

**COMPARISON BETWEEN BINARY PARTICLES SWARM OPTIMIZATION (BPSO) AND BINARY ARTIFICIAL BEE COLONY (BABC) FOR NONLINEAR AUTOREGRESSIVE MODEL STRUCTURE SELECTION OF CHAOTIC DATA**

A. Zabidi<sup>1</sup>, I. M. Yassin<sup>1,\*</sup>, N. M. Tahir<sup>1</sup>, Z. I. Rizman<sup>2</sup> and M. Karbasi<sup>3</sup>

<sup>1</sup>Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor  
Malaysia

<sup>2</sup>Faculty of Electrical Engineering, Universiti Teknologi MARA, 23000 Dungun, Terengganu,  
Malaysia

<sup>3</sup>Kulliyah of Information and Communication Technology, International Islamic University  
Malaysia, 50728 Kuala Lumpur, Malaysia

Published online: 10 September 2017

**ABSTRACT**

This paper presents a comparison between the Binary Artificial Bee Colony (BABC) and Binary Particle Swarm Optimization (BPSO) algorithm for structure selection of a Nonlinear Auto-Regressive Model (NAR) of the chaotic Mackey-Glass time-series data. Both stochastic optimization algorithms are swarm-based in nature with BABC mimicking bee colonies and BPSO mimicking the swarming behavior of birds. Recent research has suggested that the ABC algorithm has better solution quality compared to PSO. However, research on whether this advantage applies to the structure selection case in system identification has not been investigated. We conduct extensive tests to determine the convergence performance of both algorithms.

Author Correspondence, e-mail: [ihsan.yassin@gmail.com](mailto:ihsan.yassin@gmail.com)

doi: <http://dx.doi.org/10.4314/jfas.v9i3s.57>



---

It found that the BABC had managed to significantly outperform BPSO in terms of convergence consistency with a slight advantage in terms of solution quality.

**Keywords:** system identification; nonlinear autoregressive moving average; Mackey-glass; structure selection.

## 1. INTRODUCTION

System Identification (SI) is a process of building a mathematical model of a dynamic system from observed input/output data [1-3]. It has been widely use to model observations in various fields [1-2, 4-7]. The SI design process generally involves several steps [1]:

1. Data collection: Collection of dynamic input/output data that represents the characteristics of the system [1].
2. Model selection: Selection of the general model used to represent the system. Among the choices here are Hammerstein, Wiener, Hammerstein-Wiener [8-9], Volterra[10-11], Nonlinear Auto-Regressive Moving Average with Exogeneous Inputs (NARMAX) [68] and its derivatives [12-20].
3. Model estimator: Once a general model has been selected, there are several options of model estimators to construct the model, such as polynomials, Multi-Layer Perceptrons (MLP) [67], Support Vector Regression, Adaptive Neuro-Fuzzy Inference System (ANFIS), etc. [21].
4. Structure selection: The SI model usually depends on its past input/output behavior to predict its future output. The lagged terms are called regressors. In the design process, typically the regressors are selected by an algorithm as not all regressors contribute to model accuracy. This is the subject of investigation of our paper where we compare between the BABC and BPSO algorithms for structure selection of a Nonlinear Auto-Regressive (NAR) model of the Mackey-Glass dataset.
5. Parameter estimation: After the model structure has been determined, the parameters of the model are estimated, typically using some kind of optimization algorithm such as Least Squares (LS) estimation [22-26].

6. Model validation: Finally, after the parameters have been estimated, we need to confirm whether the model represents the original system without any bias. Methods such as One Step Ahead (OSA), residual analysis, correlation tests and histogram analysis are important testing methods for this task [27].

Regarding design step (4), a vital part in SI is designing an input signal (regressor) which involve the selection of delay/lag term to represent the system. The number of lags determine the number of regressor that will be used for input. The number of lags depend on the system that need to be modeled. Although an increasing number of lag may provide many information to model the system, the size of feature may need to consider in designing an efficient model [28-30].

