An Improved Swarm Optimization for Parameter Estimation and Biological Model Selection

Afnizanfaizal Abdullah[1]*, Safaai Deris[1], Mohd Saberi Mohamad[1]*, Sohail Anwar[2]

1 Artificial Intelligence and Bioinformatics Group (AIBIG), Faculty of Computing, Universiti Teknologi Malaysia, UTM, Johor, Malaysia, 2 Pennsylvania State University, Altoona, Pennsylvania, United States of America

Abstract

One of the key aspects of computational systems biology is the investigation on the dynamic biological processes within cells. Computational models are often required to elucidate the mechanisms and principles driving the processes because of the nonlinearity and complexity. The models usually incorporate a set of parameters that signify the physical properties of the actual biological systems. In most cases, these parameters are estimated by fitting the model outputs with the corresponding experimental data. However, this is a challenging task because the available experimental data are frequently noisy and incomplete. In this paper, a new hybrid optimization method is proposed to estimate these parameters from the noisy and incomplete experimental data. The proposed method, called Swarm-based Chemical Reaction Optimization, integrates the evolutionary searching strategy employed by the Chemical Reaction Optimization, into the neighbouring searching strategy of the Firefly Algorithm method. The effectiveness of the method was evaluated using a simulated nonlinear model and two biological models: synthetic transcriptional oscillators, and extracellular protease production models. The results showed that the accuracy and computational speed of the proposed method were better than the existing Differential Evolution, Firefly Algorithm and Chemical Reaction Optimization methods. The reliability of the estimated parameters was statistically validated, which suggests that the model outputs produced by these parameters were valid even when noisy and incomplete experimental data were used. Additionally, Akaike Information Criterion was employed to evaluate the model selection, which highlighted the capability of the proposed method in choosing a plausible model based on the experimental data. In conclusion, this paper presents the effectiveness of the proposed method for parameter estimation and model selection problems using noisy and incomplete experimental data. This study is hoped to provide a new insight in developing more accurate and reliable biological models based on limited and low quality experimental data.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: afnizanfaizal@utm.my (AA); saberi@utm.my (MSM)

Introduction

Computational systems biology has become an increasingly important research area in the recent years [1], [2]. This field of research is aimed to gain better understanding of how complex biological process respond as a system within living cells. This is often facilitated by using computational models [1], [3], [4]. These models are commonly constructed based on specific mathematical formulations, such as ordinary differential equations (ODEs), to measure the quantity of certain biochemical compounds within a time unit. The development of these models usually involves two stages: network structure identification and parameter estimation [3], [4], [5]. The network structure identification stage is conducted majorly by modelling experts, in which the structure of the ODEs is mathematically verified [3]. Alternatively, the parameter estimation stage is performed to evaluate if the model parameters can accurately simulate the actual processes obtained from the experimental analyses [3], [4].

In general, biological models are equipped with a set of parameters to signify the physical properties of the systems, such as kinetic constants and reaction rates. These parameters are generally difficult to be identified in high-throughput experiments [3]. Instead, they are rather estimated based on the available experimental data. This is usually performed by calibrating the model outputs with the corresponding experimental data. In most cases, nonlinear optimization methods are utilized to find the optimal parameters that can minimize the difference between the model outputs and the corresponding experimental data. However, this is a challenging task as the models are frequently hampered by the nonlinearity of the biological processes [4], [5]. Hence, parameter estimation is usually considered as a nonlinear multimodal problem, in which the estimation processes may sometimes lead to several insignificant parameters that are less accurate if only based on the actual biological processes [5]. Furthermore, the available experimental data are often incomplete and regularly exhibit a substantial level of measurement noise [5], [6]. These limitations may cause difficulty in finding plausible parameters

that represent the actual biological processes. This is a problem of non-identifiability [7], which apprehends the tasks to uniquely estimate the unknown parameters [8], [9], [10].

Currently, there is an increase of the number of nonlinear optimization methods proposed to estimate the parameters in the biological models [1], [4], [11]. The aim of these methods is to find the optimal parameter set which can produce the model outputs that closely fit into the corresponding experimental data. In general, this problem is formulated as the fitness function, usually based on the nonlinear least squares [12]. Conventionally, derivative-based optimization methods are utilized, including maximum likelihood [13] and gradient decent [14] methods. More currently, a local optimization method, namely Extended Kalman Filter (EFK) [15] method, is employed [16]. Lillacci and Khammash [6], [10] introduced an improved EFK method that incorporates the continuous model outputs and the experimental measurements to estimate the parameters using state space searching approach. Additionally, Zheng and co-workers [17] proposed inequality constraints to improve the estimation by using the EFK method. However, both improved methods commonly require the use of model refinement phases to avoid the searching processes from being trapped into the suboptimal solutions. Furthermore, these methods need to consider the limitations of the EFK method that heavily relies on a good set of initial values for both states and parameters in the models [16].

In contrast, several previous works have presented prospective achievements by using meta-heuristic methods [5]. Rodriguez-Fernandez and co-workers [11] employed Scatter Search Algorithm (SSA) [18] to estimate the parameters in benchmark biological models. The study showed that the recombination searching strategy applied by this method was robust to measurement noise in the experimental data. Similarly, Particle Swarm Optimization (PSO) [19] and Genetic Algorithm (GA) [20] methods were also used to estimate the parameters in biological systems, which showed promising results [21], [22]. More recently, evolutionary-based meta-heuristics methods have received remarkable attentions [1], [3], [23]. Generally, these methods utilize evolutionary operations such as crossover, mutation, and selection operations to exploit the information of the solutions in the population. Tashkova and co-workers [3] suggested that the use of Differential Evolution (DE) [24] method is more practical compared to the existing meta-heuristic methods. However, it was also presented that the method may use a substantial amount of computational cost to obtain the best solution [1], [3]. Despite the capabilities, there is no guarantee that these methods will converge to the global optimum solutions [5].

To overcome these limitations, the hybrid meta-heuristics methods are utilized [2], [3], [25]. Commonly, these methods combine different searching strategies from the distinctive methods. Rodriguez-Fernandez and co-workers [26] introduced a new robust hybrid method based on the Evolutionary Strategy (SRES) [27] method. The proposed method had successfully reduced the computational time while handling the measurement noise effectively. In addition, Chen and Wang [28] introduced a new hybrid method which incorporates the DE method with a geometric mean mutation. The method was evaluated using a cellulose hydrolysis model. The experimental results showed that the method was capable to estimate the initial values of the model parameters, in which later were used for gradient-based optimization approach. We had proposed a new hybrid optimization method based on PSO and DE that showed prospective achievement in dealing incomplete and noisy experimental data [29]. In a more recent work [30], we introduced a new hybrid optimization method based on Firefly Algorithm (FA) method [31]

and DE methods. To enhance the efficiency of the computational time of the existing methods, the proposed method was used to discriminate the solutions into two sub-populations based on the current fitness values. The sub-population that contained solutions with plausible fitness was exploited for further improvement using a proposed searching strategy based on the FA and DE methods.

In this paper, a new hybrid meta-heuristic method is proposed. The method, called Swarm-based Chemical Reaction Optimization (S-CRO) method, is developed based on the combination of the FA method and a recently proposed evolutionary method, Chemical Reaction Optimization (CRO) [32]. In particular, the proposed S-CRO method is distinguished from the previously proposed method in [29], as the proposed method employs the evolutionary operations of the CRO method to enhance the swarm-based search strategy applied in the FA method, instead of using evolutionary operations of DE method to enhance the PSO method. Thus, this provides a new approach to retain the robustness over the measurement noise that exhibits the experimental data during the searching process [1], [3], [21]. The effectiveness of the proposed method in estimating parameters was evaluated using a simulated nonlinear model [33] and two
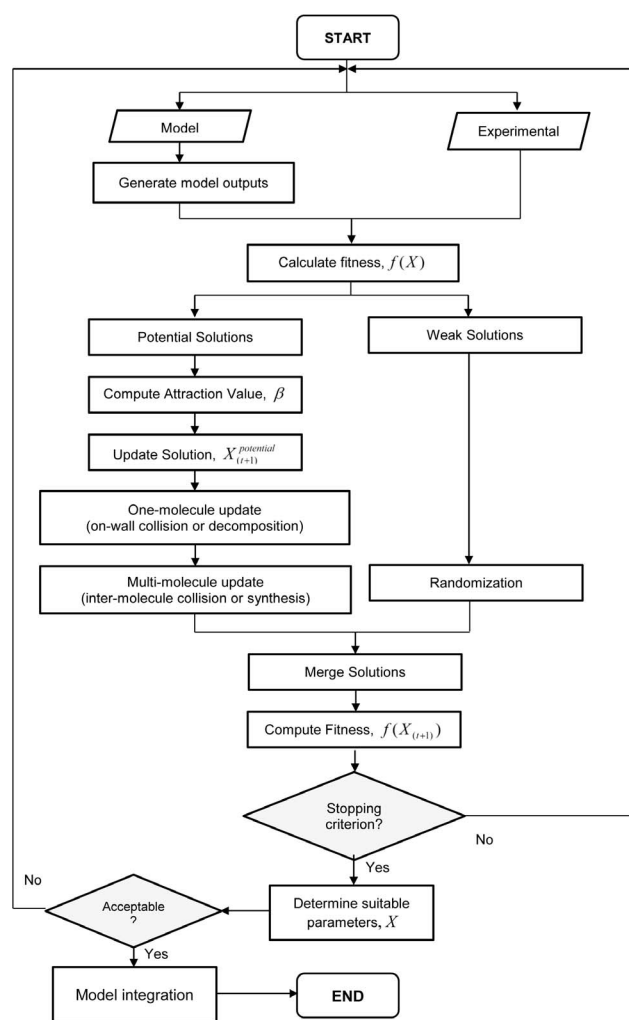

Figure 1. Parameter estimation using optimization method. The parameter estimation procedure begins with a prediction from the model and data obtained from experiments. The model predictions are generated from an ODE solver.
doi:10.1371/journal.pone.0061258.g001

biological models: synthetic transcriptional oscillators [34], and extracellular protease production [35] models. The performances of the proposed S-CRO method, in terms of convergence to better fitness values and the computational cost used, were compared with those produced by using the standard DE, FA, and CRO methods. In addition, the model outputs generated by the estimated parameters were validated using statistical analysis to address the effectiveness of the method in term of non-identifiability [30]. Furthermore, the method was also validated for model selection, which was performed using the Akaike Information Criterion (AIC) [30], [36]. The paper is organized as follows: Firstly, the problem formulation is introduced and the details of the FA, CRO, and the proposed S-CRO methods are described. The validation analyses for non-identifiability and model selection are also explained. Then, the simulation results are presented. Next, the discussion on the obtained results is addressed, which deliberates the contributions of this work. Lastly, the paper is summarized in the conclusion section.

