International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 5 (2013) pp. 571-580

A New Particle Swarm Evolutionary Optimization for Parameter Estimation of Biological Models

Afnizanfaizal Abdullah[1], Safaai Deris[1], Mohd Saberi Mohamad[1] and Siti Zaiton Mohd Hashim[2]

[1] Artificial Intelligence and Bioinformatics Group (AIBIG),
Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia,
81310 UTM, Johor Bahru, Johor, Malaysia
{afnizanfaizal, safaai, saberi}@utm.my

Sổft Computing Research Group (SCRG),
Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia,
81310 UTM, Johor Bahru, Johor, Malaysia
sitizaiton@utm.my

***Abstract*: The development of reliable biological models has become an important issue in systems biology. These models are constructed using differential algebraic equations to represent the dynamic perturbation of the biochemical quantities within the cells. However, these models heavily depend on the set of parameters that signify the physiology of the systems such as reaction rates and kinetic constants. These parameters are commonly difficult to be obtained using the experimental measurements. Due to the uncertainty of the measurements and the nonlinearity of the systems, advanced optimization methods are often necessary. In this paper, a new hybrid optimization method is introduced. The method, called Particle Swarm Evolutionary Optimization (PSEO), is proposed based on the combination of Particle Swarm Optimization (PSO) and Differential Evolution (DE) methods. The effectiveness of the proposed PSEO method on the parameter estimation problem is evaluated using two biological models, namely synthetic oscillator and microbial lactose operon models. The experimental results showed that the performances in term of better fitness value and computational speed of the proposed method have outperformed those produced by the existing methods like Particle Swarm Optimization (PSO), Differential Evolution (DE) and recently proposed hybrid Local Evolutionary PSO (LEPSO) methods.**

***Keywords*:** Hybrid optimization method, Particle Swarm Optimization, Differential Evolution, parameter estimation, biological models**.**

I. Introduction

Computational modeling plays an important role in the understanding of systems biology [1]. In this study, models are used to represent the dynamic perturbation of the biochemical reactions within living cells [2]. The models are usually constructed using a set of coupled differential algebraic equations, mostly by using ordinary differential equations (ODEs), to signify the reactions over specific range of observed time steps. To ensure the accuracy of the prediction by using the model outputs, the models heavily relied on a set of parameters that characterize the physiology behaviors of systems, such as reaction rates and kinetic constants. However, these parameters are generally difficult to be extracted from experimental analyses [1-3]. Hence, the parameters are rather approximated based on the given experimental measurements. In most cases, the nonlinear least squares techniques are used to find optimal parameters that may produce model outputs which fit closely to the corresponding experiment measurements. This task is usually hampered by the nonlinearity of the systems as well as the incompleteness of the available experimental measurements [4-5, 32].

The estimation of the parameters in the biological model is often formulated as an optimization problem. Formally, the objective function of the problem is usually intended to minimize the difference between the model outputs produced by the estimated parameters and the respective experimental measurements. As a result, the plausible solution, which is formed from the combination of parameter sets, may generate the model outputs that closely fit the experimental measurements. In the past few years, a number of previous methods have been proposed to solve this particular problem using maximum likelihood fitting [6], Bayesian estimation [7], and local optimization [8]. These methods frequently utilize derivative-based approach, mostly by transforming the problem into a linear problem and iteratively find the optimal parameters based on certain constraints. Among these methods, the improved methods that are based on the Extended Kalman Filter (EKF) method have been extensively exploited [9]. In general, the methods employ local-based optimization approach by using a set of initial values to estimate the parameters according to the state approximation techniques. Lillacci and Khammash [10] introduced an improved EFK method that incorporates continuous model outputs with discrete experimental measurements. On the other hand, Zheng and co-workers [11] proposed a new hybrid EKF method using switching Particle Swarm Optimization (PSO) [12] method to estimate these parameters. However, despite of the advantages in handling the noisy and incomplete experimental measurements, these methods frequently depend on further estimation refinements to avoid

the methods from being trapped into the local suboptimal solutions.

Otherwise, the global optimization methods have shown potential achievements, especially in term of robustness on the initial values. The methods basically use stochastic population-based searching strategy to find the plausible solutions. Firstly, a set of randomly formed solutions is initialized. Then, these solutions are subjected for recombination operations to update the fitness values. These steps are repeated until a maximum number of iterations is reached or the specific fitness value is met. Rodriguez-Fernandez et al. [2] proposed a new method based on the Scatter Search [13] algorithm to estimate the parameters of several benchmark models. Recently, the evolutionary-based optimization methods have received an increasing attention for the problem solution as these methods are capable to handle measurement noise more effectively [14]. A real coded- Genetic Algorithm (RCGA) method is introduced and the performances in term of fitness evaluation and convergence speed for the parameter estimation problem are presented [15]. Conversely, Differential Evolution (DE) methods [16] have been broadly used. Nevertheless, these methods majorly lack of the use of computational cost due to the fact that these methods normally utilize a substantial amount of computational times to converge to better fitness values [17]. Moreover, there is no guarantee that these methods are capable to converge to the global best solutions [18, 30, 32].

Therefore, hybrid optimization methods are introduced to alleviate these challenges, in which several methods are combined so that the advantages of one method can be used to improve the limitations of the others [19-20]. Previously, Rodriguez-Fernandez et al. [21], Ashyraliyev et al. [22], and Fomekong-Nanfack et al. [23] had introduced hybrid optimization methods based on Stochastic Ranking Evolutionary Strategy (SRES) [24] to improve the local optimization methods. More recently, improved PSO methods, with the combination of local simplex search method and statistical inference [25], has been proposed to robustly estimate the parameters with the availability of a significant level of measurement noise. In addition, a cooperative based strategy is introduced to the standard DE method to control the behavior of evolutionary operations, thus utilizing the computational time more efficiently [19]. Abdullah et al. [3] introduced a new hybrid optimization method that is based on the Clonal Selection Algorithm (CSA) and DE methods to handle both measurement noise and incompleteness problems in the experimental measurements. This method is coupled with the statistical analysis to validate the consistency of the model outputs compared to the experimental measurements.

