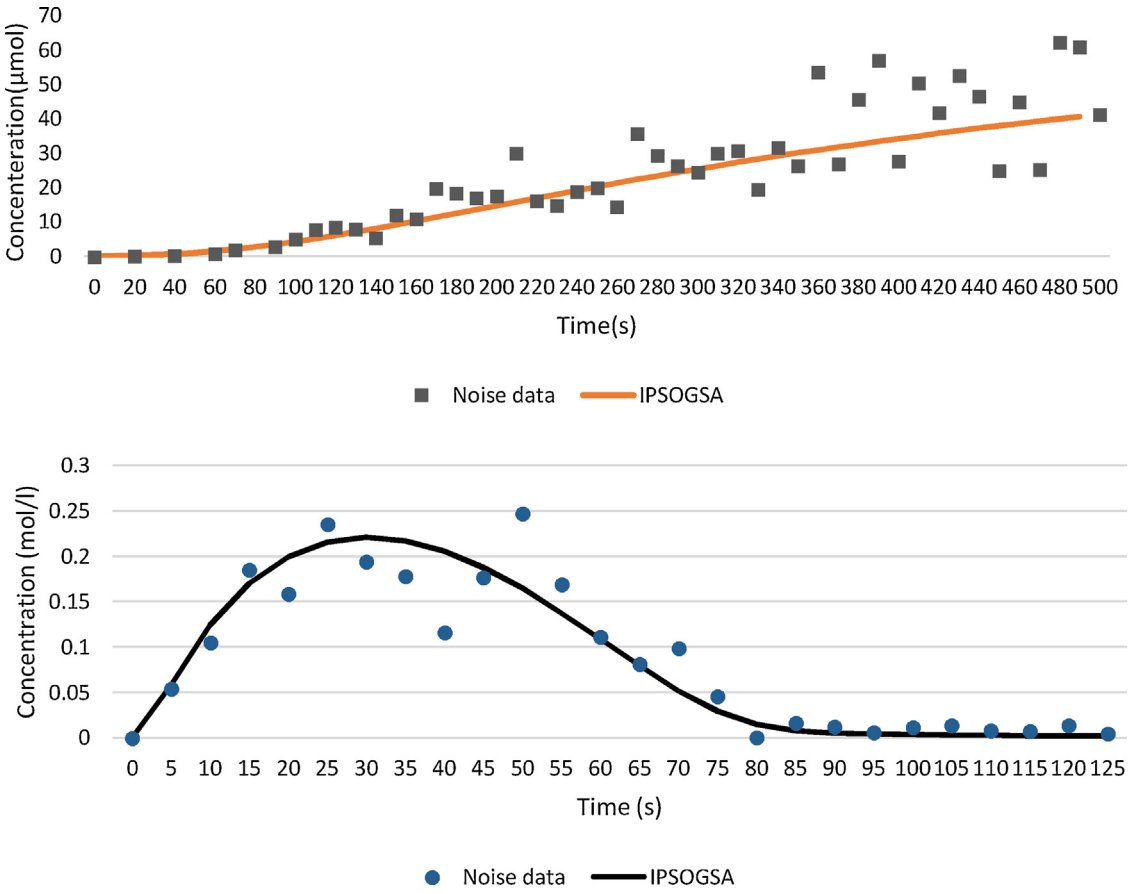

Fig. 5. The time series concentration fitting graph (Note: The upper panel shows the estimated Ile metabolite concentration while the lower panel shows the estimated HSP metabolite concentration graph).

other algorithms in both case studies, especially in the fitted model with experimental data.

Based on the previous discussion, this study shows that IPSOGSA is able to obtain a near-optimal set of kinetic parameter values for both metabolites. The result also indicates that the estimated model is consistent with the noisy data. Meanwhile, the other three algorithms failed to achieve the required consistency, especially PSO and GSA, both of which failed to avoid the local optimal problem and this led to a poor performance in terms of the STD and RMSE scores. Furthermore, the convergence graph results proved that IPSOGSA has the ability to find the feasible solution areas and, therefore, can accelerate the process of finding near-optimal solutions. This feature proved that IPSOGSA is more promising than other standard algorithms.