Methods

Problem Formulation

The parameter estimation of the biological models can be formulated as follows. Suppose a model contains a biochemical compound, s, that is formed as $s(u,X)$, which consists of a set of parameters, $X = \{x_1, x_2, x_3, ..., x_D\}$, where D is the total number of parameters, and the input signal, u. Thus, the reaction rate of the compound s is given as follows

$$\begin{cases} \dfrac{ds}{dt} = g(s(u,X), \tau) \\ s_{(\tau_0)} = s_{(0)} \\ y = h(s(u,X), \tau) + \varepsilon \end{cases} \quad (1)$$

where g and h are the nonlinear functions, τ is the sampling time, y is the model output and ε is the measurement noise, which is generated by random Gaussian noise with zero mean [6,10,30]. Thus, the parameter estimation problem is aimed to find the optimal parameter set, $\hat{X}$, which minimizes the difference between the model output, y, and the corresponding experimental data, $y^{\exp}$. This is commonly performed by using the nonlinear least squared error function, $f(X)$, defined as follows:

$$f(X) = \arg\min \sum_{d=1}^{D} \sum_{n=1}^{N} \left(y_n^{\exp} - y_n(x_d) \right)^2 \quad (2)$$

where N is the total number of samples [30]. This function is considered as the fitness function in most optimization methods. Since the experimental data is hampered by the measurement noise and is often incomplete, finding the plausible parameters that may minimize this equation is difficult. Figure 1 shows the general framework of solving parameter estimation problem using nonlinear optimization methods.

Firefly Algorithm

The FA method is a swarm-based meta-heuristics method [31]. The method is inspired by the natural social behaviours of a firefly population. In nature, the fireflies produce flashing light, which is generated by bioluminescence chemical reactions. The light is used to attract mating partners. The fireflies also use the light as a communication medium to prevent potential preys. In the FA

method, the solutions are formulated as the fireflies which carry a vector of variables used to compute the fitness functions. The vector of ith solution, X_i, is formed as follows:

$$X_i = \{x_{i1}, x_{i2}, x_{i3}, ..., x_{iD}\} \quad (3)$$

where D is the dimension size of the problem. Each ith solution computes the individual fitness value, calculated by a specific fitness function, such as non-linear least squared errors. The fitness value can be represented as the light intensity of the natural firefly. The fitness value of the current ith solution is compared with the

---

Swarm-based Chemical Reaction Optimization (S-CRO) algorithm

Input: Randomly initialized D number of vectors (parameter set of the model): $X_i = \{x_{i1}, x_{i2}, x_{i3}, ..., x_{iD}\}$

Output: Vectors of the global best solution: X_{best}

Begin

For $i \leftarrow 1$ to NP solutions

$\quad X \leftarrow x_i^L + rand \times \left(x_i^U - x_i^L \right);$ // Initialize population

$\quad f(X_i) \leftarrow f(x_{i1}, x_{i2}, x_{i3}, ..., x_{iD});$ // Evaluate fitness value

End For

$\quad X_{best} \leftarrow select\_best(X);$

For $t \leftarrow 1$ to maximum number of iterations

   // Sort population based on the fitness value

$\quad X^{sub1} \leftarrow select\_potential\_solution(X);$

$\quad X^{sub2} \leftarrow select\_weak\_solution(X);$

For $i \leftarrow 1$ to number of $X^{potential}$ solutions

For $j \leftarrow 1$ to number of $X^{potential}$ solutions

If $(f(X_{i(t)}^{sub1}) > f(X_{j(t)}^{sub1}))$ **then**

$\quad r_{ij} \leftarrow \left\| X_{i(t)}^{sub1} - X_{j(t)}^{sub1} \right\|$ // calculate distance

$\quad \beta \leftarrow \beta_0 e^{-\mu r_{ij}^2};$ // calculate attractiveness

$\quad X_{i(t+1)}^{sub1} \leftarrow \left\{ x_{id(t)}^{sub1} + \beta \left(x_{jd(t)}^{sub1} - x_{id(t)}^{sub1} \right) + rand_1 \left(rand_2 - \dfrac{1}{2} \right) \right\}$

   // apply Chemical Reaction Optimization (CRO)

If $(rand_3 < 0.5)$ **then**

If $(rand_3 \leq 0.33)$ **then**

$\quad X_{i(t+1)}^{sub1} \leftarrow on\_wall\_collision(X_{i(t+1)}^{sub1});$

Else

$\quad X_{i(t+1)}^{sub1} \leftarrow decomposition(X_{i(t+1)}^{sub1});$

Else

If $(rand_3 > 0.7)$ **then**

$\quad X_{i(t+1)}^{sub1} \leftarrow molecule\_collision(X_{i(t+1)}^{sub1}, X_{j(t)}^{sub1});$

Else

$\quad X_{i(t+1)}^{sub1} \leftarrow synthesis(X_{i(t+1)}^{sub1}, X_{j(t)}^{sub1});$

End If

End For

End For

For $i \leftarrow 1$ to number of X^{sub2} solutions

$\quad X_{i(t+1)}^{sub2} \leftarrow \{x_{best(t)} + rand_4 \times \left(x_{id(t)}^{sub2} - x_{best(t)} \right);$

End For

$\quad X \leftarrow combine(X^{sub1}, X^{sub2});$

$\quad X_{best(t+1)} \leftarrow select\_best(X);$

$\quad t \leftarrow t + 1;$

End For

End Begin

Figure 2. Swarm-based Chemical Reaction Optimization (S-CRO) Algorithm. The proposed S-CRO method is composed of three main steps as indicated by the shaded sections. The first step sorts the population according to fitness into two groups: potential and weak solution groups. In the second step, the potential solutions are subjected to evolutionary operations. In the third step, a random vector update is performed to the weak solutions in order to allow the method to escape from the suboptimal solutions more effectively.
doi:10.1371/journal.pone.0061258.g002
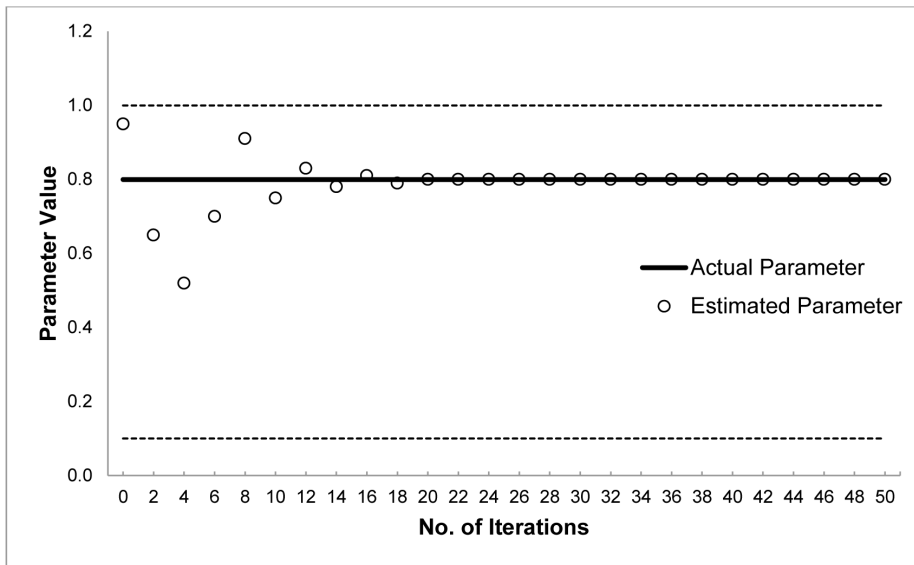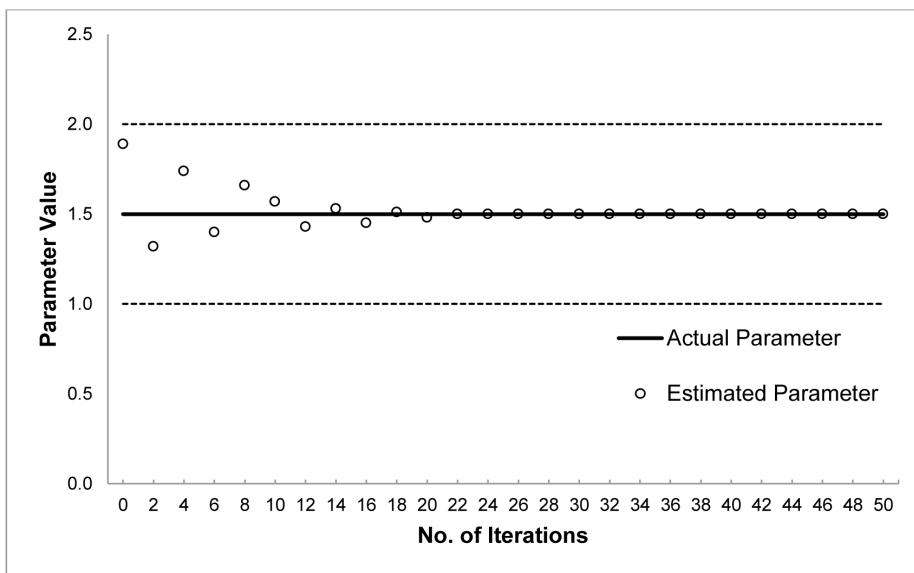
Figure 3. The estimated parameter k_1 and k_2 for simulated nonlinear model over the number of iterations by the proposed S-CRO method. The plots show the parameter estimation of the simulated nonlinear model. The dashed lines represent the upper and lower boundaries values, bold lines represent the actual parameter values, and the circles represent the estimated parameter values. Graph A represents the estimated parameter k_1 and graph B represents the estimated parameter k_2.
doi:10.1371/journal.pone.0061258.g003

jth neighbouring solutions. If the fitness value of the neighbouring solution is better than the current solution, the distance, r_{ij}, is computed using the standard distance function, such as *Euclidean* distance, as follows [31]:

$$r_{ij} = \left\| X_{i(t)} - X_{j(t)} \right\| = \sqrt{\sum_{d=1}^{D} \left(x_{id} - x_{jd} \right)^2} \qquad (4)$$

Using this information, the attractive value of each ith solution is further calculated using the following equation [31]

$$\beta = \beta_0 e^{-\mu r_{ij}{}^2} \qquad (5)$$

where for each ith firefly with its corresponding jth neighbor, β is the attractive value, μ is the predefined light absorption coefficient and β_0 is the initial attractive value [31]. Then, this attractive value is used to update the vector of the ith solution:

$$X_{i(t)} = X_{i(t)} + \beta\left(X_{j(t)} - X_{i(t)}\right) + rand_1\left(rand_2 - \frac{1}{2}\right) \qquad (6)$$

where $rand_1$ and $rand_2$ are uniformly distributed random values between 0 to 1 [31]. Thus, this permits the population to move towards the solution that represents the current best fitness value and exploits the searching space more effectively [31]. The searching process is repeated until the maximum number of iterations is reached.