In this paper, a new hybrid optimization method is introduced to estimate the parameters in the complex and nonlinear biological models. The proposed method, named Particle Swarm Evolutionary Optimization (PSEO), is proposed based on the combination of Particle Swarm Optimization (PSO) and Differential Evolution (DE) methods. Generally, the method employs the recombination strategy adopted from the DE method to enhance the neighboring searching strategy by the PSO method. In addition, selective technique is also introduced to discriminate the population of the solutions based on the fitness values. Thus, the approach is aimed to reduce the computational times substantially as only a number of solutions with potential fitness values are selected for the recombination process. On the other hand, the solutions with least substantial fitness values are subjected for separate recombination according to randomly chosen potential solutions so that the method is capable to escape the local suboptimal solutions more effectively. The effectiveness of the proposed method is evaluated using two biological models: synthetic oscillator and microbial lactose operon models. The synthetic oscillator model [26] is proposed based on the cell-free concept which neglects the necessity of experimental measurements. Instead, the microbial lactose operon model [27] contains a large number of parameters and involves noisy and incomplete experimental measurements. The performances of the methods in term of finding better fitness values and the use of computational cost are compared with those generated by the existing PSO, DE and hybrid Local Evolutionary PSO (LEPSO) [19] methods. In addition, statistical analysis is also employed to evaluate the reliability of the model outputs produced by using the estimated parameters over the given experimental measurements.

The outline of the paper is presented as follow. The Method section explains the problem formulation, standard PSO, DE, and the proposed PSEO methods. Furthermore, statistical analysis, used to evaluate the reliability of the model outputs produced by the estimated parameters, is also presented. Then, the Experimental Results section describes the results of the performances in term of convergence behaviors and computational cost from the proposed HCSR method. These performances are later compared with the performances of the existing PSO, DE, and hybrid LEPSO methods. Next, the contribution of this work is discussed in the Discussion section.

II. Methods

A. Problem Formulation

The parameter estimation of biological models can be formulated as a nonlinear optimization problem. Suppose the concentration of a component in the model is given as $m(x, u, t)$, in which $x = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_p\}$ is the set of p model parameters, u is the input signal, and τ is the observed time step. Hence, the reaction rate of the involved concentration is given as follows

$$\frac{d[m]}{dt} = h(m(x, u, \tau)) \tag{1}$$

where $h(\cdot)$ is the nonlinear function. The concentration is used to simulate the output, y, as follows

$$y = g(m(x, u, \tau)) + \varepsilon \tag{2}$$

where $g(\cdot)$ is another nonlinear function and ε is the measurement noise superimposed in the model to represent the actual noise exhibited in the actual experimental data, y^{exp}. In general, the parameter estimation problem is aimed to find the optimal parameter set that minimizes the difference between the model output and its corresponding experimental

data, represented as the following equation

$$f(x_{opt}) = \min \arg \sum_{m=1}^{M}\sum_{n=1}^{N}\left(y_{mn} - y_{mn}^{exp}\right)^2 \quad (3)$$

where $x_{opt} = \{\theta_{opt1}, \theta_{opt2}, \theta_{opt3}, \ldots, \theta_{optp}\}$ is the solution provided by the optimization methods that contains a set of p optimal parameters, and M and N are the total numbers of involved concentrations in the model and samples observed, respectively.

B. Particle Swarm Optimization

Particle swarm optimization (PSO) is a stochastic population-based optimization method [12]. The main idea of this method is generally inspired based on the natural behavior of animal foraging activity. PSO promotes the movement of possible solutions, called particles, around the search space. Each particle travels in a specific dimensional space based on the previous experiences of finding the local best solutions, by itself or its neighbors. The movement of a particle, i, is based on its velocity, v_i, and position, x_i, derived as followings [12]:

$$v_i = wv_i + C_aR_1(x_{lbest} - x_i) + C_bR_2(x_{gbest} - x_i) \quad (4)$$

$$x_i = x_i + v_i \quad (5)$$

where w is weight constant, C_a is exploitation coefficient, C_b is exploration coefficient, R_1 and R_2 are random numbers between 0 and 1, while x_{lbest} and x_{gbest} are current local best particle and current global best particle at iteration t, respectively. The termination criteria are satisfied when either a maximum number of iteration is reached or an acceptable fitness value is met. The current global best position is taken as the best solution for the particular run.

C. Differential Evolution

The Differential Evolution (DE) algorithm is also a stochastic population-based optimization method [16], developed based on evolutionary operation similar to Genetic Algorithm (GA). However, in DE, mutation is performed to create a trivial population and this population is crossover with its original counterpart in order to produce offspring population. DE starts with a population of d dimension search vectors, or called chromosomes. Thus, the ith chromosome, X, of the whole population at specific generation t is given by the following form:

$$X_{i(t)} = \{x_{i1(t)}, x_{i2(t)}, \ldots, x_{id(t)}\} \quad (6)$$

where d is the total number of dimension. For each generation, a range of search space has to be set for finding good solution. Thus, at initial generation or t equal to 0, each chromosome is initialized, with lower and upper bound, x_j^L and x_j^U respectively, for jth dimension, as the following equation:

$$x_{ij(t=0)} = x_j^L + R \cdot (x_j^U - x_j^L) \quad (7)$$

where R is a random number generated between 0 and 1.

In DE, mutation operation is used to create a trivial chromosome, v_{ij}. The mutation operation is based on the differentiation of neighborhood chromosome and is executed as the following:

$$v_{ij(t+1)} = x_{best(t)} + F \cdot (x_{r1(t)} - x_{r2(t)}) \quad (8)$$

where $x_{best(t)}$ denotes the current best chromosome, F is scaling factor, $x_{r1(t)}$ and $x_{r2(t)}$ are randomly chosen chromosomes. Using this trivial chromosome, a new offspring chromosome, y_{ij}, is created by performing a crossover operation between the trivial and parent chromosomes. The crossover operation is accomplished using the following rule:

$$y_{ij(t)} = \begin{cases} v_{ij(t)} \; if \; R < CR \\ x_{ij(t)} \; Otherwise \end{cases} \quad (9)$$

where CR is crossover constant and R is a random number between 0 and 1. Hence, the number of offspring chromosomes is the same with the number of parent chromosomes. In order to keep population number constant, a simple selection is executed to decide which chromosome will survive in the next generation. The selection is performed based on the calculated fitness value of each chromosome as the following rule:

$$X_{i(t+1)} = \begin{cases} Y_{i(t)} \; if \; f(Y_{i(t)}) \leq f(X_{i(t)}) \\ X_{i(t)} \; if \; f(Y_{i(t)}) > f(X_{i(t)}) \end{cases} \quad (10)$$

This rule implies that if the offspring chromosome produces a better fitness value, the current parent chromosome will be replaced. Otherwise, the parent chromosome will remain in the population for the next generation.