The structure selection stage in SI is intended to reduce the structure size by selecting only the most significant regressors and discarding ones that contribute the least to classification accuracy. The Orthogonal Least Squares (OLS) algorithm [31-32] has been originally used for structure selection purpose in SI [33]. However, the algorithm tends to select incorrect regressor terms when the data is contaminated by noise sequence [34] and it has since been proven that optimization algorithms such as BPSO was able to outperform the OLS in the selection process [33]. However, the consistency of the convergence has not been explored, in which this is a major concern in SI. Our motivation for comparing of BABC and BPSO is that complex or multi dynamic system may require adequate lag terms to represent, thus increasing solution space (the number of candidate solutions) in the process. Since the solution space is extremely large, an inferior optimization algorithm will simply be trapped in local minima thus degrading the solution quality.

The Artificial Bee Colony (ABC) algorithm [35-36] mimics the intelligent behavior of bee colonies in searching for food source (potential solutions) [35-37]. In a bee colony, the bees are divided per their specific roles in the hive. Scout bees explore new areas for food (solutions) and quality, and returns the information back to the hive like a global search process. Onlooker bees receive this information regarding the food sources and goes on to harvest nectar from the sources while looking for other areas around the food source (local search). The ABC algorithm has a unique feature called the limit, where if after a certain

number of iterations with no improvement in results, the search process will be reset to explore new solutions [38-39]. This gives the ability for ABC to escape the local minima and explore new solutions globally when the need arises.

There have been many comparisons made between the performance of BABC and other optimization algorithms such as Genetic Algorithm (GA) [40], Genetic Programming (GP) [40-41], Evolutionary Strategy (ES) [42], Evolutionary Programming (EP) [40, 43], PSO and also Ant Colony Optimization (ACO) [44-45]. In those comparisons, ABC has been proven to outperform the other algorithms in solving multimodal and multidimensional optimization problems with using less control parameters [35-37] and the ability to escape from local minima [38].

## 2. THEORETICAL BACKGROUND

### 2.1. Nonlinear Autoregressive Model (NAR)

The NAR model represents the output behavior of a system based on its past outputs:

$$y(t) = f^d[y(t-1), y(t-2), \dots, y(t-n_y)] + \varepsilon(t) \quad (1)$$

where  $f^d$  is the estimated model,  $y(t-1), y(t-2), \dots, y(t-n_y)$  are lagged input terms and  $\varepsilon(t)$  are the white noise residuals.

The NAR model can be constructed using various methods [46-49, 50-55], although the polynomial approach is the only method that can explicitly define the relationship between the input/output data.

The polynomial representation of the NAR model for a given time series is:

$$y(t) = \sum_{m=1}^{n_p} P_m \theta_m + \varepsilon(t) \quad (2)$$

where  $n_p$  is the number of terms in the polynomial expansion,  $P_m$  is the  $m$ -th regression term with  $P_1 = 1$ , and  $\theta_m$  is the  $m$ -th regression parameter.  $P_m$  is formed by a combination of input, output and residual terms. In matrix form, identification involves the formulation and solution of the LS problem:

$$P\theta + \varepsilon = y \quad (3)$$

where  $P$  is a  $n \times m$  regressor matrix,  $\theta$  is a  $m \times 1$  coefficient vector and  $y$  is the  $n \times 1$  vector of actual observations.  $P$  is arranged such that its columns represent the  $m$  lagged regressors.  $\varepsilon$  is the white noise residuals.

The NAR identification process is done in two steps namely model structure selection and parameter estimation. Structure selection involves selecting which columns in  $P$  that best describes the observations,  $y$ . After a subset of  $P$  has been selected, the parameter estimation step estimates the parameters of the function  $f^d(\blacksquare)$  that gives the best fit for  $y$ .

## 2.2. ABC Algorithm

A bee colony consists of three groups of bees that fulfill specific roles within the colony: scout bees are responsible for searching for potential food sources, while employed bees are tasked with collecting honey from discovered food sources. Onlooker bees stay in the hive and rotate their roles with scout and employed bees as the scout and employed bees relay information through a special dance that indicate the direction and quality of available food sources [56-57]. The ABC algorithm was designed based on the cooperative behavior of natural bees in the swarm.

In the ABC implementation [39, 58], initially scout bees are sent out to scout the problem space. The position of the scout bees follows:

$$\theta_{ij} = \theta_{jmin} + r \cdot (\theta_{jmax} - \theta_{jmin}) \quad (4)$$

where  $\theta_{ij}$  = Position of  $i_{th}$  bees at  $j_{th}$  dimension and  $r$  = A random number.