Artificial Chemical Reaction Optimization

The Chemical Reaction Optimization (CRO) is another meta-heuristic method, which is based on the chemical reactions of molecules to reach low energy stable state [32]. The method manipulates the reactions involving molecules including collision, synthesis and diffusion. In these reactions, the energy is transferred to a stable state is reached. In this method, these molecules are formulated as solutions. Each solution holds two properties: potential and kinetic energies [32]. The potential energy represents the fitness value calculated using the fitness function. On the other hand, the kinetic energy, K_E, represents a tolerance measurement for the solution to be transformed into a less favourable solution, thus permitting the method to escape the local optima more effectively [32].

In this method, the searching process can be divided into two major actions: single and multi-molecule reactions. The single-molecule reaction usually involves only one solution to be improved using on-wall collision or decomposition processes [32]. Biologically, the on-wall collision occurs when a molecule bumps into a cell wall and then bounces into another direction within the cell. The offspring solutions are mostly less distinctive compared to the parent solution before the collision [32]. For the ith solution, the solution intends to gain better fitness from the neighbouring jth solution. The vectors in this solution are updated only if the following rule is met

$$f(X_i) + K_{E_i} \geq f(X_j) \qquad (7)$$

in which the following equation is formed

$$K_{E_j} = \left(f(X_i) + K_{E_i} - f(X_j)\right) \times a \qquad (8)$$

where $a \in [LR, 1]$, in which LR is the loss rate that limits the maximum percentage of kinetic energy lost [32]. Alternatively, the decomposition process occurs when a molecule is diffused into two or more molecules after the collision with the cell wall. The resultant molecules are supposedly to be much different compared to the original molecule. This process is executed if the following rule is met:

$$f(X_i) + K_{E_i} \geq f(X_{\bar{i}}) + f(X_{\bar{\bar{i}}}) \qquad (9)$$

where $X_{\bar{i}}$ and $X_{\bar{\bar{i}}}$ are the offspring solutions from the original solution, X_i, produced after the collision. Based on this rule, a new variable is assigned as follows:

$$temp = f(X_i) + K_{E_i} - f(X_{\bar{i}}) - f(X_{\bar{\bar{i}}}) \qquad (10)$$

which is used to generate new kinetic energies, $K_{E_{\bar{i}}}$ and $K_{E_{\bar{\bar{i}}}}$, as follow:

$$K_{E_{\bar{i}}} = temp \times k \qquad (11)$$

$$K_{E_{\bar{\bar{i}}}} = temp \times (1 \times k) \qquad (12)$$

Table 1. Performance comparison of DE, FA, CRO, and S-CRO methods.

	DE	FA	CRO	S-CRO
Noise: 5%				
Average Fitness Value	3.92×10^{-7}	8.35×10^{-5}	8.87×10^{-7}	7.21×10^{-9}
Standard Deviation	4.35×10^{-7}	7.11×10^{-5}	6.51×10^{-7}	8.43×10^{-9}
Computational Time (s)	83.1	103.4	91.9	72.2
Noise: 10%				
Average Fitness Value	1.82×10^{-3}	5.53×10^{-4}	7.01×10^{-3}	8.91×10^{-6}
Standard Deviation	2.11×10^{-3}	4.21×10^{-4}	5.25×10^{-3}	1.05×10^{-6}
Computational Time (s)	98.8	110.8	107.5	81.9
Noise: 15%				
Average Fitness Value	2.95×10^{-2}	8.22×10^{-1}	6.31×10^{-2}	5.15×10^{-4}
Standard Deviation	2.01×10^{-2}	7.42×10^{-1}	5.77×10^{-2}	1.27×10^{-4}
Computational Time (s)	117.8	200.8	135.5	97.9

where k is a uniform random number between 0 and 1 [32]. The value is used to generate two newly formed solutions and which are then added into the population.

For multi-molecule reaction, there are two important processes, namely inter-molecule collision and synthesis. The inter-molecule collision involves two solutions that collide with each other and bounce away in two separate directions. The effect of the energy change of the solutions is similar to those in the on-wall collision, except that this process involves two solutions instead of a single solution. The process is performed if the following rule is met [32]:

$$f(X_i) + K_{E_i} + f(X_j) + K_{E_j} \geq f(X_i') + f(X_j') \tag{13}$$

in which the following variable is produced

$$temp = \left(f(X_i) + K_{E_i} + f(X_j) + K_{E_j} \right) - \left(f(X_i') + f(X_j') \right) \tag{14}$$

to generate new kinetic energies, $K_{E_i'}$ and $K_{E_j'}$, as follow

$$K_{E_i'} = temp \times k \tag{15}$$

$$K_{E_j'} = temp \times (1 - k) \tag{16}$$

Thus, these values are used to generate two newly formed solutions and are added into the population. Otherwise, a synthesis process is performed, which involves two solutions to be combined together after the collision. This process is executed if the following rule is accepted:

$$f(X_i) + K_{E_i} + f(X_j) + K_{E_j} \geq f(X_i') \tag{17}$$

Based on this rule, the kinetic energy of the newly produced solution, X_i', is generated as follows:
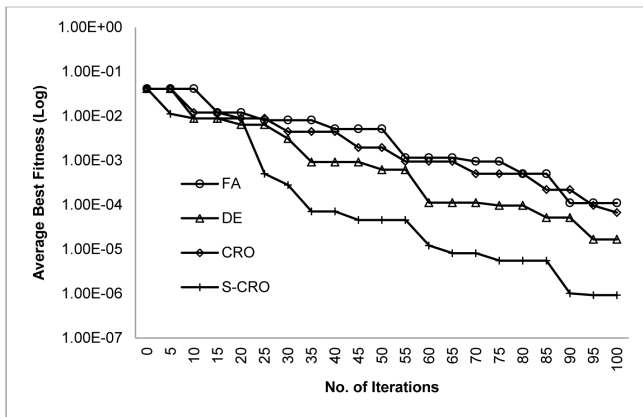
A
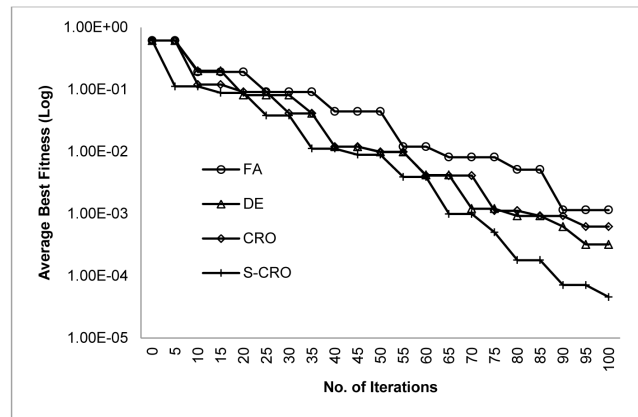
B
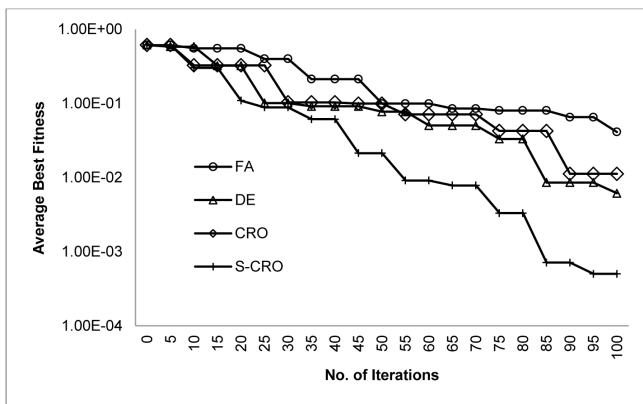
C

Figure 4. Convergence behaviours of the DE, FA, CRO, and S-CRO methods for the synthetic transcriptional oscillators model. The plots show the average best fitness values of DE, FA, CRO, and the proposed methods in each iteration. Graph A, B, and C represents the convergence behaviours for 5%, 10%, and 15% measurement noise, respectively.
doi:10.1371/journal.pone.0061258.g004
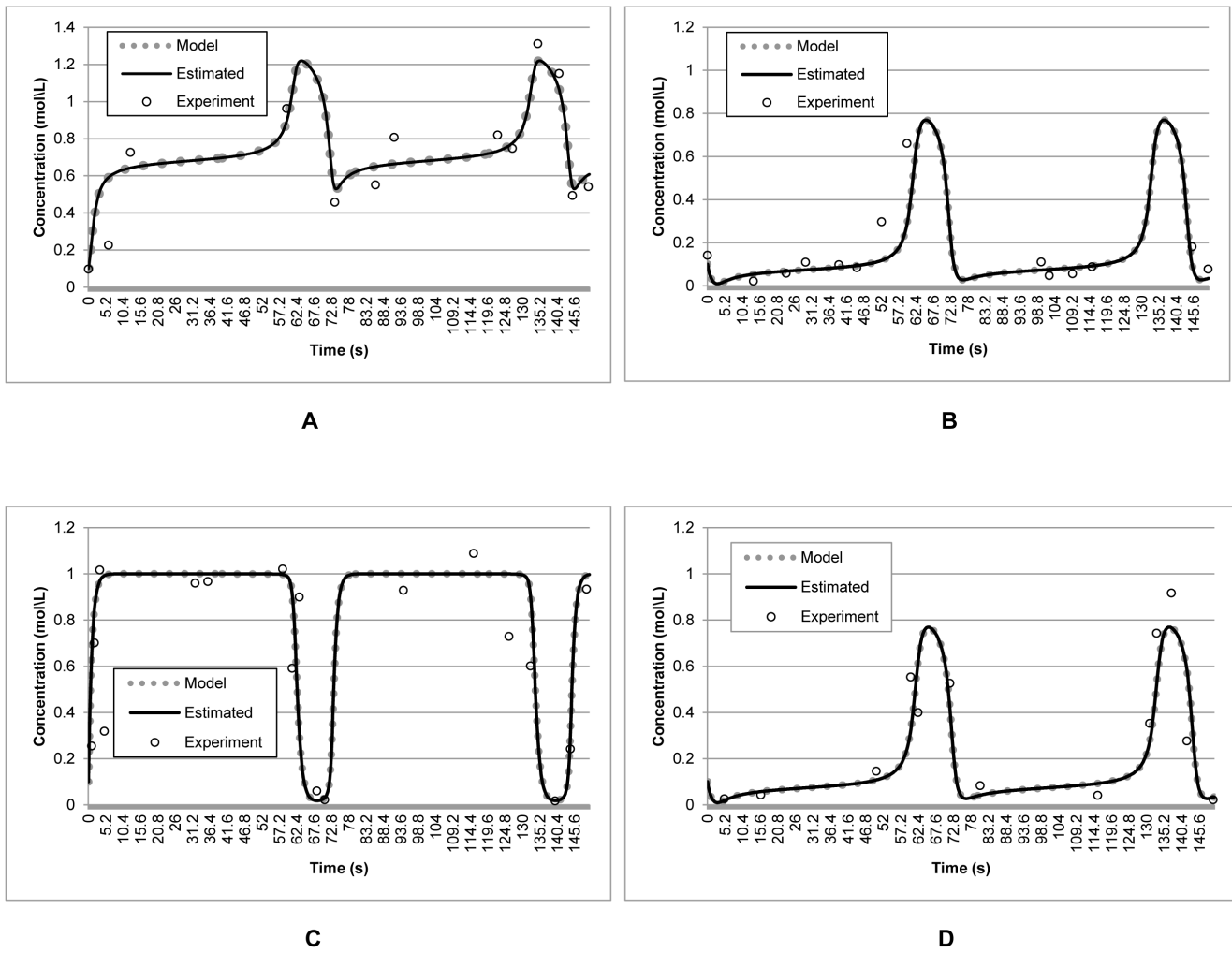
Figure 5. Data fit of model outputs produced by the estimated parameters and the corresponding experimental measurements for the synthetic transcriptional oscillators model. The data points (circles) represent synthetic measurements obtained by adding Gaussian noise to the model prediction (dotted line). The straight lines represent the reconstructed model using the parameters estimated by the proposed S-CRO method. Graph A, B, C, and D represents concentrations of RNA activation, RNA inhibition, ON-state switch Sw21, and ON-state switch Sw12, respectively.
doi:10.1371/journal.pone.0061258.g005