D. Local Evolution Particle Swarm Optimization

In this section, the local evolutionary PSO (LEPSO) method [19] is described. The aim of this work is to improve local best particle searching. The evolutionary operations of DE are employed into local particle searching in PSO so that the neighborhood solutions of the particles can be utilized efficiently. In addition, greedy selection performed by DE is used to find better solutions from the existing particles. Thus, gradient-based searching can be performed within the population of local best particles. The outline of the proposed method is illustrated in Figure 1. Suppose the searching space is a finite d dimensional and p particles in a population. The ith particle denotes a vector X_i, giving a fitness function as the following:

$$f(X_i) = f(x_{i1}, x_{i2}, \ldots, x_{id}) \quad (11)$$

In this paper, minimization problem is considered. Thus, the global optimum is given by lowest fitness value produced by the global best particle. In the proposed method, DE evolutionary operations are used to improve the current local best particles. For ith particle, the current local best vector, X_{li}, is denoted as:

$$X_{li} = \{x_{li1}, x_{li2}, \ldots, x_{lid}\} \quad (12)$$

```
1: Begin
2: Initiate population, Xi = {X1,X2,X3, ...,XN}
3: Evaluate fitness of each particle, f(Xi)={x1,x2,x3,...,xp}
4: For t < maximum number of generations {
6:    For i = 1 to N number of particles {
8:       Xlbest = find_local_best(Xi); //find local best
9:       Vi = mutate(Xlbest); //perform mutation
10:      Yi =crossover(Vi, Xlbest); //perform crossover
11:      Evaluate f(Yi)={y1,y2,y3, ...,yp}
12:      if f(Xlbest) ≥ f(Yi){
14:         Xlbest = replace(Xlbest, Yi); //replace Xlbest with Yi
15:      } // end if
16:   }// end for
17:
18:   Xgbest = find_global_best(Xi); //find global best
20:
21:   For i = 1 to N number of particles {
22:      Update velocity;
23:      Update position;
24:   }// end for
25:   t = t + 1;
26: } // end for
27: End Begin
```

Figure 1. LEPSO Algorithm.

Then, a new particle, V_{li}, is produced by performing mutation operation. The operation is executed using two randomly selected neighbor particles, x_{r1} and x_{r2}, and the current global best particle, x_{gbest}, as follow:

$$v_{li(t+1)} = x_{gbest(t)} + F \cdot (x_{r1(t)} - x_{r2(t)}) \tag{13}$$

$$V_{li} = [v_{li1}, v_{li2}, \ldots, v_{lid}] \tag{14}$$

where F is the scaling factor. After the new particle is generated, the offspring particle, Y_{li}, of ith particle is created, based on the following rule:

$$y_{li(t)} = \begin{cases} v_{li(t)} \; if \; R < CR \\ x_{li(t)} \; Otherwise \end{cases} \tag{15}$$

where CR is a crossover constant and R is a random number between 0 and 1. Conventionally, the values of F and CR are predefined in the initialization step. Finding the suitable values of these parameters is a challenging task especially when the tradeoff between reliability and efficiency is concerned [29]. In this work, values between 0.4 to 0.9 for F and 0.4 to 1.0 for CR are used [17].

In order to keep the population size constant, a selection of better particles among parent and offspring particles is executed. The proposed method implements a greedy selection by comparing the fitness value of every original particle with its corresponding offspring given by the following rule:

$$X_{li(t+1)} = \begin{cases} V_{li(t)} \; if \; f(V_{li(t)}) \leq f(X_{li(t)}) \\ X_{li(t)} \; if \; f(Y_{li(t)}) > f(X_{li(t)}) \end{cases} \tag{16}$$

As a result, particles that produce better fitness value may survive and will be used in the next generation. A set of local best particle is collected based on the best fitness values achieved so far by each particle personally. The best fitness values among these particles are then selected as the current global best. The particle that produces this value is chosen as the current global best particle. The positions of local and global particle are used to update the velocity and position of each particle in the population using Equation 4 and 5, respectively. The whole process is iteratively executed until the maximum number of iterations is reached.

E. Proposed Particle Evolutionary Swarm Optimization

This paper presents a new hybrid optimization method which is based on the PSO and DE methods. The proposed Particle Evolutionary Swarm Optimization (PSEO) method starts with the vector update based on the swarm searching strategy adopted from the standard PSO method. The update is performed as in Equation 4 and 5. Uniquely, the proposed method employs two stages of evolutionary operations adopted from the DE method to enhance the swarm based searching strategy utilized by the PSO method. The first stage involves two important steps: population ranking and evolutionary improvement. The population ranking step starts with the evaluation of the fitness value of each solution in the population. Next, these solutions are sorted based on the fitness values. The sorted population is later separated into two groups, namely potential and weak sub-populations. The potential sub-population consists of the solutions that hold substantial fitness values. For ith solutions in the potential population, the sub-population is given as follows:

$$X_{i(t)}^{potential} = \left\{x_{i1(t)}^{potential}, x_{i2(t)}^{potential}, \ldots, x_{id(t)}^{potential}\right\} \tag{17}$$

On the other hand, the weak sub-population contains the solutions that yield low fitness values. The ith solutions in the weak population is given as follows:

$$X_{i(t)}^{weak} = \left\{x_{i1(t)}^{weak}, x_{i2(t)}^{weak}, \ldots, x_{id(t)}^{weak}\right\} \tag{18}$$

The potential sub-population is used to undergo the next evolutionary improvement step adopted from the DE method. The step involves mutation and crossover operations. The mutation operation is performed by using the following equation:

$$\hat{X}_{i(t+1)}^{potential} = \left\{x_{gbest(t)} + F \cdot (x_{r1(t)}^{potential} - x_{r2(t)}^{potential})\right\} \tag{19}$$

which may produce a new population of offspring solutions, $\hat{X}_{i(t+1)}^{potential}$, that are used for the crossover operation given by the following rule

$$\hat{X}_{i(t+1)}^{potential} = \begin{cases} \hat{x}_{id(t)}^{potential} \; if \; R < CR \\ x_{id(t)}^{potential} \; Otherwise \end{cases} \tag{20}$$

In order to keep the sub-population size constant, a selection is performed to choose which solutions are plausible. The selection operation is executed based on the following rule:

$$X_{i(t+1)}^{potential} = \begin{cases} \hat{X}_{i(t+1)}^{potential} \; if \; f(\hat{X}_{i(t+1)}^{potential}) \leq f(X_{i(t)}^{potential}) \\ X_{i(t)}^{potential} \; if \; f(\hat{X}_{i(t+1)}^{potential}) > f(X_{i(t)}^{potential}) \end{cases} \tag{21}$$