The scout bees evaluate the fitness of the solution (termed nectar amount), and this information is shared with onlooker bees waiting in the hive. After the initial search, all scout bees now become employed bees. The employed bees go to the food sources (solutions) in its memory and determines the neighboring food sources to evaluate the nectar amount. If the neighboring food source contains a better solution, the new position is kept. Otherwise the old position is maintained [36, 39, 56-57, 59]. The equation to select neighboring food source is given by:

$$x_{ij}(t+1) = \theta_{ij}(t) + \phi(\theta_{ij}(t) - \theta_{kj}(t)) \quad (5)$$

where  $x_i$  = The position of the onlooker bee,  $t$  = The iteration number,  $\theta_k$  = The randomly chosen employed bee,  $j$  = The dimension of the solution and  $\emptyset(\blacksquare)$  = A series of random variable in the range  $[-1, 1]$ .

The info of the new or existing nectar amount then is relayed to awaiting onlookers when the employed bees return. Onlookers bees then select a food source depend on nectar amount relayed. If the nectar amount increases (solution approaching objective), the probability which that food source is selected is higher. The employed bees which carrying high nectar amount will attract onlooker bees toward it food sources position. The probability of onlookers bees selecting best nectar (solution) in the area is depend on:

$$P_i = \frac{F(\theta_i)}{\sum_{k=1}^S F(\theta_k)} \quad (6)$$

where  $P_i$  = The probability of selecting the  $i^{\text{th}}$  employed bee,  $S$  = The number of employed bees,  $\theta_i$  = The position of the  $i^{\text{th}}$  employed bee and  $F(\theta_i)$  = The fitness value.

After selecting potential food source from employed bees, the onlooker bee goes toward the direction and evaluate the neighboring food source. Similar to employed bees, if the neighboring food source contains a better solution, the new position is kept [36, 39, 56-57, 59]. Otherwise, the old position is maintained. The process is repeated between employed and onlooker bees until the food source is finished. Once this happens, scout bees will now be sent to discover new food sources. In ABC, the activation of scout bees is controlled by how many iterations in which no better-quality food sources are discovered [39].

### 2.3. Binary Artificial Bee Colony Algorithm (BABC)

In order to binarize the ABC algorithm, we follow the concept outlined by [60] by representing the bee positions as “probabilities of change” rather than the actual solution. Suppose that the structure selection problem is defined as:

$$F_{reduced} \subseteq F \quad (7)$$

where  $F$  consists of  $n$  columns representing each regressor. To select a feature subset,  $F_{reduced}$  using BABC, a binary string of length  $1 \times n$  is defined, so that each regressor column has a bit assigned to it. The initial value of the binary string can be randomly defined

during initialization. A value of 1 given to the binary string indicates that the column is chosen for the construction of  $F_{reduced}$ , while the value of 0 means that the column is ignored.

In the swarm, each particle carry a  $1 \times m$  vector in solutions  $x_{ij}$ . This vector contains the “probabilities of change” defined earlier. During optimization, the vector elements will be used as a reference to alter the binary string from its initial state. For example, assuming the initial bit state is 0, if the particle vector element is more than 0.5, the binary bit will be changed to 1, otherwise the bit is kept the same.

### 3. METHODOLOGY

All experiments were performed on a personal computer with 3.10GHz Intel Xeon E3-1220 v3 microprocessor and 4GB RAM. The operating system was Linux Mint XFCE version 17.1 with MATLAB 2014a as the development platform. The flowchart for feature selection process is shown Fig.1 and parameter setting for BABC and BPSO for the first test is shown in Table 1. These parameters were selected to test the robustness of both algorithms under different initialization and exploration conditions.

The dataset used is Mackey-Glass (chaotic time-series differential equation) [61-63] (MG). The lag used was reported as 17 in [64]. For the purpose of this experiment, the lag space was expanded to 20.

The dataset was preprocessed prior to the experiment:

1. No magnitude scaling, 50:50 training and testing division ratio using block division method (PP1).
2. Magnitude scaling between -1 and 1, 50:50 training and testing division ratio using block division method (PP2).
3. Magnitude scaling between -1 and 1, 50:50 training and testing division ratio using interleaving division method (PP3).

The regressor matrix was created based on the model order of two. A total of 230 regressor terms were generated. The number of possible combination of regressor is  $2 \times 10^{230}$ .