$$K_{E_{\bar{i}}} = f(X_i) + K_{E_i} + f(X_j) + K_{E_j} - f(X_{\bar{i}}) \qquad (18)$$

Table 2. Estimated parameters by DE, FA, CRO, and S-CRO methods using the noisy and incomplete experimental data (15% white Gaussian noise).

| Parameter	Actual	DE	FA	CRO	S-CRO
k_1 (μM)	0.57	0.52	1.18	0.52	0.56
k_2	1.5	1.90	1.31	2.09	1.5
k_3 (μM)	2.5	2.31	3.32	2.45	2.5
k_4	6.5	5.91	6.03	7.21	6.5
k_5	6.5	6.4	5.95	6.4	6.45

doi:10.1371/journal.pone.0061258.t002

As a result, the value of $K_{E_{\bar{i}}}$ is substantially large compared to K_{E_i} and K_{E_j} as the value of $f(X_{\bar{i}})$ is expected to be equal to $f(X_i)$ or $f(X_j)$ [32]. This process is important to allow the method to escape the local optima more effectively. The process is iterated until the maximum number of iterations is reached.

Swarm-based Chemical Reaction Optimization (S-CRO)

In this paper, a new hybrid optimization method is proposed based on the CRO and FA methods. The method is developed to introduce the combinatorial searching strategy employed by the evolutionary operations in the CRO method to the swarm-based search strategy of the FA method. This is due to the fact that the evolutionary operations are practical to handle the measurement noise in the experimental data [1], [3], [30]. Basically, the ith

Table 3. Non-identifiability validation (15% white Gaussian noise).

	RNA activator	RNA inhibitor	ON-state switch Sw21	ON-state switch Sw12
Real Variance (ξ_n)	9.28×10^{-2}	8.04×10^{-2}	1.33×10^{-1}	1.28×10^{-2}
Variance Point ($\hat{\sigma}_n^2$)	9.26×10^{-2}	8.04×10^{-2}	1.32×10^{-1}	1.27×10^{-2}
Variance Interval	$[8.43 \times 10^{-2}, 1.02 \times 10^{-1}]$	$[7.32 \times 10^{-2}, 8.89 \times 10^{-2}]$	$[1.21 \times 10^{-1}, 1.47 \times 10^{-1}]$	$[1.25 \times 10^{-2}, 1.37 \times 10^{-2}]$
χ^2Test	**Pass**			

solution, X_i, are formulated as follows:

$$X_i = \{x_{i1}, x_{i2}, x_{i3}, ..., x_{iD}\} \tag{19}$$

where $d = \{1,2,3,...,D\}$ is the number of parameters to be estimated. A number of NP solutions are used. The vectors of each solution are initiated randomly within the search space as the following equation:

$$X_i = X_i^L + rand_3 \times (X_i^U - X_i^L) \tag{20}$$

where $rand_3$ is a uniformly distributed random value between 0 to 1, while X_i^L and X_i^U are the predefined lower and upper bound values, respectively. The fitness value of each solution is evaluated. Based on the value, the solution with best fitness value among the population is selected as the current global best solution, X_{best}.

The S-CRO method incorporates initial selection step, in which the population is sorted based on the fitness values. Then, this sorted population is divided into two major sub-populations. The first sub-population, X^{sub1}, contains a set of solutions that generate potential fitness values whereas the other sub-population, X^{sub2}, consists of solutions that hold least substantial fitness values [29],[30]. The solutions in the first sub-population are submitted for neighbouring improvement step. In this step, the fitness value of the ith solution is compared with its neighbouring solutions. If the value of the jth neighbouring solution is better than the ith solution, the distance of these solution, r_{ij}, is computed. Then, the attractiveness value, β, is calculated. According to this value, the

vectors of the ith solution are updated using equation (6). Next, the ith solution is subjected for evolutionary combinatorial step. This is performed by applying the evolutionary operations adopted from the CRO method. Firstly, a random number is generated and if the value is less than 0.5, the ith solution is submitted for the on-wall collision (if the value is less or equal to 0.33) or decomposition processes (if the value is greater than 0.33) [32]. Otherwise, if the random value is higher than 0.5, the inter-molecule collision (if the value is greater than 0.7) or synthesis processes (if the value is less or equal to 0.7) are executed into the solution [32]. Since these processes involve two solutions, the ith solution and another randomly chosen neighbouring solution are used.

Conversely, the solutions in the second sub-population, X^{sub2}, are subjected for random update step. This is performed to ensure that the fitness values of these solutions are improved for the next iterations. Moreover, this step is also implemented to permit the method to escape the local optima more efficiently. The random update is executed using the following equation:

$$X_{i(t+1)}^{sub2} = X_{best(t)} + rand_4 \times (X_{i(t)}^{sub2} - X_{best(t)}) \tag{21}$$

where $rand_4$ is a random value between 0 and 1 [30]. Different to our work in [29], this step requires vectors of the current best solution to assist the randomization process so that the newly formed weak solutions may consist of potential vectors that will produce better fitness for the next iterations. After this step, the first and second sub-populations are merged to form the updated population. The steps are repeated until the maximum number of

Table 4. Model selection validation (15% white Gaussian noise).

Model		RNA activator	RNA inhibitor	ON-state switch Sw21	ON-state switch Sw12
Real Variance (ξ_n)		9.28×10^{-2}	8.04×10^{-2}	1.33×10^{-1}	1.28×10^{-2}
Z_1	Point ($\hat{\sigma}_n^2$)	9.26×10^{-2}	8.04×10^{-2}	1.32×10^{-1}	1.27×10^{-2}
	Interval	$[8.43 \times 10^{-2}, 1.02 \times 10^{-1}]$	$[7.32 \times 10^{-2}, 8.89 \times 10^{-2}]$	$[1.21 \times 10^{-1}, 1.47 \times 10^{-1}]$	$[1.25 \times 10^{-2}, 1.37 \times 10^{-2}]$
	AIC	$\mathbf{-2.84 \times 10^4}$			
	χ^2Test	**Pass**			
Z_2	Point ($\hat{\sigma}_n^2$)	6.61×10^{-1}	5.11×10^{-1}	2.56×10^{-1}	9.29×10^{-2}
	Interval	$[6.01 \times 10^{-1}, 7.30 \times 10^{-1}]$	$[4.64 \times 10^{-1}, 5.64 \times 10^{-1}]$	$[2.33 \times 10^{-1}, 2.83 \times 10^{-1}]$	$[8.45 \times 10^{-2}, 1.03 \times 10^{-1}]$
	AIC	-2.26×10^4			
	χ^2 Test	Fail			

Table 5. Parameter values of extracellular protease production model.

Parameter	Value	Lower Boundary	Upper Boundary
$k_{syn}(\text{s}^{-1})$	0.04	0.02	0.06
$k_{deg}(\text{s}^{-1})$	0.0004	0.0002	0.0006
$k_{dephos}(\text{s}^{-1})$	0.15	0.05	0.30
$k_{syn1}(\text{s}^{-1})$	0.04	0.02	0.06
$k_{phos}(\text{s}^{-1})$	0.004	0.002	0.006
$k_a(\text{s}^{-1})$	0.025	0.010	0.035
$k_d(\text{s}^{-1})$	0.1	0.05	0.20
$k_{r1}(\text{s}^{-1})$	7	5	9
$I_{ro}(\#/\text{s})$	0.02	0.005	0.04
$k_{Dim}(\text{s}^{-1})$	12	10	14
$k_r(\text{s}^{-1})$	7	5	9
$k_{degm}(\text{s}^{-1})$	0.0099	0.0001	0.0100
$I_{rMax}(\#/\text{s})$	0.4	0.2	0.6
I_o	0.004	0.002	0.006
$I_{Max}(\#/\text{s})$	0.048	0.02	0.06
$R(\text{s}^{-1})$	7	5	9
$k_b(\text{s}^{-1})$	7	5	9

A cell volume is assumed to be 10^{-15} litre and the copy numbers (#) is used as a unit for effective concentration. 602 # corresponds to 1 μM/l concentration [28].
doi:10.1371/journal.pone.0061258.t005

iterations is reached. Figure 2 illustrates the algorithm of the proposed S-CRO method.