Alternatively, the second stage involves weak solution

enhancement by using random vector update. This is performed to avoid the method from being trapped into the suboptimal solutions. The random vector update is executed using the following equation:

$$X_{i(t+1)}^{weak} = \{x_{id(t)}^{weak} + rand \cdot (x_{r1(t)}^{potential} - x_{r2(t)}^{potential})\} \quad (22)$$

where $rand$ is a random number between 0 and 1, while $x_{r1(t)}^{potential}$ and $x_{r2(t)}^{potential}$ are vectors of two randomly choosen solutions from the potential sub-population. By doing this, the weak solutions are improved using the information gathered from the potential sub-population. After that, the potential and weak sub-population is merged back to produce an improved population. These stages are repeated and the whole procedure is iterated until the maximum number of iterations is reached. The outline of the proposed PSEO method is given in the Figure 2.

```
1: Begin
2: Initiate population, Xi = {X1,X2,X3,...,XN}
3: Evaluate fitness of each particle, f(Xi)={x1,x2,x3,...,xp}
4: For t < maximum number of iterations {
6:    For i = 1 to N number of solutions {
8:       Xi = evaluate_fitness();
9:    }// end for
10:
11:   Xgbest = find_global_best(Xi); //find global best
12:   sort (Xi )    //sort population based on the fitness
14:   X^potential = split(X, N^potential);
15:   X^weak = split(X, N^weak);
16:
17:   For i = 1 to N^potential {
18:         Update velocity;
20:         Update position;
21:         X̂^potential_i(t+1) = mutate(X^potential_i(t))
22:         X̂^potential_i(t+1) = crossover(X̂^potential_i(t+1), X^potential_i(t))
23:         X^potential_i(t+1) = select(X̂^potential_i(t+1), X^potential_i(t))
24:   }// end for
25:
26:   For i = 1 to N^weak {
27:         X^weak_i(t+1) = {x^weak_id(t) + rand • (x^potential_r1(t) − x^potential_r2(t))}
28:   }// end for
29:
30:   X = merge(X^potential, X^weak);
31:   t = t + 1;
32: } // end for
33:
34: End Begin
```

Figure 2. PSEO Algorithm.

III. Results

A. Synthetic Oscillator

In this paper, a synthetic oscillator model is used to evaluate the performance of the proposed method in estimating small numbers of model parameters. The model is proposed by [26] to simulate the nonlinearity of the deoxyribonucleic acid (DNA) regulatory. This model is designed to modularly fit the synthetic circuit, which allows the model to be used without the dependency on the experimental measurements [26]. The model includes the ribonucleic acid (RNA) activation and inhibition, whereas the gene switches include ON switch Sw21 and Sw12. The following equations are constructed to represent the involved reactions in the model [26]:

$$\frac{d[active]}{dt} = k_1[ON\ Sw21] + k_2[ON\ Sw12] - [active] \quad (23)$$

$$\frac{d[inhibit]}{dt} = k_3[ON\ Sw12] - [inhibit] \quad (24)$$

$$\frac{d[ON\ Sw21]}{dt} = \frac{1}{1^{k_4}+[inhibit]^{k_4}} + [ON\ Sw21] \quad (25)$$

$$\frac{d[ON\ Sw12]}{dt} = \frac{[active]^{k_5}}{1^{k_5}+[active]^{k_5}} + [ON\ Sw12] \quad (26)$$

where k_1, k_2, k_3, k_4, and k_5 are the model parameters with the values of 0.57, 1.5, 2.5, 6.5, and 6.5, respectively [26]. Table 1 presents the comparative results on the average fitness values found by using the existing PSO, DE, LEPSO and proposed PSEO methods. For this experiment, each method is executed 100 times independently and the average best fitness value, accompanying the corresponding standard deviation, is recorded. The population size and the maximum number of iterations are 50 and 200, respectively. For PSO, LEPSO, and PSEO methods, the inertia weight is tuned to 0.5, while the self-exploitation and swarm-exploration rates are both set to 3.5. On the other hand, for the DE, LEPSO, and PSEO methods, the mutation rate is set to 0.7.

	PSO	DE	LEPSO	PSEO
Average Best Fitness	1.74×10^{-3}	5.04×10^{-4}	2.20×10^{-5}	1.52×10^{-7}
Standard Deviation	2.12×10^{-3}	7.21×10^{-3}	2.19×10^{-4}	2.11×10^{-7}
Computational Time (s)	115.3	91.1	210.7	79.3

Table 1. Comparison of average best fitness values and computational costs.

According to the table, the average fitness value for the proposed PSEO method is smaller compared to the other methods. This shows that the proposed method is capable to find a solution that yields better fitness values than those found by using the other methods. Furthermore, the standard deviation of the average best fitness value found by using the PSEO method is relatively small. This suggests that the method can consistently find the best fitness values toward different runs. Figure 3 presents the convergence behavior of the method. Based on this figure, it is suggested that the proposed method is also capable to escape the suboptimal solution more effective compared to the other methods. This is important to ensure that the method is reliable in searching better fitness values toward the iterations, thus avoiding the method to find least substantial solutions.

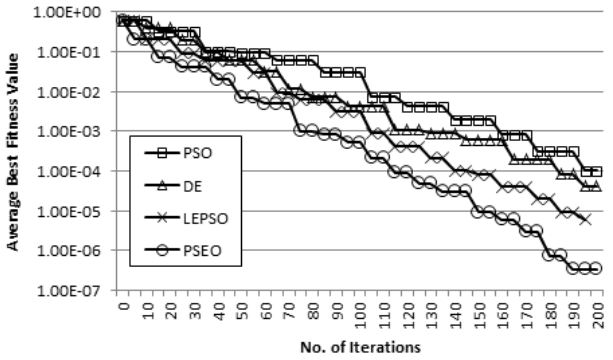

Figure 3. Convergence behavior of the PSO, DE, LEPSO and PSEO methods.

	RNA activation	RNA inhibition	ON switch Sw21	ON switch Sw12
Real Point	1.29×10^{-1}	1.46×10^{-2}	1.12×10^{-1}	9.21×10^{-2}
Error Variance Point	1.29×10^{-1}	1.47×10^{-2}	1.13×10^{-1}	9.23×10^{-2}
Error Variance Intervals	$[1.12\times10^{-1}, 1.45\times10^{-1}]$	$[1.28\times10^{-2}, 1.69\times10^{-2}]$	$[9.90\times10^{-2}, 1.31\times10^{-1}]$	$[8.08\times10^{-2}, 1.06\times10^{-1}]$
χ^2 test	**Valid**			

Table 2. Statistical validation of the model outputs and the experimental measurements.