4. Conclusion

IPSOGSA was employed successfully in this study to reduce the space to obtain near-optimal solutions. This algorithm is driven by its ability to exploit the high-probability feasible solution areas, which can subsequently be employed to obtain the best result. In terms of model fitting accuracy, the estimated results of the IPSOGSA were more closely matched with the noisy data than the PSOGSA, GSA, and PSO algorithms. In addition, the IPSOGSA demonstrated the ability to avoid bad solutions or local optima, and this also contributed to the strong performance of the estimated results. In addition, the analysis of the results shows that IPSOGSA was able to perform well in comparison to the other algorithms, although it was hampered by noise and incomplete experimental data. Moreover, the results of the IPSOGSA indicated that the small RMSE and PRESS value can represent a good estimated result. In contrast, for RMSE, the results estimated by IPSOGSA were more reliable and accurate than those obtained through the application of the other standard algorithms. In the large scale and complex models, the parameter estimation process is computationally expensive and is indeed time-consuming. Thus, the proposed method will not be applicable to more complex models due to time limitation. Henceforth, IPSOGSA still needs to be tested in a larger kinetic model that consists of more complex evolution equations. In conclusion, IPSOGSA is an estimation algorithm that can be highly useful in areas of research that deal with noisy data; for example, in the domains of electronic and electrical engineering. Furthermore, IPSOGSA can also be implemented in other biochemical pathways to obtain the near-optimal parameter values that can reduce the model errors. The majority of the estimation algorithms require the control parameters to be set. Therefore, a better approach, such as enabling these control parameters to self-tune as a means of obtaining more accurate and reliable results, is required to overcome this limitation.

Acknowledgements

We would like to thank the Malaysian Ministry of Higher Education for supporting this research through two Fundamental Research Grant Schemes (Grant number: R.J130000.7823.4F615 and R.J130000.7828.4F720). We would also like to thank the editors and reviewers for their helpful suggestions. Lastly, we owe a debt of gratitude to Universiti Teknologi Malaysia for sponsoring this research via the GUP Research Grant (Grant Number: Q.J130000.2528. 11H05).

References

Allen, D.M., 1974. The relationship between variable selection and data augmentation and a method for prediction. Technometrics 16 (1), 125, http://dx.doi.org/10.2307/1267500.

Baker, S.M., Schallau, K., Junker, B.H., 2010. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. J. Integr. Bioinform. 7 (3), 1–9, http://dx.doi.org/10.2390/biecoll-jib-2010-133.

Banga, J.R., 2008. Optimization in computational systems biology. BMC Syst. Biol. 2 (February (2008)), 47, http://dx.doi.org/10.1186/1752-0509-2-47.

Bartoli, A., 2009. On computing the prediction sum of squares statistic in linear least squares problems with multiple parameter or measurement sets. Int. J. Comput. Vis. 85 (2), 133–142, http://dx.doi.org/10.1007/s11263-009-0253.

Chassagnole, C., Fell, D.A., Raïs, B., Kudla, B., Mazat, J.P., 2001. Control of the threonine-synthesis pathway in *Escherichia coli*: a theoretical and experimental approach. Biochem. J. 356 (Pt. 2), 433–444, http://dx.doi.org/10.1042/0264-6021:3560433.

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol. Bioeng. 79 (1), 53–73, http://dx.doi.org/10.1002/bit.10288.

Chong, C.K., Mohamad, M.S., Deris, S., Shamsir, M.S., Choon, Y.W., Chai, L.E., 2013. Threonine biosynthesis pathway simulation using IBMDE with parameter estimation. PAKDD, 191–200, http://dx.doi.org/10.1007/978-3-642-40319-4_17.