Identifiability Analysis

To demonstrate the effectiveness of the proposed method in estimating accurate and reliable parameters, a statistical analysis based on the error variance of the random variables of noise is used [6], [10], [30]. Suppose a model is represented as follows:

$$\begin{cases} \dfrac{ds}{dt} = g(s(u,X),\tau) \\ s_{(\tau_0)} = s_{(0)} \\ y_n = h_n(s(u,X),\tau) + \varepsilon_n \end{cases} \quad (22)$$

where $n = \{1,2,3,...,N\}$ is the number of samples. Thus, the measurement noise is obtained using the following equation:

$$\varepsilon_n = y_n - h_n(s(u,X),\tau) \quad (23)$$

By executing the methods, an estimated parameter set, $\hat{X}$, is found. Hence, if $\hat{X}$ is near to X, then $h_n(s(u,\hat{X}),\tau)$ is close to the output $h_n(s(u,X),\tau)$, then the variance of $\hat{\varepsilon}_n$ is supposedly close to the variance of ε_n. Let $\hat{\sigma}_n^2$ be the variance of $\hat{\varepsilon}_n$. The point estimate of variance $\hat{\sigma}_n^2$ is computed as follows:

$$\hat{\sigma}_n^2 \approx \xi_n = \frac{1}{2}\sum_{n=1}^{N}(\hat{\varepsilon}_n)^2 \quad (24)$$

Subsequently, the interval estimates of $\hat{\sigma}_n^2$ is corresponded to the confident level of $\gamma = 1-\delta$ [6], [10], [30], which is formed using the following equation

$$\frac{N\xi_n}{\chi_{N,1-\frac{\delta}{2}}} \le \hat{\sigma}_n^2 \le \frac{N\xi_n}{\chi_{N,\frac{\delta}{2}}} \quad (25)$$

with a confidence level of $100\gamma\%$. In other words, if the real variance σ_n^2 is not lie within these intervals, then the model output y_n could not have been generated by the estimated parameter set, $\hat{X}$. Therefore, the parameter set $\hat{X}$ is considered as not plausible for the given experimental data with a confidence level of $100\gamma\%$ [30]. In this paper, a significance level, δ, of 0.05 is set, in which giving the confidence level of 95% [6], [10], [30].

Model Selection

Due to the various uncertainties of the experimental environment, it is important to choose a plausible model that may perform consistent predictions according to the given experimental data [6], [37]. In this paper, the model selection is conducted to assess which environment is more feasible to fit the model prediction. The validation is performed using two approaches. The first approach is suggested by [6] and [10], which is presented in the previous sub-section. Let two distinctive models of the form in Equation (1), which are constructed as follows:

$$Z_1: \begin{cases} \dfrac{ds_{z_1}}{dt} = g_{z_1}\left(s_{z_1}(u,X),\tau\right) \\ s_{z_1(\tau_0)} = s_{z_1(0)} \\ y = h_{z_1}\left(s_{z_1}(u,X),\tau\right) + \varepsilon_1 \end{cases} \quad (26)$$

$$Z_2: \begin{cases} \dfrac{ds_{z_2}}{dt} = g_{z_2}\left(s_{z_2}(u,X),\tau\right) \\ s_{z_2(\tau_0)} = s_{z_2(0)} \\ y = h_{z_2}\left(s_{z_2}(u,X),\tau\right) + \varepsilon_2 \end{cases} \quad (27)$$

For both models, the same experimental data are used [6], [10], [30]. Later, the variance points and intervals are computed using these data.

The second approach is applied from [30] and [36], in which the Akaike Information Criterion (AIC) is employed. The AIC validation test is calculated using the following equation:

$$AIC = N\ln\left(\frac{MSE}{N}\right) + 2D \quad (28)$$

where MSE is the mean squared error that generated from the best fitness value, while N and D are the number of samples and estimated parameters, respectively [30]. Generally, this equation

Table 6. Performance comparison of DE, FA, CRO, and S-CRO methods.

	DE	FA	CRO	S-CRO
Noise: 5%				
Average Fitness Value	1.89×10^{-5}	4.14×10^{-3}	8.73×10^{-5}	7.21×10^{-7}
Standard Deviation	2.11×10^{-5}	4.99×10^{-3}	5.67×10^{-5}	1.15×10^{-7}
Computational Time (s)	109.3	181.9	120.9	93.4
Noise: 10%				
Average Fitness Value	1.67×10^{-2}	1.55×10^{-1}	9.34×10^{-2}	4.77×10^{-5}
Standard Deviation	1.82×10^{-2}	1.98×10^{-1}	8.71×10^{-2}	3.99×10^{-5}
Computational Time (s)	161.5	256.6	183.9	107.4
Noise: 15%				
Average Fitness Value	5.12×10^{1}	9.81×10^{2}	7.87×10^{1}	8.11×10^{-2}
Standard Deviation	8.78×10^{1}	8.22×10^{2}	7.93×10^{1}	5.52×10^{-2}
Computational Time (s)	243.9	351.3	279.92	150.5

doi:10.1371/journal.pone.0061258.t006

implies that the model that has smaller AIC value is considered as the best model [36].

Results

Simulated Nonlinear Model

The proposed S-CRO method was firstly evaluated using the simulated nonlinear model [33]. This was important to show the effectiveness of the proposed method in finding the accurate parameters. The time series data was generated based on the following discrete equations:

$$A_{(\tau+1)} = k_1 A_{(\tau)} + A_{(\tau)} B_{(\tau)} + 0.1 + 0.01 \varepsilon_{1(\tau)} \tag{29}$$

$$B_{(\tau+1)} = k_2 B_{(\tau)} + A_{(\tau)} B_{(\tau)} + 0.1 + 0.01 \varepsilon_{2(\tau)} \tag{30}$$

$$y_{(\tau)} = B_{(\tau)} + 0.01 \varepsilon_{3(\tau)} \tag{31}$$

where the values of parameter k_1 and k_2 are 0.8 and 1.5, respectively, while ε_1, ε_2, and ε_3 are the independent zero mean noise [16], [33]. The time series data was produced by running the simulation of this model for 800 time points. The upper and lower boundaries of the parameter k_1 and k_2 were set as follow: $0.1 \leq k_1 \leq 1.0$ and $1.0 \leq k_1 \leq 2.0$. The initial attraction, β_0, and the light absorption coefficient, μ, were fixed to 0.5 and 0.01, respectively. The proposed S-CRO method was executed with 50 iterations. Figure 3 illustrates the results of the estimation by using the S-CRO method. These results showed that the proposed method was capable to accurately estimate the parameters within a relatively small number of iterations.

Small Scale Model: Synthetic Transcriptional Oscillators

The performance of the proposed method for estimating the parameters in biological models is first evaluated using a synthetic model of transcriptional oscillators [34]. This model is basically a cell-free model, in which the prediction can be studied without the prior knowledge of the existing *in vivo* experiments [34], [38]. The model is proposed to simulate the complex networks of the

regulatory perturbation of *deoxyribonucleic acid* (DNA) templates. The model is used to fit in the arbitrary synthetic circuits in a modular fashion [34]. The model is constructed based on the following reactions:

$$\frac{dA}{dt} = k_1 C + k_2 D - A \tag{32}$$

$$\frac{dB}{dt} = k_3 D - B \tag{33}$$

$$\frac{dC}{dt} = \frac{1}{1^{k_4} + B^{k_4}} + C \tag{34}$$

$$\frac{dD}{dt} = \frac{A^{k_5}}{1^{k_5} + A^{k_5}} + D \tag{35}$$

where A and B are the ratios of the RNA activator and inhibitor, respectively, meanwhile C and D are the fractions of ON-state switch Sw21 and switch Sw12, respectively [34]. The values of the parameters $k_1, k_2, k_3, k_4,$ and k_5, are 0.57 μM, 1.5, 2.5 μM, 6.5, and 6.5, respectively [34]. The model is downloaded from Biomodels database [39].

Table 1 presents the performance comparison of the proposed S-CRO method over the standard DE, FA, and CRO methods in terms of average best fitness values and the efficiency of these methods in utilizing the computational cost. In this paper, the experimental data is generated *in silico*, in which the model predictions are added with 5%, 10%, and 15% of white Gaussian noise [6], [10]. The methods were executed with 100 independent runs in a same workstation powered by Intel Core i5 1.5 GHz of central processing unit (CPU) and 4.0 GB of memory using a 64-bit platform. Each method used 20 solutions and 100 iterations. For the DE method, the mutation and crossover coefficient were set to 2.5 and 1.5, respectively. For the FA and S-CRO method, the initial attraction, β_0, and the light absorption coefficient, μ, were fixed to 0.5 and 0.01, respectively. The lower and upper
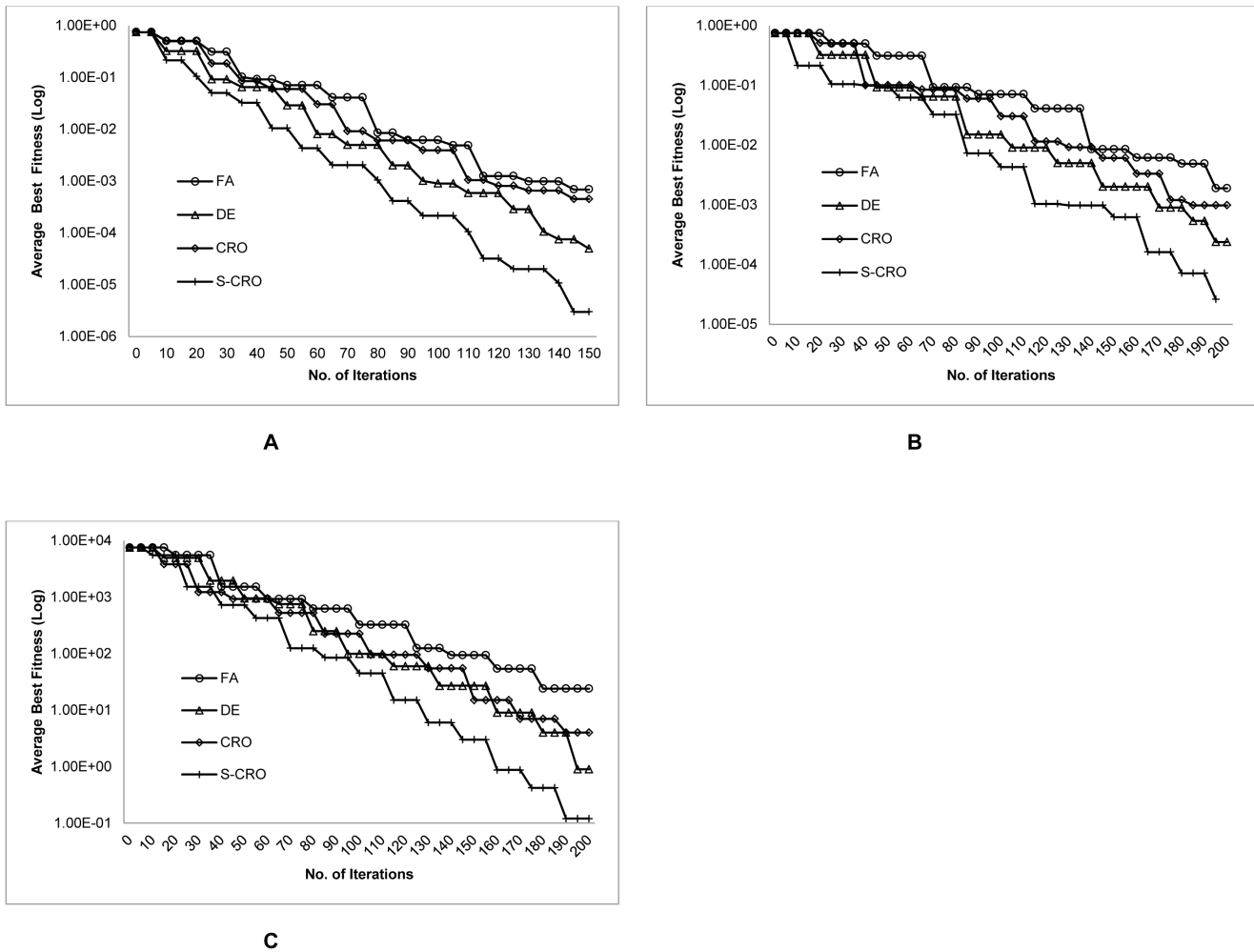
A