To demonstrate the effectiveness of the parameters estimated by the proposed method, the model outputs produced by the estimated parameters are compared with the incomplete and noisy experimental measurements and the original model. Table 2 presents the statistical analysis computed using error variance point and intervals based on the model output produced by the estimated parameters and the corresponding experimental measurements [3, 10]. For this validation, a significant level of 95% is used [3]. The results show that the error variance point of each concentration is close to the real point and most importantly, these points lie within the intervals. This provides an evidence that the model outputs produced by the estimated parameters are valid based on the experimental measurements with a significant level of 95%. It is clearly shown that the method is capable to estimate parameters that are reliable and consistent to those from the original models even though with the presence of the incomplete and noisy experimental measurements. This supports the evidence that the method is robust to the measurement noise. Figure 4 to 7 illustrate the data fit of the model outputs produced by the reconstructed model using the estimated parameters, the original model, and the respective experimental measurements.

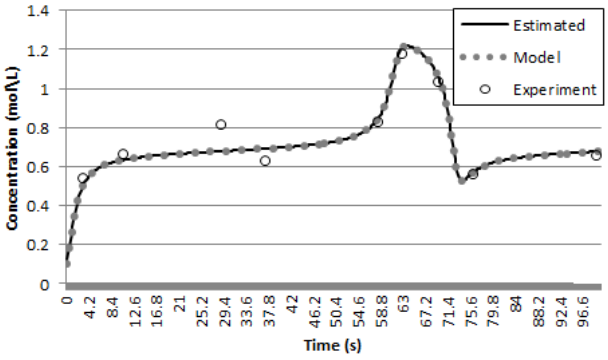

Figure 4. The RNA activation concentration.

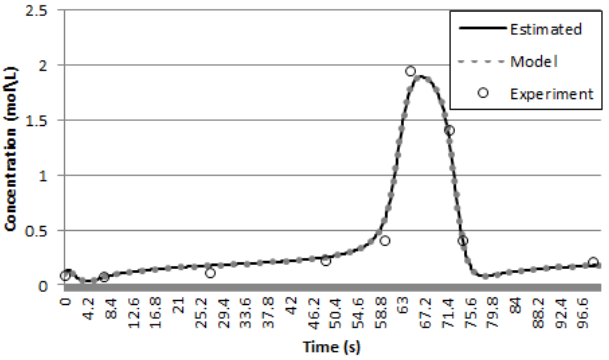

Figure 5. The RNA inhibition concentration.

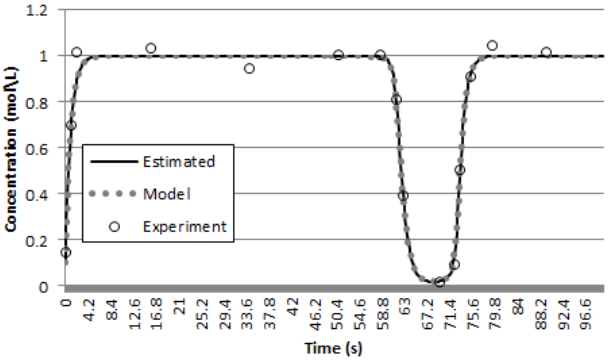

Figure 6. The ON switch Sw21 concentration.

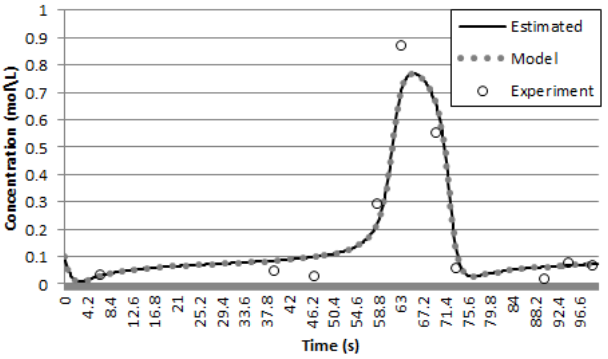

Figure 7. The ON switch Sw12 concentration.

B. *Microbial Lactose Operon*

To show the effectiveness of the proposed method in dealing with a huge number of model parameters, a model of microbial lactose operon is used. The model is developed by [27] to simulate the feedback regulation of lactose operon in the *Escherichia coli* bacteria. This model represents the cellular metabolism without the presence of glucose. Fundamentally, the lactose is fed into the cell by the *permease* enzyme, in which is extracted by the *β-galactosidase* to produce *allolactose*. Later, the *allolactose* is bonded onto the lactose repressor to permit the transcription process by the messenger RNA (mRNA). In general, the model is constructed by using the following reaction:

$$\frac{d[mRNA]}{dt} = k_1 - ((k_2+k_3)[mRNA]) + \frac{[mRNA_P]}{k_4} \quad (27)$$

$$\frac{d[B]}{dt} = \frac{[P_p]}{k_5} - ((k_6+k_7)[B]) \quad (28)$$

$$\frac{d[Al]}{dt} = k_7[B]\left(\frac{[L_i]}{k_8+[L_i]}\right) - k_9[P]\left(\frac{[Al]}{k_{10}+[Al]}\right) - ((k_{11}+k_{12})[Al]) \quad (29)$$

$$\frac{d[P]}{dt} = \frac{[P_p]}{k_5+k_{12}} - ((k_{13}+k_9)[P]) \quad (30)$$

where B, Al, P, $mRNA_p$, B_p, L_i, and P_p are the concentrations of *β-galactosidase*, *allolactose*, *permease*, partial mRNA, partial *β-galactosidase*, lactose internal, and partial *permease*, respectively. The parameter k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10}, k_{11}, k_{11}, and $k_{13}$are set to the values of 0.000000725, 0.411, 0.0226, 0.1, 2.0, 0.000833, 17600, 0.97, 21500, 1.95, 0.52, 0.83, and 0.65, respectively [27].

	PSO	DE	LEPSO	PSEO
Average Best Fitness	3.11×10^{-3}	2.57×10^{-4}	2.15×10^{-5}	1.21×10^{-6}
Standard Deviation	5.35×10^{-3}	1.10×10^{-3}	1.21×10^{-4}	5.05×10^{-6}
Computational Time (s)	150.3	113.2	201.7	95.5

Table 3. Comparison of average best fitness values and computational costs.