Fig.1. Optimization process

Table 1. Optimization algorithm parameter setting

	BPSO	BABC
Fitness criterion	AIC, FPE, MDL	AIC, FPE, MDL
Swarm size	10, 20, 30, 40, 50	10, 20, 30, 40, 50
Max Iterations	500, 1000, 1500	500, 1000, 1500
Initial Seed	0, 10000, 20000	0, 10000, 20000
Limit (% from total iteration)	N/A	25, 50, 75, 100

Apart from the first test, an extended test is design to examine the performance between BPSO and BABC comprehensively in Table 2.

Table 2. Optimization algorithm parameter setting (extensive test)

	BPSO	BABC
Fitness criterion	AIC	AIC
Swarm size	50	50
Max Iterations	5000	5000
Initial Seed	0, 10000,....100000	0, 10000,....100000
Limit (% from total iteration)	N/A	20, 40, 60, 80

After the regressor matrix was created, the BABC and BPSO algorithms were used to select the best possible structure guided by the Akaike Information Criterion (AIC), Final Prediction Error (FPE) and Model Descriptor Length (MDL) [12] as the fitness function. Several tests,

namely the One Step Ahead (OSA) prediction, residual plot, correlation tests and residual histogram analysis, were performed to validate the model.

Correlation tests measure the correlation between two time-series sequences at different points in time. They are useful indicators of dependencies and co relatedness between two sequences. Correlation tests are done by shifting the signals at different lags and measuring the correlation coefficients (degree of correlation) between them. Correlation tests are used to validate the model by determining the whiteness of its residuals. A residual sequence exhibits white noise characteristics if tests Eq. (8) to Eq. (12) hold [53]:

$$\theta_{\varepsilon\varepsilon}(\tau) = E[\varepsilon(t - \tau)\varepsilon(t)] = \delta(\tau) \quad (8)$$

$$\theta_{\varepsilon^2\varepsilon^2}(\tau) = E[\varepsilon^2(t - \tau)\varepsilon^2(t)] = \delta(\tau) \quad (9)$$

$$\theta_{y\varepsilon}(\tau) = E[y(t - \tau)\varepsilon(t)] = 0, \forall\tau \quad (10)$$

$$\theta_{y^2\varepsilon}(\tau) = E[(y^2(t - \tau) - \bar{y}^2(\tau))\varepsilon(t)] = 0, \forall\tau \quad (11)$$

$$\theta_{y^2\varepsilon^2}(\tau) = E[(y^2(t - \tau) - \bar{y}^2(\tau))\varepsilon^2(t)] = 0, \forall\tau \quad (12)$$

where  $\theta_{x_1x_2}(\tau)$  = correlation coefficient between signals  $x_1$  and  $x_2$ ,  $E[\blacksquare]$  = mathematical expectation of the correlation function,  $\varepsilon(t)$  = model residuals =  $y(t) - \hat{y}(t)$ ,  $\tau$  = lag space,  $y(t)$  = observed output at time  $t$  and  $\delta(\tau)$  = Kronecker delta defined as:

$$\delta(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & \tau \neq 0 \end{cases} \quad (13)$$

The confidence band reveals the significance of the correlation, and a significantly large correlation is indicated by one or more coefficients lying outside the confidence band. In correlation tests, the 95% confidence band is required because there is a finite amount of data length available [65]. The model is accepted if the correlation coefficients lie within the 95% confidence limits, defined as  $\pm 1.96/n$  with  $n$  is the number of data points in the sequence.

#### 4. RESULTS AND DISCUSSION

---

Table 3 shows the summary of best result acquired using BPSO and BABC using all parameter combinations. In BPSO, the limit parameter is missing as this parameter does not exist in the algorithm.