B

C

Figure 6. Convergence behaviours of the DE, FA, CRO, and S-CRO methods for the extracellular protease production model. The plots show the average best fitness values of DE, FA, CRO, and the proposed methods in each iteration. Graph A, B, and C represents the convergence behaviours for 5%, 10%, and 15% measurement noise, respectively.
doi:10.1371/journal.pone.0061258.g006

boundaries of the parameters were set as follows: $0.1 \leq k_1 \leq 1.0$, $1.0 \leq k_2 \leq 2.0$, $1.5 \leq k_3 \leq 3.0$, $5.0 \leq k_4 \leq 7.0$, and $5.0 \leq k_5 \leq 7.0$.

The results exposed that the proposed S-CRO method was capable to find better average fitness values compared to the other methods. In addition, the small number of standard deviation suggested that the fitness values found by the S-CRO method were also consistent for the independent runs. More importantly, the results showed that the proposed method evaluated within an acceptably small amount of computational time. This supports the evidence that the evolutionary operations incorporated with the swarm-based search strategy applied by the S-CRO method could utilize the computational cost more effectively than the other methods. Furthermore, Figure 4 shows the convergence behaviours of the methods in finding the average best fitness values. Initially, the DE and S-CRO methods presented a competitive achievement but as the iterations progressed, the proposed method began showing its advantage due to its capability in converging more frequently. This suggests that the random update step implemented by the S-CRO method was effective to permit the method for escaping the sub-optimal solutions.

To demonstrate the effectiveness of the method in estimating the plausible parameters using the noisy and incomplete exper-

imental measurements, the model outputs produced by the estimated parameters were compared with those produced by the actual parameters and the experimental measurements. Figure 5 shows that the outputs produced by the reconstructed model were close to those produced by the actual parameters. This shows that the proposed S-CRO method was robust to the noisy and incomplete experimental data. Moreover, Table 2 shows the estimated parameters by the S-CRO method compare to the other methods using noisy and incomplete experimental data. To address the reliability of the S-CRO method, the statistical analysis for non-identifiable parameters is presented. The results are shown in Table 3. According to the analysis, the real variance errors for the RNA activator and inhibitor, as well as ON-state switch Sw21 and Sw12 were 9.28×10^{-2}, 8.04×10^{-2}, 1.33×10^{-1}, and 1.28×10^{-2}, respectively. Overall, the variance points computed using the outputs of the reconstructed model were close to the real variance. Prominently, these variance points lay within the variance intervals, which suggest that the model outputs were valid based on the given noisy and incomplete experimental data with 95% confidence level.

The S-CRO method was also verified for estimating plausible parameters between two different models. To elucidate this
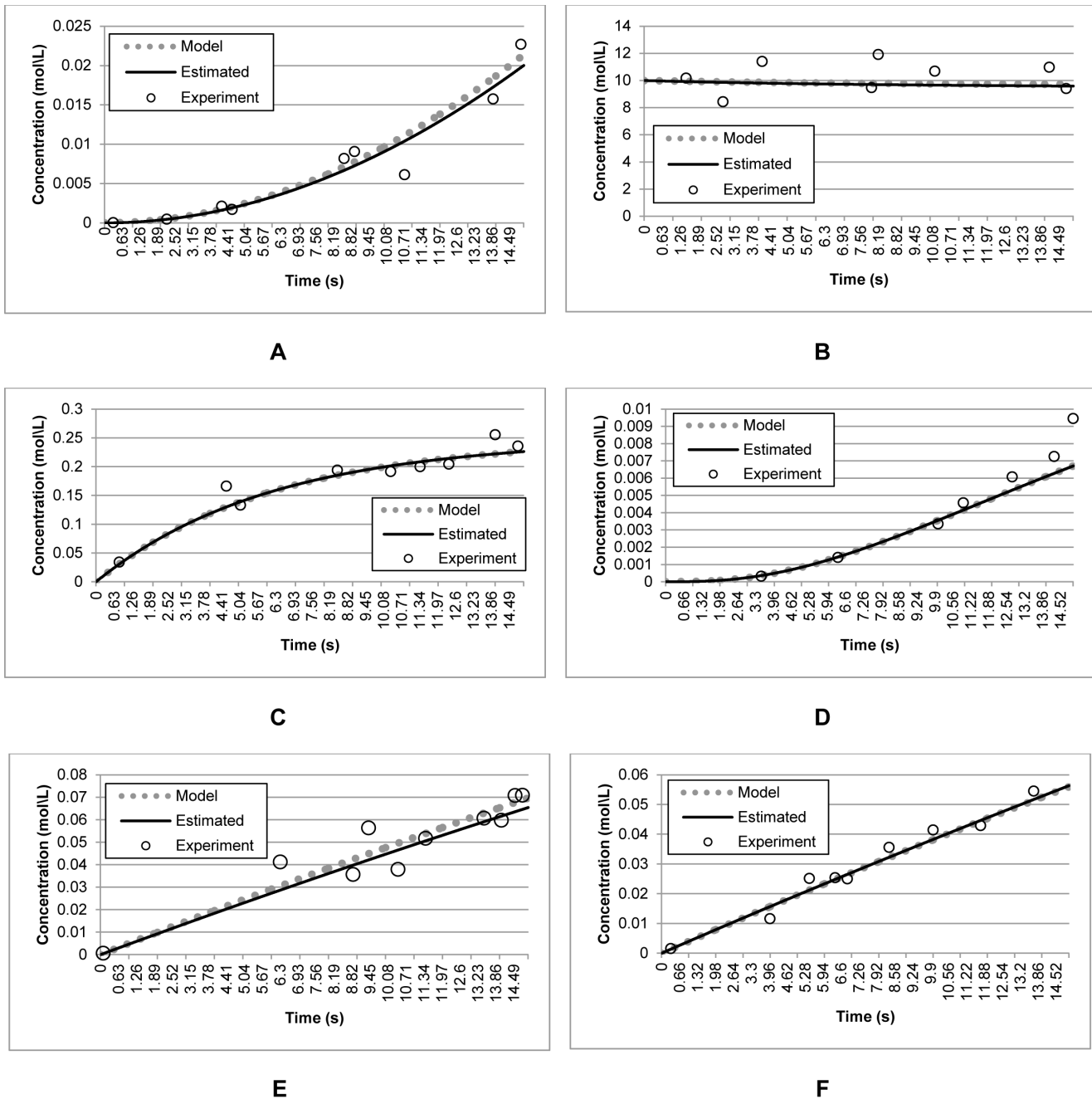
Figure 7. Data fit of model outputs produced by the estimated parameters and the corresponding experimental measurements the extracellular protease production model. The data points (circles) represent synthetic measurements obtained by adding Gaussian noise to the model prediction (dotted line). The straight lines represent the reconstructed model using the parameters estimated by the proposed S-CRO method. Graph A, B, C, D, E, and F represents concentrations of AprE, DegU, DegUP, Dim, mAprE, and mDegU, respectively.
doi:10.1371/journal.pone.0061258.g007

capability, the model defined in Equation 32–35 was modified by changing the values of k_1, and k_3 parameters to zero. Hence, the original and modified models were named as Z_1 and Z_2, respectively. Table 4 presents the results that compare these models based on the same experimental data. Based on this table, the error variance points computed in the modified model were mostly differed from the real variance points. Moreover, the real variance points did not lie within the calculated variance intervals. Additionally, the validation test showed that the AIC values of the model Z_1 were smaller than the model Z_2. This proved that the

proposed S-CRO method was able to discriminate the parameters of these two different models using the same experimental data.

Large Scale Model: Extracellular Protease Production

Naturally, bacterial cells like *Bacillus subtilis* are capable to produce their own nutrient and converge to the steady growth phase by implementing several adaptation strategies. The most substantial strategy used by these bacteria is large scale extracellular protease secretion [35]. Commonly, this process is performed by *subtilisin* (AprE) and *bacillopeptidase* (bpr) genes that encode the

Table 7. Estimated parameters by DE, FA, CRO, and S-CRO methods using the noisy and incomplete experimental data (15% white Gaussian noise).