Table 3 shows the comparative results on the average fitness values found by using the methods. Similar to the first experiment, each method is executed 100 times independently and the average best fitness value, accompanying the corresponding standard deviation, is recorded. The population size and the maximum number of iterations are 50 and 200, respectively. For PSO, LEPSO, and PSEO methods, the inertia weight is altered to 0.7, while the self-exploitation and swarm-exploration rates are both changed to 4.5. Conversely, for the DE, LEPSO, and PSEO methods, the mutation rate is raised to 0.9. The results for this experiment present that the proposed PSEO method report the smallest average best fitness value than the other methods. Similarly, the standard deviation value is also small, which suggests that the method is still capable to consistently find plausible solutions in different runs. Furthermore, the results show that the computational cost used by the proposed method is relatively smaller than the other methods. This suggests that the method is practical in utilizing the computational cost more effectively compared to the other methods. Figure 8 presents the convergence behavior of the proposed methods with the existing PSO, DE, and PSEO methods. Generally, the method shows a capability in escaping the suboptimal solution, similar with the result that has been presented in the first experiment.

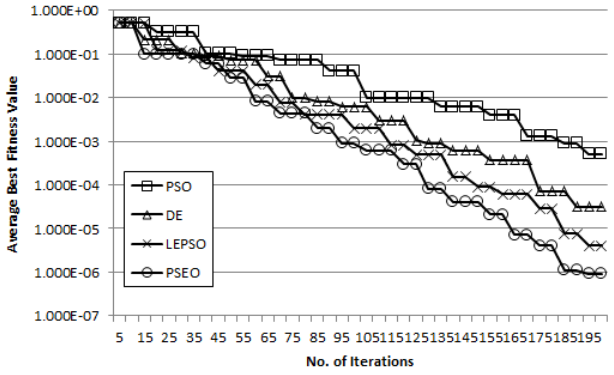

Figure 8. Convergence behavior of the PSO, DE, LEPSO, and PSEO methods.

Table 4 demonstrates the statistical analysis that shows the consistency of the model outputs produced by the estimated parameters based on the experimental measurements. Similar with the first experiment, a significant level of 95% is also used. The results present that the error variance points generally lie within the given intervals, which proves that the estimated parameters may produce valid model outputs for the experimental measurements with a 95% significant level. Figure 9 to 12 show the data fit of the model outputs produced by the estimated parameters, the original model, and the experimental measurements. Overall, the results advise that the proposed PSEO method is capable to estimate parameters reliably and robustly using the incomplete and noisy experimental measurements.

	mRNA	*β-galactosidase*	*Allolactose*	*Permease*
Real Point	1.21×10^{-7}	6.58×10^{-9}	3.18×10^{-2}	2.08×10^{-5}
Error Variance Point	1.22×10^{-7}	6.57×10^{-9}	3.18×10^{-2}	2.09×10^{-5}
Error Variance Intervals	$[1.09\times10^{-7}, 1.37\times10^{-7}]$	$[5.88\times10^{-9}, 7.38\times10^{-9}]$	$[2.85\times10^{-2}, 3.58\times10^{-2}]$	$[1.87\times10^{-5}, 2.35\times10^{-5}]$
χ^2 test	Pass			

Table 4. Statistical validation of the model outputs and the experimental measurements.

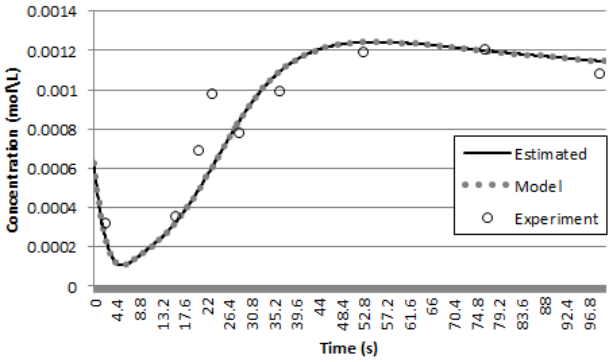

Figure 9. The mRNA concentration.

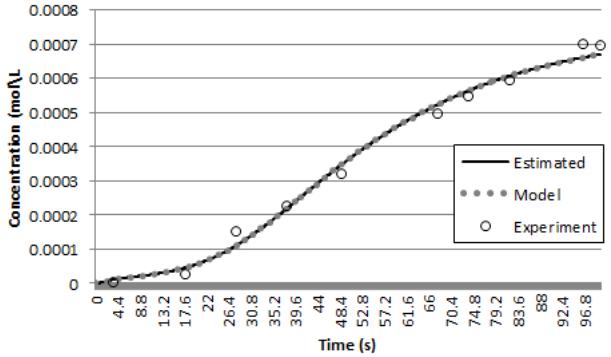

Figure 10. The *β-galactosidase* concentration.

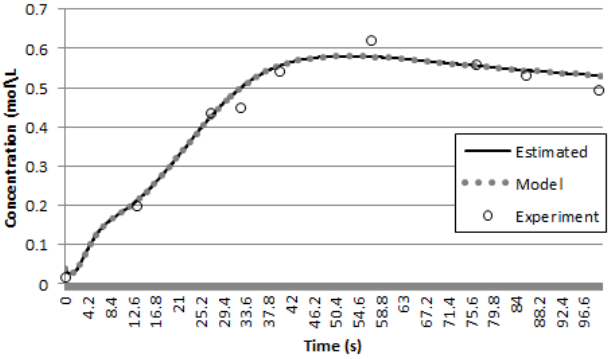

Figure 11. The *allolactose* concentration.

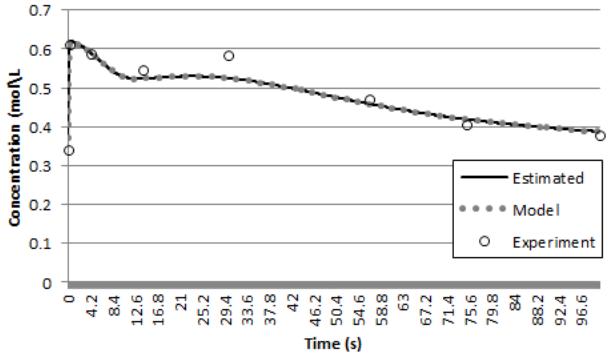

Figure 12. The *permease* concentration.

IV. Discussion

Parameter estimation problem has become an important key in the development of the biological models. The problem is focused on finding optimal parameters that can minimize the difference between the experimental measurements and the model outputs. Due to the nonlinearity of the model, it is very difficult to extract the parameters through in vivo experiments [1-3]. Therefore, in most cases, the problem is formulated as a nonlinear optimization problem, in which metaheuristics methods have been initiated and shown as a prospective approach lately [2-3]. However, these methods are commonly hindered by the need of a substantially huge amount of computational time and in some cases, the methods do not guarantee to converge to the global optimum solutions [17, 30]. This may lead to the inconsistency of finding better solutions. As a result, the estimation may provide parameters that may not be plausible to the given experimental measurements [31].