Chong, C.K., Mohamad, M.S., Deris, S., Shamsir, M.S., Chai, L.E., Choon, Y.W., 2014. Using an improved bee memory differential evolution algorithm for parameter estimation to simulate biochemical pathways. J. Biol. Syst. 03 (22(01)), 101–121, http://dx.doi.org/10.1142/s0218339014500065.

Curien, G., Bastien, O., 2009. Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. Mol. Syst. 5 (271), 271, http://dx.doi.org/10.1038/msb.2009.29.

Fernández Slezak, D., Suárez, C., Cecchi, G.A., Marshall, G., Stolovitzky, G., 2010. When the optimal is not the best: parameter estimation in complex biological models. PLoS One 5 (10), http://dx.doi.org/10.1371/journal.pone.0013283.

Fong, S.S., 2014. Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. Comput. Struct. Biotechnol. J. 11 (18), 28–34, http://dx.doi.org/10.1016/j.csbj.2014.08.005 (Mini Review).

Galazzo, J.L., Bailey, J.E., 1990. Fermentation pathway kinetics and metabolic flux control in suspended and immobilized Saccharomyces cerevisiae. Enzyme Microb. Technol. 12 (3), 162–172, http://dx.doi.org/10.1016/0141-0229(90)90033-M.

Ismail, M.A., Deris, S., Mohamad, M.S., Abdullah, A., 2015. A newton cooperative genetic algorithm method for In Silico optimization of metabolic pathway production. PLoS One 10 (5), 1–19, http://dx.doi.org/10.1371/journal.pone.0126199.

Karaboga, D., Akay, B., 2009. A comparative study of artificial bee colony algorithm. Appl. Math. Comput. 214 (1), 108–132, http://dx.doi.org/10.1016/j.amc.2009.03.090.

Kumar, Y., Sahoo, G., 2014. A review on gravitational search algorithm and its applications to data clustering & classification. Int. J. Intell. Syst. Appl. 6 (6), 79–93, http://dx.doi.org/10.5815/ijisa.2014.06.09.

Leong, Y.Y., Chong, C.K., Choon, Y.W., En, L., Deris, S., Illias, R.M., . . . Illias, R.M., . . . Omatu, S., . . . Mohamad, M.S., 2013. Simulation of fermentation pathway using bees algorithm. Adv. Distrib. Comput. Artif. Intell. J. 1 (2), 13–19, http://dx.doi.org/10.14201/ADCAIJ2012121319.

Lillacci, G., Khammash, M., 2010. Parameter estimation and model selection in computational biology. PLoS Comput. Biol. 6 (3), http://dx.doi.org/10.1371/journal.pcbi.1000696.

Mirjalili, S., Hashim, S.Z.M., 2010. A new hybrid PSOGSA algorithm for function optimization. Proceedings of ICCIA 2010 – 2010 International Conference on Computer and Information Application 1, 374–377, http://dx.doi.org/10.1109/iccia.2010.6141614.

Ng, S.T., Chong, C.K., Choon, Y.W., Chai, L.E., Deris, S., Illias, R.M., Shamsir, M.S., Mohamad, M.S., 2013. Estimating kinetic parameters for essential amino acid production in arabidopsis thaliana by using particle swarm optimization. Jurnal Teknologi 64 (1), 73–80, http://dx.doi.org/10.11113/jt.v64.1737.

Pham, D.T., Ghanbarzadeh, A., Koç, E., Otri, S., Rahim, S., Zaidi, M., 2006. The bees algorithm—a novel tool for complex optimisation problems. In: Intelligent Production Machines and Systems. Elsevier, pp. 454–459, http://dx.doi.org/10.1016/B978-008045157-2/50081-X.

Rashedi, E., Nezamabadi-pour, H., Saryazdi, S., 2009. GSA: a gravitational search algorithm. Inform. Sci. 179 (13), 2232–2248, http://dx.doi.org/10.1016/j.ins.2009.03.004.