Parameter	Value	DE	FA	CRO	S-CRO
k_{syn} (s^{-1})	0.04	0.047	0.031	0.042	0.04
k_{deg} (s^{-1})	0.0004	0.00032	0.00051	0.00049	0.00039
k_{dephos} (s^{-1})	0.15	0.22	0.13	0.25	0.149
k_{syn1} (s^{-1})	0.04	0.051	0.025	0.058	0.04
k_{phos} (s^{-1})	0.004	0.0041	0.0052	0.0038	0.004
k_a (s^{-1})	0.025	0.0249	0.0245	0.023	0.0249
k_d (s^{-1})	0.1	0.099	0.095	0.122	0.099
k_{r1} (s^{-1})	7	6.51	7.01	7	7
I_{ro} (#/s)	0.02	0.022	0.02	0.015	0.02
k_{Dim} (s^{-1})	12	13.1	12.03	12.8	11.98
k_r (s^{-1})	7	7.01	5.21	6.95	7
k_{degm} (s^{-1})	0.0099	0.0102	0.0085	0.0092	0.0099
I_{rMax} (#/s)	0.4	0.39	0.38	0.39	0.4
I_o	0.004	0.0032	0.0054	0.0036	0.0039
I_{Max} (#/s)	0.048	0.053	0.037	0.052	0.0479
R (s^{-1})	7	6.5	6.89	6.5	7
k_b (s^{-1})	7	6.72	7.21	6.8	7

involved enzymes to secrete and degrade proteins from the environment. These genes are majorly expressed by DegS–DegU two-component system [35]. The DegS sensor protein is needed to phosphorylate the DegU protein so that the AprE gene expression is triggered [35]. The model is given by the following reactions:

$$\frac{dA}{dt} = k_{syn}E - k_{deg}A \tag{36}$$

$$\frac{dB}{dt} = k_{dephos}C - k_{deg}B + k_{syn1}F - k_{phos}B \tag{37}$$

$$\frac{dC}{dt} = k_{phos}C - 2(k_a C^2) + 2(k_d D) - k_{deg}C - k_{dephos}C \tag{38}$$

$$\frac{dD}{dt} = k_a C^2 + k_d D - k_{deg}D \tag{39}$$

$$\frac{dE}{dt} = \frac{k_{r1}}{R + k_{r1}}$$

$$\left(\frac{I_{ro}\frac{D}{k_{Dim}} + 1}{1 + \frac{D}{k_{Dim}} + \frac{D^2}{k_{Dim}^2} + \frac{R}{k_r}} + \frac{I_{rMax}D^2}{k_{Dim}^2\left(1 + \frac{D}{k_{Dim}} + \frac{D^2}{k_{Dim}^2} + \frac{R}{k_r}\right)} \right) \tag{40}$$

$$- k_{degm}E$$

$$\frac{dF}{dt} = \left(\frac{I_o k_b + I_{max}D}{D + k_b} \right) - k_{degm}F \tag{41}$$

Where A, B, C, D, E, and F are the concentrations of AprE, DegU, DegUP, Dim, mAprE, and mDegU, respectively [35]. The model contains 17 parameters and the values of these parameters are listed in Table 5. The model is obtained from Biomodels database [39].

Table 8. Non-identifiability validation (15% white Gaussian noise).

	AprE	DegU	DegUP	Dim	mAprE	mDegU
Real Variance (ξ_n)	1.47×10^{-5}	1.59×10^{1}	4.54×10^{-3}	1.83×10^{-6}	3.01×10^{-4}	1.76×10^{-4}
Variance Point ($\hat{\sigma}_n^2$)	1.47×10^{-5}	1.59×10^{1}	4.55×10^{-3}	1.83×10^{-6}	3.01×10^{-4}	1.76×10^{-4}
Variance Interval	$[1.38 \times 10^{-5},$ $1.58 \times 10^{-5}]$	$[1.48 \times 10^{1}, 1.70 \times 10^{1}]$	$[4.25 \times 10^{-3},$ $4.88 \times 10^{-3}]$	$[1.71 \times 10^{-6},$ $1.96 \times 10^{-6}]$	$[2.81 \times 10^{-4},$ $3.23 \times 10^{-4}]$	$[1.65 \times 10^{-4},$ $1.89 \times 10^{-4}]$
χ^2Test	**Pass**					

Table 9. Model selection validation (15% white Gaussian noise).

Model		AprE	DegU	DegUP	Dim	mAprE	mDegU
Real Variance (ξ_n)		1.47×10^{-5}	1.59×10^{1}	4.54×10^{-3}	1.83×10^{-6}	3.01×10^{-4}	1.76×10^{-4}
Z_1	Point ($\hat{\sigma}_n^2$)	1.47×10^{-5}	1.59×10^{1}	4.55×10^{-3}	1.83×10^{-6}	3.01×10^{-4}	1.76×10^{-4}
	Interval	$[1.38 \times 10^{-5}, 1.58 \times 10^{-5}]$	$[1.48 \times 10^{1}, 1.70 \times 10^{1}]$	$[4.25 \times 10^{-3}, 4.88 \times 10^{-3}]$	$[1.71 \times 10^{-6}, 1.96 \times 10^{-6}]$	$[2.81 \times 10^{-4}, 3.23 \times 10^{-4}]$	$[1.65 \times 10^{-4}, 1.89 \times 10^{-4}]$
	AIC	$\mathbf{-3.21 \times 10^{4}}$					
	χ^2Test	**Pass**					
Z_2	Point ($\hat{\sigma}_n^2$)	1.48×10^{-5}	1.59×10^{1}	9.59×10^{-3}	2.31×10^{-5}	3.01×10^{-4}	1.79×10^{-4}
	Interval	$[1.38 \times 10^{-5}, 1.58 \times 10^{-5}]$	$[1.48 \times 10^{1}, 1.70 \times 10^{1}]$	$[8.97 \times 10^{-3}, 9.62 \times 10^{-3}]$	$[2.16 \times 10^{-5}, 2.48 \times 10^{-5}]$	$[2.81 \times 10^{-4}, 3.23 \times 10^{-4}]$	$[1.67 \times 10^{-4}, 1.92 \times 10^{-4}]$
	AIC	-2.10×10^{4}					
	χ^2Test	Fail					

doi:10.1371/journal.pone.0061258.t009

For this simulation, all competitive methods were run using 50 solutions and 200 iterations. For the DE method, the mutation and crossover coefficient were amplified to 3.0 and 2.5, respectively. For the FA and S-CRO methods, the initial attraction and the light absorption coefficient were altered to 0.7 and 0.05, respectively. Table 6 describes the comparison of performance among the DE, FA, CRO, and S-CRO methods. Again, the S-CRO method had shown better average fitness values while maintaining the achievement of finding these values consistently by having the small number of standard deviation. Similar to the previous model, the experimental data for this model is also generated by adding 5%, 10%, and 15% of white Gaussian noise to the model predictions [6], [10]. Based on these results, it is proven that the evolutionary combinatorial step of the proposed method is practical in handling noisy and incomplete experimental data. Similar to the former simulation, the S-CRO method also presented better computational cost utilization compared to the other methods. This can be seen from the relatively small amount of computational time consumed. It was also suggested that the discrimination strategy employed in the initial selection step might have contributed to this result. This was due to the fact that only a specific number of solutions were considered to be evaluated using the method. The convergence behaviours of the involved methods are presented in Figure 6. According to this figure, it is clearly observed that the proposed method converged to the average best fitness values more rapidly than the other methods. Although the performance of the DE method was quite competitive, the random update step of the S-CRO method allowed the method to escape the sub-optimal solution more frequently.

The capability of the proposed S-CRO method in handling the noisy and incomplete experimental measurements is presented in Figure 7. In general, the parameters that had been estimated by the proposed method might have generated the model outputs which closely fitted with those produced by the actual parameters, even though the noisy and incomplete experimental data were used. Table 7 describes the comparison of the estimated parameters by the proposed method over the existing methods. On the other hand, the results of the statistical analysis employed for validating these parameters are described in Table 8. For this model, the real variance error computed were 1.47×10^{-5}, 1.59×10^{1}, 4.54×10^{-3}, 1.83×10^{-6}, 3.01×10^{-4}, and 1.76×10^{-4} for the concentrations of AprE, DegU, DegUP, Dim, mAprE, and mDegU, respectively. Similar to the results presented in the former experiment, the variance points calculated using the model outputs

produced by the estimated parameters were close to the values of the real variance points. Essentially, these variance points lay within the computed variance intervals. This proved that the model outputs produced by these parameters were valid with 95% confidence level. Therefore, the S-CRO method had been considered as robust to the noisy and incomplete experimental data.

The results of the model selection are shown in Table 9. In this simulation, the model in Equation 36–41 was adjusted by changing the values of three parameters, k_{syn}, k_d, and $k_{\deg m}$ to zero. Similar to the former simulation, the original and modified models were denoted as Z_1 and Z_2, correspondingly. The results presented that the real error variance points of the DegUP, and Dim concentrations did not lie in the computed variance intervals for the modified model. This suggests that the estimated parameters for the modified model might have produced the model outputs that were not valid to the corresponding experimental measurements with 95% confidence level. Moreover, the results also showed that the AIC values of the original model were smaller than the modified model. Again, this simulation showed the effectiveness of the S-CRO method in selecting a plausible model using the given experimental measurements.

Discussion

Computational systems biology plays an important role in understanding the dynamics of biological systems. This is due to the fact that the biological components involved in the systems often interact with each other to perform specific functions. Therefore, the analyses of individual components are restrictive and impractical [1], [2], [3]. However, this study is commonly hampered by the imperfection of the experimental data obtained in the *in vivo* experimental setups [3], [4], [9]. As a result, the investigations of the complex cellular processes are frequently difficult and ineffective [1], [8]. To elucidate this challenge, a computational modelling approach is exploited. This approach focuses on the design and development of computational models to represent the dynamics behaviours of the biological systems. This is performed by constructing mathematical formulation, namely ODEs, to derive the processes over a specific range of times. These models often depend on a set of parameters that represent the physiological properties of the systems, such as the reaction rates and kinetic constants. These parameters are normally unavailable

in the experimental data. Thus, these parameters are rather estimated by fitting the model output with the corresponding experimental data using nonlinear least squares techniques. As the experimental measurements are noisy and incomplete, the estimation of these parameters is usually challenging and often needs the use of practical nonlinear optimization methods [1], [4], [5].

Recent studies have shown a number of optimization methods to estimate the parameters in the biological models. The local optimization methods, especially those that are developed based on the EFK methods, have presented potential achievements in dealing with the experimental measurements [6], [17]. Nevertheless, these methods generally need to be incorporated with the global optimization methods since the EFK methods are only practical to estimate parameters based on the initial values [16]. Due to these limitations, a number of previous works had considered the use of meta-heuristics methods as the methods are generally robust to the measurement noise. Recently, Evolutionary Computation (EC) methods such as GA and DE methods are pondered due to their effectiveness in finding plausible parameters using noisy and incomplete experimental data [1], [5]. Despite of this advantage, the meta-heuristics methods commonly require a significantly huge amount of computational times [1]. This disadvantage often hinders the methods to converge the search for better fitness values frequently. Therefore, hybrid meta-heuristics methods are commonly exploited to overcome these drawbacks [2], [3], [25].