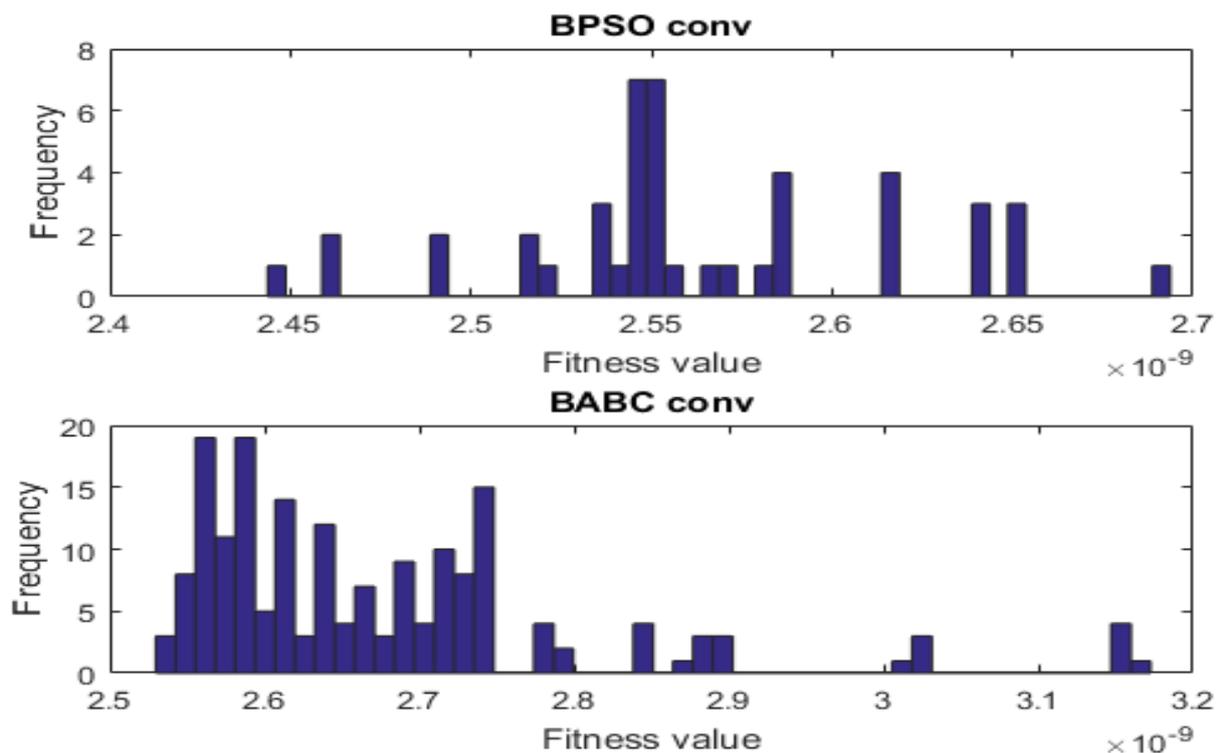
The Akaike Information Criterion (AIC), Model Descriptor Length (MDL) and Final Prediction Error (FPE) was used as the fitness functions to guide the optimization performance. These information criteria simultaneously seek to minimize the error produced by the model while punishing model structures which are too large.

The lowest (best) fitness achieved for BPSO and BABC were  $2.44 \times 10^{-09}$  and  $2.53 \times 10^{-09}$  respectively using the AIC criterion and PP1 preprocessing method (no magnitude scaling and no interleaving). By selecting AIC, the best MSE achieved by BPSO was better than BABC since BPSO produce the lowest MSE in training and testing dataset. The number of training and testing violations during correlation shows that BABC performed better since the number of violation is smaller during training and testing phase. In the selection of terms, the number of regressor selected by BABC was higher than BPSO. Based on these observations, we could not determine any significant difference in performance between BABC and BPSO as the results were almost similar. The convergence of both optimization algorithms is summarized in Table 3.

Fig.2 is a collection of result using AIC and PP1 as constant properties while others were varied (swarm size, max iterations, and initial random seed). For BABC the limit property was added as another property to be tested, but this is absent in BPSO. Therefore, the possible combination of properties for BPSO was 45 and for BABC was 180 (higher due to limit property added into the combination). As can be seen, the convergence distribution pattern was comparable. If examined closely, the distribution of solutions for BABC appears to be more focused, while BPSO appears to be more scattered. The number of possible solutions for this problem (230 regressors) was massive which is  $2 \times 10^{230}$ . From here, we theorized that 1) BABC has some positive effect in terms of convergence and clustering of results. However, 2) as the solution space was large, BABC may require more time to mature and converge.