Raue, A., Steiert, B., Schelker, M., Kreutz, C., Maiwald, T., Hass, H., Vanlier, J., Tonsing, C., Adlung, L., Engesser, R., Mader, W., Heinemann, T., Hasenauer, J., Schilling, M., Hofer, T., Klipp, E., Theis, F., Klingmuller, U., Schoberl, B., Timmer, J., 2015. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. Bioinformatics 31 (21), 3558–3560, http://dx.doi.org/10.1093/bioinformatics/btv405.

Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R., 2006a. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. BMC Bioinf. 7, 483, http://dx.doi.org/10.1186/1471-2105-7-483.

Rodriguez-Fernandez, M., Mendes, P., Banga, J.R., 2006b. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. Biosystems 83 (2–3 SPEC. ISS), 248–265, http://dx.doi.org/10.1016/j.biosystems.2005.06.016.

Schaber, J., 2012. Easy parameter identifiability analysis with COPASI. Biosystems 110 (3), 183–185, http://dx.doi.org/10.1016/j.biosystems.2012.09.003.

Schilling, C., Bogomolov, S., Henzinger, T.A., Podelski, A., Ruess, J., 2016. Adaptive moment closure for parameter inference of biochemical reaction networks. Biosystems, http://dx.doi.org/10.1016/j.biosystems.2016.07.005.

Schmidt, H., 2007. SBaddon: high performance simulation for the systems biology toolbox for MATLAB. Bioinformatics 23 (5), 646–647, http://dx.doi.org/10.1093/bioinformatics/btl668.

Shanhe, J., Zhicheng, J., 2014. An improved HPSO-GSA with adaptive evolution stagnation cycle. In: Proceedings of the 33rd Chinese Control Conference, IEEE, pp. 8601–8606, http://dx.doi.org/10.1109/chicc.2014.6896444.

Shi, Y., Eberhart, R.C., 1999. Empirical study of particle swarm optimization. Proceedings of the 1999 Congress on Evolutionary Computation-CEC99, 1945–1950, http://dx.doi.org/10.1109/CEC.1999.785511.

Sun, J., Garibaldi, J.M., Hodgman, C., 2012. Parameter estimation using metaheuristics in systems biology: a comprehensive review. IEEE/ACM Trans. Comput. Biol. Bioinform. 9 (1), 185–202, http://dx.doi.org/10.1109/TCBB.2011.63.

Tarpey, T., 2000. A note on the prediction sum of squares statistic for restricted least squares. Am. Stat. 54 (2), 116–118, http://dx.doi.org/10.1080/00031305.2000.10474522.

Tashkova, K., Korošec, P., Šilc, J., Todorovski, L., Džeroski, S., 2011. Parameter estimation with bio-inspired meta-heuristic optimization: modeling the dynamics of endocytosis. BMC Syst. Biol. 5 (1), 159, http://dx.doi.org/10.1186/1752-0509-5-159.

Villaverde, A.F., Egea, J.A., Banga, J.R., 2012. A cooperative strategy for parameter estimation in large scale systems biology models. BMC Syst. Biol. 6, 75, http://dx.doi.org/10.1186/1752-0509-6-75.

Villaverde, A.F., Henriques, D., Smallbone, K., Bongard, S., Schmid, J., Cicin-Sain, D., Crombach, A., Saez-Rodriguez, J., Mauch, K., Balsa-Canto, E., Mendes, P., Jaeger, J., Banga, J.R., 2015. BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. BMC Syst. Biol. 9 (1), 8, http://dx.doi.org/10.1186/s12918-015-0144-4.

Yang, X.-S., 2009. Firefly algorithms for multimodal optimization. Proceedings of the 5th International Conference on Stochastic Algorithms: Foundations and Applications, 169–178, http://dx.doi.org/10.1007/978-3-642-04944-6_14.