In this paper, a new hybrid optimization method based on the FA and CRO methods is proposed. The new method, called S-CRO method, is developed by incorporating the evolutionary operations adopted from the CRO method to improve the swarm-based search strategy employed by the FA method. The evolutionary operations are often considered practical to handle measurement noise and incompleteness of the experimental data during the estimation of the model parameters [1], [3]. In general, the method is developed to investigate the effectiveness of the new evolutionary strategy, applied using the CRO method into the swarm-based search strategy of the FA method. Thus, this can provide a new approach to handle noisy and incomplete experimental data in the parameter estimation problem. Furthermore, the S-CRO method also introduces a step to rank the population based on the fitness values and divide this population into two sub-populations. This is performed to reduce the computational cost faced by most conventional meta-heuristics methods [3]. The effectiveness of the proposed method, specifically in the parameter estimation problem, was verified by using a simulated nonlinear model, and two biological models: synthetic transcriptional oscillators and extracellular protease production models. The performance of the proposed method was compared with those from the existing DE, FA, and CRO methods. In addition, the proposed S-CRO method was tested for non-identifiability and model selection. These tests were crucial to validate the capability of the proposed method in estimating reliable and identifiable parameters based on the experimental data [6], [9], [17], [30], [36].

The simulation results showed that the proposed method was capable to consistently find better fitness values than the other methods. This provides evidence that the evolutionary operations incorporated with the swarm-based search strategy is practical to handle uncertainty in the experimental data. More importantly, the proposed method also requires an acceptably small amount of computational time. This shows that the initial selection step employed by the method to discriminate the solutions that hold potential fitness values with those that have incompetent fitness

values is indeed practical to reduce the computational time. Also, it was observed that the parameters estimated using the proposed S-CRO method could generate model outputs that are valid according to the experimental data. The results showed that the outputs produced by the reconstructed models fitted well with the outputs from the actual parameters even though noisy and incomplete experimental data were used. Different from the work presented in [30], the present method considered the parameter boundaries before the estimation. By doing this, the estimation of the parameters had been improved especially in a model with substantially large number of parameters. In addition, the statistical analysis based on the error variance points and intervals supported that these outputs were produced by the valid parameters estimated by the using proposed method. In terms of model selection, the results presented that the outputs of the modified models had failed the validation test, which suggest that the method is also capable to estimate plausible parameters based on given experimental measurements.

Inclusively, the proposed S-CRO method had shown prospective achievement on estimating parameters. The proposed searching strategy that incorporates the evolutionary operations adopted from the CRO method had presented its effectiveness in handling the measurement noise and incompleteness of the experimental data. Additionally, the initial selection step employed by the proposed S-CRO method had also shown its prominent potential, especially in term of utilizing the computational time. The simulation results suggested that the proposed method is capable of estimating both small and large numbers of parameters. Due to the achievements in the practical non-identifiability, it is preferable to extend the capability of the proposed method in handling structural non-identifiability problem. This is because the problem often involves advance knowledge on the model structure [7], [9], [38], which can lead to further discoveries in selecting feasible routes of the pathways that are particularly important in the commercialized biotechnology engineering.

Conclusions

In this paper, a new hybrid optimization method is proposed to estimate the parameters of the biological models. The proposed method, S-CRO method, incorporates the evolutionary operations based on the CRO method to enhance the swarm-based searching strategy employed by the FA method. The method is developed to improve the parameter estimation capability of the current optimization methods, especially when noisy and incomplete experimental measurements are involved. The method also utilizes an initial selection step that selects the solutions with feasible fitness values in order to enhance the utilization of computational cost. The effectiveness of the proposed S-CRO method was validated using simulated nonlinear, synthetic transcriptional oscillators, and extracellular protease production models. The simulation results suggested that the proposed method is capable to consistently find better fitness values compared to the other existing methods. Furthermore, the tests also presented that the parameters estimated by using the S-CRO method can produced model outputs that are valid to the corresponding experimental data. Also, the proposed method was tested for non-identifiability and model selection, which showed that the method is capable to estimate reliable parameters and select appropriate models based on the given experimental data.

Acknowledgments

Author Contributions

Conceived and designed the experiments: AA. Performed the experiments: AA SD. Analyzed the data: MSM SA. Contributed reagents/materials/analysis tools: AA. Wrote the paper: AA.

References

1. Sun J, Garibaldi J, Hodgman C (2012) Parameter estimation using meta-heuristics in systems biology: a comprehensive review. IEEE/ACM Trans Comput Biol Bioinform 9: 185–2022.
2. Abdullah A, Deris S, Hashim SZM, Jamil HM (2009) Graph partitioning method for functional module detections of protein interaction network. Int Conf Comput Technol Dev 1: 230–234.
3. Tashkova K, Korosec P, Silc J, Todorovski L, Dzeroski S (2011) Parameter estimation with bio-inspired meta-heuristic optimization: modelling the dynamics of endocytosis. BMC Syst Biol 5: 159–185.
4. Chou I, Voit EO (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci 219: 57–83.
5. Balsa-Canto E, Banga JR, Egea JA, Fernandez-Villaverde A, Hijas-Liste GM (2011) Global optimization in systems biology: stochastic methods and their applications. Adv Syst Biol 736: 409–424.
6. Lillacci G, Khammash M (2010) Parameter estimation and model selection in computational biology. PLoS Comput Biol 6: e1000696.
7. Balsa-Canto E, Alonso AA, Banga JR (2008) Computational procedures for optimal experimental design in biological systems. IET Syst Biol 2: 163–172.
8. Slezak DF, Suarez C, Cecchi GA, Marshall G, Stolovitzky G (2010) When the optimal is not the best: parameter estimation in complex biological models. PLoS ONE 5: e13283.
9. Raue A, Kreutz C, Maiwald T, Klingmuller U, Timmer J (2011) Addressing parameter identifiability by model-based experimentation. IET Syst Biol 5: 120–130.
10. Lillacci G, Khammash M (2012) A distribution-matching method for parameter estimation and model selection in computational biology. Int J Robust Nonlinear Control 22: 1065–1081.
11. Rodriguez-Fernandez M, Egea JA, Banga JR (2006) Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. BMC Bioinformatics 7: 483–501.
12. Yuting Z, Ganesh S (2010) Mathematical modeling: bridging the gap between concept and realization in synthetic biology, J Biomed Biotechnol, doi:10.1155/2010/541609.
13. Lloyd-Smith JO (2007) Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PLoS ONE 2: e180.
14. Ashyraliyev M, Jaeger J, Blom JG (2008) Parameter estimation and determinability analysis applied to Drosophila gap gene circuits. BMC Syst Biol 2: 83–102.
15. Costa PJ (1994) Adaptive model architecture and extended Kalman-Bucy filters, IEEE Trans Aerospace Electron Syst 30: 525–533.
16. Sun X, Jin L, Xiong M (2008) Extended Kalman Filter for estimation of parameters in nonlinear state-space models of biochemical networks. PLoS ONE 3: e3758.
17. Zeng N, Wang Z, Li Y, Du M, Liu X (2012) A Hybrid EKF and Switching PSO Algorithm for Joint State and Parameter Estimation of Lateral Flow Immunoassay Models. IEEE/ACM Trans Comput Biol Bioinform 9: 321–329.
18. Glover F, Laguna M, Marti R (2000) Fundamentals of scatter search and path relinking. Control Cybern 39: 653–684.
19. Kennedy J, Eberhart R (1995) Particle swarm optimization. IEEE Int Conf Neural Netw 4: 1942–1948.
20. Goldberg DE (1988) Genetic algorithms and machine learning. Mach Learn 3: 95–99.
21. Besozzi D, Cazzaniga P, Mauri G, Pescini D, Vanneschi L (2009) A comparison of genetic algorithms and particle swarm optimization for parameter estimation in stochastic biochemical systems. Evol Comput, Mach Learn Data Mining Bioinformatics: 116–127.
22. Tutkun N (2009) Parameter estimation in mathematical models using the real coded genetic algorithms. Expert Syst Appl 36: 3342–3345.
23. Buhry L, Grassia F, Giremus A, Grivel E, Renaud S, et al. (2011) Automated parameter estimation of the hodgkin-huxley model using the differential evolution algorithm: application to neuromimetic analog integrated circuits. Neural Comput 23: 2599–2625.
24. Storn R, Price K (1997), Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim, 11: 341–359.
25. Balsa-Canto E, Peifer M, Banga JR, Timmer J, Fleck C (2008) Hybrid optimization method with general switching strategy for parameter estimation. BMC Syst Biol 2:26.
26. Rodriguez-Fernandez M, Mendes P, Banga JR (2006) A hybrid approach for efficient and robust parameter estimation in biochemical pathway. Biosyst 83: 248–265.
27. Runarsson TP, Yao X (2000) Stochastic ranking for constrained evolutionary optimization. IEEE Trans Evol Comput 4: 284–294.
28. Chen YT, Wang FS (2012) Determination of kinetic parameters for enzymatic cellulose hydrolysis using hybrid differential evolution. Int J Chem Reactor Eng, 9: 1–16.
29. Abdullah A, Deris S, Mohamad MS, Hashim SZM (2013) A new particle swarm evolutionary optimization for parameter estimation of biological models. Int J Comput Inf Syst Ind Manag Appl 5: 571–580.
30. Abdullah A, Deris S, Anwar S, Arjunan SNV (2013) An Evolutionary Firefly Algorithm for the Estimation of Nonlinear Biological Model Parameters. PLoS ONE 8: e56310.
31. Yang XS (2009) Firefly algorithms for multimodal optimization. Stoch Algorithm Found Appl 5792: 169–178.
32. Lam AYS, Li VOK (2010) Chemical-reaction-inspired metaheuristic for optimization. IEEE Trans Evol Comput 14: 381–399.
33. Germani A, Manes C, Palumbo P (2003), Polynomial extended Kalman filtering for discrete time nonlinear stochastic systems. IEEE Conf Decis Control 1: 886–891.
34. Kim J, Winfree E (2011) Synthetic in vitro transcriptional oscillators. Mol Syst Biol 7: 1–15.
35. Veening JW, Igoshin OA, Eijlander RT, Nijland R, Hamoen LW, et al. (2008) Transient heterogeneity in extracellular protease production by Bacillus subtilis. Mol Syst Biol 4: 184–209.
36. Miao H, Dykes C, Demeter LM, Wu H (2009) Differential equation modeling of HIV viral fitness experiments: model identification, model selection, and multimodel inference. Biometrics 65: 292–300.
37. Lages NF, Cordeiro C, Sousa Silva M, Ponces Freire A, Ferreira AEN (2012) Optimization of time-course experiments for kinetic model discrimination. PLoS ONE 7: e32749.
38. Bujara M, Schümperli M, Billerbeck S, Heinemann M, Panke S (2010) Exploiting cell free systems: implementation and de-bugging of a system of biotransformations. Biotechnol Bioeng 106: 376–389.
39. Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, et al. (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res 34: D68.