**Table 3:** Summary of BPSO and BABC best result using AIC, FPE and MDL fitness  
criteria

	BPSO			BABC		
	AIC	FPE	MDL	AIC	FPE	MDL
Magnitude scaling used?	0	0	0	0	0	0
Interleaving used?	0	0	0	0	0	0
Initial Random Seed	20000	0	0	10000	20000	10000
Limit	none	none	none	0.25	0.25	1
No Particles	50	50	50	50	40	50
Max Iterations	1500	1000	1000	1000	1500	1500
Fitness	2.44E-09	2.74E-09	3.50E-09	2.53E-09	2.73E-09	3.48E-09
Number of correlation violations (training set)	33	28	26	29	28	32
Number of correlation violations (testing set)	38	36	32	35	29	32
Training MSE	2.42E-09	2.38E-09	2.73E-09	2.45E-09	2.51E-09	3.02E-09
Testing MSE	2.78E-09	2.90E-09	2.97E-09	2.90E-09	2.82E-09	3.52E-09
No Regressors selected	117	127	101	122	114	91



**Fig.2.** Distribution of results using AIC criteria and PP1 preprocessing method

In the second more extensive test, the maximum number of iterations were increased to 5000, the number of particles was set to 50 and the number of initial random seeds tested was expanded. A summary of the results is shown in Table 4. The best fitness achieved was  $2.43 \times 10^{-9}$  for BABC, while for BPSO was  $2.44 \times 10^{-9}$ . As the results were very near, we considered them as similar. Interestingly, as can be seen from Table 4, BABC managed to find this best solution nine times out of eleven (81.8%) while BPSO managed to only find this solution only two times out of eleven trials (18.2%).

Based on the above observation, we proceed to examine the convergence properties of both algorithms at initial seed 0 (Fig.3). Although the BPSO algorithm was able to quickly find better solutions during the initial part of optimization (less than 4000 iterations), it appears to stall at higher iterations (saturating at above 4000 iterations). However, BABC holds a distinct advantage here as it has the limit mechanism which resets the search process as it saturates. This limit property allows the BABC algorithm to find the solution at a much more consistent rate than BPSO.

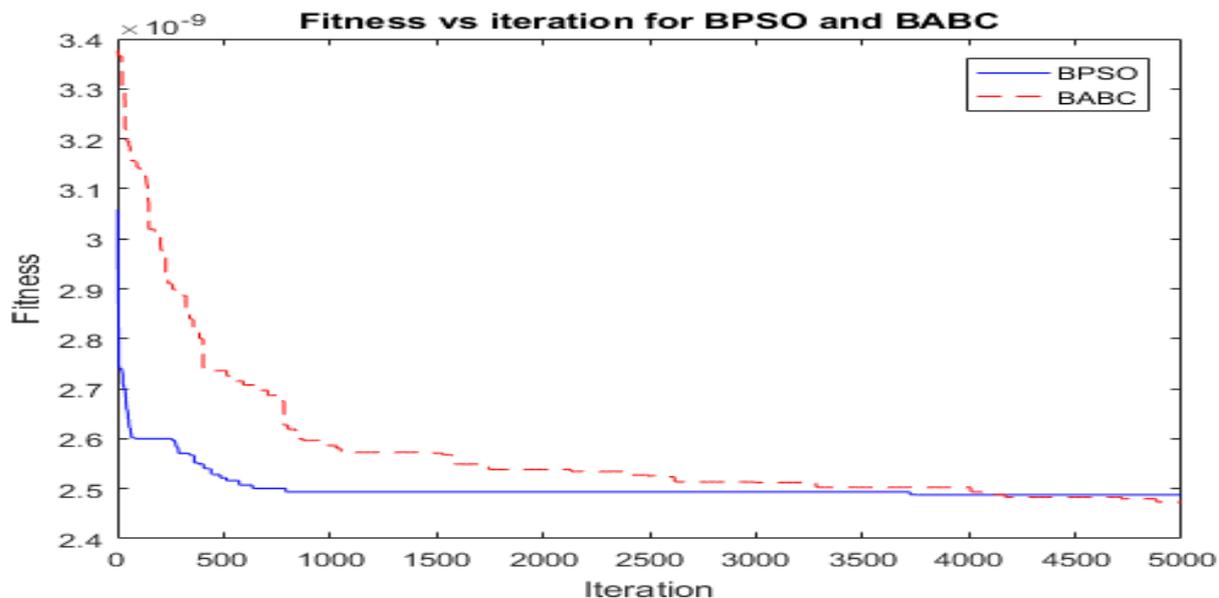
**Table 4.** Fitness result for BPSO and BABC (extensive test)