In this paper, a new hybrid optimization method called PSEO method is introduced. The method is developed based on the combination of the PSO and DE methods. Compared to the recently proposed LEPSO method, the present method incorporates a selective process during the iterations. The process involves the solutions to be ranked based on the fitness values. Then, the ranked population is separated to produce two sub-populations. The first sub-population consists of a set of solutions that yield potential fitness value, whereas the second sub-population contains the solutions that hold least plausible fitness values. The first sub-population, namely the potential sub-population, is subjected for neighboring improvement using the evolutionary operations adopted from the DE method. The improvement involves a mutation of each solution and the crossover with two neighboring solutions in the sub-population. Thus, only a certain number of solution is considered to be evaluated, which may be useful in utilizing the computational time. On the other hand, the second sub-population, named the weak sub-population, is used for random update. This strategy is to improve the weak solutions and at the same time to permit the method to escape the suboptimal solutions more effectively. To demonstrate the effectiveness of the method, two biological models have been selected for parameter estimation problem. These models, namely the synthetic oscillator and microbial lactose operon models, are used to evaluate the potential of the method in dealing with the incomplete and noisy experimental measurements, as well as to observe the capability of the method in handling both small and large number of model parameters.

The experimental results show that the proposed method is capable to estimate parameters in both models. The method had found better average best fitness value for 100 independent runs. The result present that the standard deviations of the best fitness values found by the method are relatively small, which suggested that the method can consistently find these values compared to the existing PSO, DE, and LEPSO methods. More interestingly, the method only requires a small amount of computational time to estimate the parameters. This provided an evidence that the method can utilize the computational cost more efficiently by using the proposed searching strategy. Moreover, the

convergence behaviors presented by both experiments show that the method is capable to escape the suboptimal solution more effectively than the other methods. To present the effectiveness of the estimated parameters in order to produce model outputs that are acceptable to the given experimental measurements, a statistical analysis based on the error variance points and intervals is utilized. The analysis shows that the error variance points of the model outputs produced by the estimated parameters lie within the intervals. This proves that the estimated parameters are valid for the given experimental measurements. Furthermore, the data fit between the experimental measurements, original model and the reconstructed model using the estimated parameters are presented. Generally, the estimated parameters generate model outputs that are consistent to the original model while using incomplete and noisy experimental measurements.

The proposed PSEO method has presented a potential achievement on estimating parameters using incomplete and noisy measurement noise. More likely, the proposed selective searching strategy has shown practicality for estimating parameters in the models that have both small and large number of parameters. However, another important issue in the development of biological models is not addressed in the experiments, which to show the effectiveness of the method for handling practical non-identifiable parameters. This problem is crucial due to the fact that each parameter in the model may produce unique stimulation according to the provided experimental measurements [28]. This issue will eventually lead to model selection problem [10]. Therefore, in the future, the proposed PSEO method will be used to evaluate the non-identifiable parameters and then will be verified for selecting plausible models using these parameters.

V. Conclusion

In this paper, a new hybrid optimization method based on PSO and DE methods is presented. The proposed method, called PSEO method, is used to find plausible parameters in the biological models according to the given noisy and incomplete experimental measurements. The PSEO method is developed based on the idea of introducing the evolutionary operations such as selection, mutation, and crossover applied in the standard DE method to the swarm-based searching strategy adopted by the original PSO method. Different from the other hybrid optimization methods, the PSEO method employs selective sub-population to enhance the current solutions for the next iterations. The experimental results showed that the proposed method is effective in estimating the parameters in the biological models by fitting the model prediction with the corresponding experimental measurements. Besides the overcoming the premature convergence suffered by the standard counterparts, the proposed method also improves the convergence speed compared to another recently proposed LEPSO method. More importantly, the proposed PSEO method is capable to handle the incompleteness of the experimental measurements during the estimation statistically.

Acknowledgment

This work is supported by Malaysia Ministry of Science, Technology and Innovation (MOSTI), organized by Malaysia Genome Institute (MGI) project number 07-05-MGI-GMB011 and managed by Universiti Teknologi Malaysia (UTM) grant number 73744.

References

[1] J. Sun, J. Garibaldi and C. Hodgman, “Parameter Estimation Using Metaheuristics in Systems Biology: A Comprehensive Review,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no.1, pp. 185-202, 2012.

[2] M. Rodriguez-Fernandez, J.A. Egea and J.R. Banga, "Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems," *BMC Bioinformatics*, vol.7, no.1, pp. 483-501, 2006

[3] A. Abdullah, D. Safaai., and S. Anwar, “Hybrid evolutionary clonal selection for parameter estimation of biological model,” *International Journal of Computer Applications in Engineering Sciences*, vol. 1, no.3, pp.313-319, 2011.

[4] D.F. Slezak, C. Suarez, G.A. Cecchi, G. Marshall and G. Stolovitzky, "When the optimal is not the best: parameter estimation in complex biological models," *PLoS One*, vol.5, no.10, pp. e13283, 2010

[5] L. Buhry, F. Grassia, A. Giremus, E. Grivel, S. Renaud and S. Saïghi, "Automated Parameter Estimation of the Hodgkin-Huxley Model Using the Differential Evolution Algorithm: Application to Neuromimetic Analog Integrated Circuits," *Neural computation*, vol.23, no.10, pp. 2599-2625, 2011.

[6] J.O. Lloyd-Smith, "Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases," *PLoS One*, vol. 2, no.2, pp. e180, 2007

[7] L.M.A. Bettencourt and R.M. Ribeiro, "Real time bayesian estimation of the epidemic potential of emerging infectious diseases," *PLoS One*, vol.2, no.5, pp. e2185, 2008.

[8] M. Ashyraliyev, J. Jaeger and J.G. Blom, "Parameter estimation and determinability analysis applied to Drosophila gap gene circuits," *BMC Systems Biology*, vol. 2, no.1, pp. 83-102, 2008.

[9] X. Sun, L. Jin and M. Xiong, “Extended kalman filter for estimation of parameters in nonlinear state-space models of biochemical networks,” *PLoS One*, vol.3, no.11, pp. e3758, 2008.

[10] G. Lillacci and M. Khammash, "Parameter estimation and model selection in computational biology,” *PLoS Computational Biology*, vol.6, no.3, pp. e1000696, 2010.

[11] N. Zeng, Z. Wang, Y. Li, M. Du and X. Liu, "A Hybrid EKF and Switching PSO Algorithm for Joint State and Parameter Estimation of Lateral Flow Immunoassay Models," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no.2, pp. 321-329, 2012.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE International Conference on Neural Networks*, vol. 4, pp. 1942-1948, 1995

[13] F. Glover, M. Laguna and Marti, "Fundamentals of scatter search and path relinking," *Control and Cybernetics*, vol. 39, no. 3, pp. 653-684, 2000

[14] Y. Jin and J. Branke, "Evolutionary optimization in uncertain environments-a survey," *IEEE Transactions on*

Evolutionary Computation, vol. 9, no.3, pp. 303-317, 2005

[15] N. Tutkun, "Parameter estimation in mathematical models using the real coded genetic algorithms," *Expert Systems with Applications*, vol. 36, no. 2, 3342-3345, 2009.

[16] R. Storn and K. Price, "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, 1997.

[17] S. Das, A. Abraham and A. Konar, "Particle swarm optimization and differential evolution algorithms: technical analysis, applications and hybridization perspective," *Studies in Computational Intelligence*, vol. 116, pp.1-38, 2008

[18] N. Noman and H. Iba, "Accelerating differential evolution using an adaptive local search," *IEEE Transaction on Evolutionary Computation*, vol.12, no.1, pp.107-125, 2008

[19] A. Abdullah, S. Deris, S.Z.M. Hashim, M.S. Mohamad, and S.N.V. Arjunan, "An improved local best searching in Particle Swarm Optimization using Differential Evolution," *11th International Conference on Hybrid Intelligent Systems*, pp. 115-120, 2011

[20] W.H. Ho and A.L.F. Chan, "Hybrid Taguchi-Differential Evolution Algorithm for Parameter Estimation of Differential Equation Models with Application to HIV Dynamics," *Mathematical Problems in Engineering*, 2011, doi: 10.1155/2011/514756

[21] M. Rodriguez-Fernandez, P. Mendes and J.R. Banga,"A hybrid approach for efficient and robust parameter estimation in biochemical pathways," *Biosystems*, vol. 83, no. 2-3, 248-265, 2006

[22] M. Ashyraliyev, Y. Fomekong-Nanfack, J.A. Kaandorp and J.G. Blom, "Systems biology: parameter estimation for biochemical models," *FEBS Journal*, vol. 276, no. 4, pp. 886-902, 2009

[23] Y. Fomekong-Nanfack, J.A. Kaandorp and J. Bloom, "Efficient parameter estimation for spatio-temporal models of pattern formulation: case study of Drosophilia melanogaster," *Bioinformatics*, vol.23, pp.3356-3363, 2007

[24] T.P. Runarsson and X. Yao, "Stochastic ranking for constrained evolutionary optimization," *IEEE Transactions on Evolutionary Computation*, vol.4, no.3, pp.284-294, 2000

[25] M. Forlin, D. Slanzi and I. Poli,"Combining Probabilistic Dependency Models and Particle Swarm Optimization for Parameter Inference in Stochastic Biological Systems," *Proceedings of the 2011 2nd International Congress on Computer Applications and Computational Science*, pp. 437-443, 2012

[26] J. Kim and E. Winfree, "Synthetic in vitro transcriptional oscillators," *Molecular Systems Biology*, vol. 7, no. 1, pp. 1-15, 2011

[27] N. Yildirim and M.C. Mackey, "Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data," *Biophysical Journal*, vol. 84, no. 5, pp. 2841-2851, 2003.

[28] E. Balsa-Canto, A.A. Alonso and J.R. Banga, "Computational procedures for optimal experimental design in biological systems," *IET Systems Biology*, vol.2, no. 4, pp.163-172, 2008.

[29] P. Kaelo and M.M. Ali, "Differential evolution algorithms using hybrid mutation," *Comput Optim Appl*, vol. 37, no. 2, pp. 231-246, 2007.

[30] F. Mahar, S.S.A. Ali and A. Hussain,"Design of Fixed Order Controller by Using Particle Swarm Optimization and System Simulation,"*International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp.459-466, 2012.

[31] D. A. de Waal, S. E. S. Campher and J. V. du Toit,"Estimating Those Transformations That Produce the Best-fitting Additive Model: Smoothers Versus Universal Approximators," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp.9-17, 2012.

[32] A. I. S. Nascimento and C. J. A. Bastos-Filho,"Designing Cellular Networks using Particle Swarm Optimization and Genetic Algorithms," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 4, pp.496-505, 2012.

Author Biographies

Afnizanfaizal Abdullah received the B.Sc. and M.Sc. degrees in computer science from Universiti Teknologi Malaysia, Skudai, Malaysia, in 2007 and 2010, respectively. He is currently working toward the Ph.D. degree in computer science at Universiti Teknologi Malaysia, Skudai, Malaysia. His current research interests include computational systems and synthetic biology, in particular, estimating parameters and model selection of biological models using hybrid metaheuristics algorithms and methods.

Safaai Deris received the M.E. degree in industrial ingineering and the D.Eng. degree in computer and system sciences from Osaka Prefecture University, Sakai, Japan, in 1989 and 1997, respectively. He is currently a Professor of Artificial Intelligence and Software Engineering at the Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Skudai, Malaysia. His recent academic interests include the application of intelligent techniques in scheduling and bioinformatics. He is also a reviewer for several refereed journals.

Mohd Saberi Mohamad received the B.Sc. and M.Sc. degrees in computer science from Universiti Teknologi Malaysia, Skudai, Malaysia, in 2002 and 2005, respectively. He received the Ph.D. degree in intelligent systems for bioinformatics from Osaka Prefecture University, Sakai, Japan, in 2010. He is currently a Lecturer at the Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia. His journal papers in the field of gene selections using computational intelligence approaches have been published with the total number of more than ten papers. His research interests include computational methods such as particle swarm optimizers, hybrid approaches, genetic algorithms, support vector machines, and neural networks.

Siti Zaiton Mohd Hashim received her B.Sc. in computer science in 1990, M.Sc. in computer science in 1997 and Ph.D. (Softcomputing in Control) in 2005 from University of Hartford, USA, University of Bradford, UK, and Sheffield University, UK consecutively. Her research interests are soft computing and intelligent systems. She is currently the head of Office of Postgraduate Studies, Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia. She is now actively doing research on RFID middleware and soft computing fundamental research.