Seed	BAB			Seed	BAB		
	BPSO Conv	C Limit	BABC Conv		BPSO Conv	C Limit	BABC Conv
0	2.49E-09	0.2	2.49E-09	60000	2.50E-09	0.2	2.46E-09
0		0.4	2.50E-09	60000		0.4	2.47E-09
0		0.6	2.48E-09	60000		0.6	2.47E-09
0		0.8	2.47E-09	60000		0.8	2.47E-09
1000							
0	2.55E-09	0.2	2.48E-09	70000	2.48E-09	0.2	2.48E-09
1000							
0		0.4	2.46E-09	70000		0.4	2.46E-09
1000							
0		0.6	2.47E-09	70000		0.6	2.48E-09
1000							
0		0.8	2.47E-09	70000		0.8	2.48E-09
2000							
0	2.44E-09	0.2	2.46E-09	80000	2.52E-09	0.2	2.48E-09
2000							
0		0.4	2.45E-09	80000		0.4	2.50E-09
2000							
0		0.6	2.45E-09	80000		0.6	2.48E-09
2000							
0		0.8	2.46E-09	80000		0.8	2.48E-09
3000							
0	2.49E-09	0.2	2.44E-09	90000	2.58E-09	0.2	2.49E-09
3000							
0		0.4	2.44E-09	90000		0.4	2.46E-09
3000							
0		0.6	2.44E-09	90000		0.6	2.47E-09

---

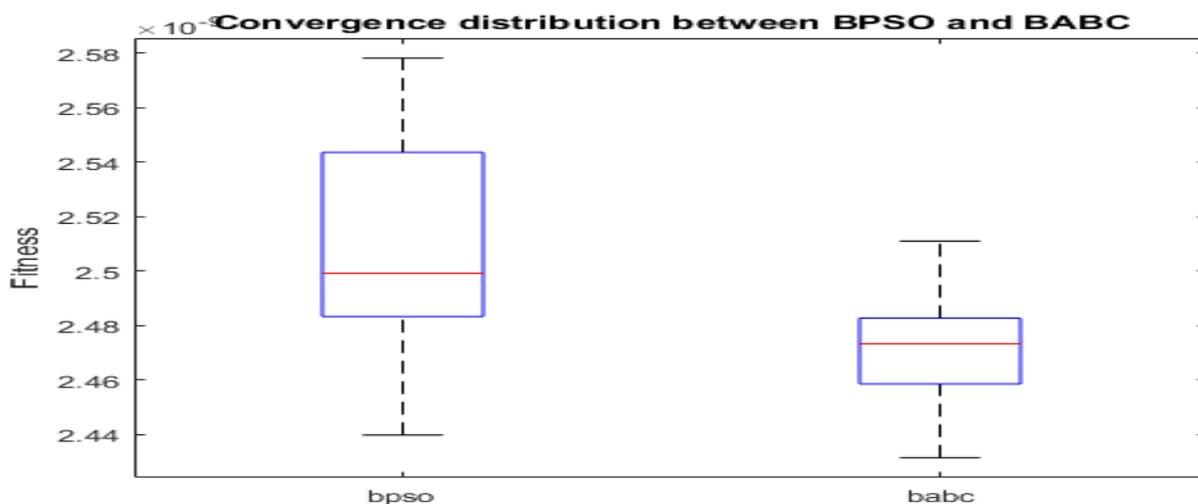
0							
3000							
0		0.8	2.44E-09	90000		0.8	2.47E-09
4000				10000			
0	2.57E-09	0.2	2.48E-09	0	2.50E-09	0.2	2.48E-09
4000				10000			
0		0.4	2.48E-09	0		0.4	2.43E-09
4000				10000			
0		0.6	2.46E-09	0		0.6	2.44E-09
4000				10000			
0		0.8	2.47E-09	0		0.8	2.44E-09
5000							
0	2.48E-09	0.2	2.51E-09				
5000							
0		0.4	2.49E-09				
5000							
0		0.6	2.50E-09				
5000							
0		0.8	2.50E-09				

---



**Fig.3.** Expanded test for BPSO and BABC convergence

By combining all possible combination of properties, the fitness result across all combination of BPSO and BABC were visualized using box plot shown in Fig.4. In the extended test, BABC achieved the lowest number of fitness for maximum and minimum number in box plot scale. Other than that, first quarter, median and third quarter of the box plot scale shows that BABC achieved the lowest fitness with very small variance compared to BPSO. From this, we can conclude that BABC has higher chance in finding lowest possible number of fitness compared to BPSO.

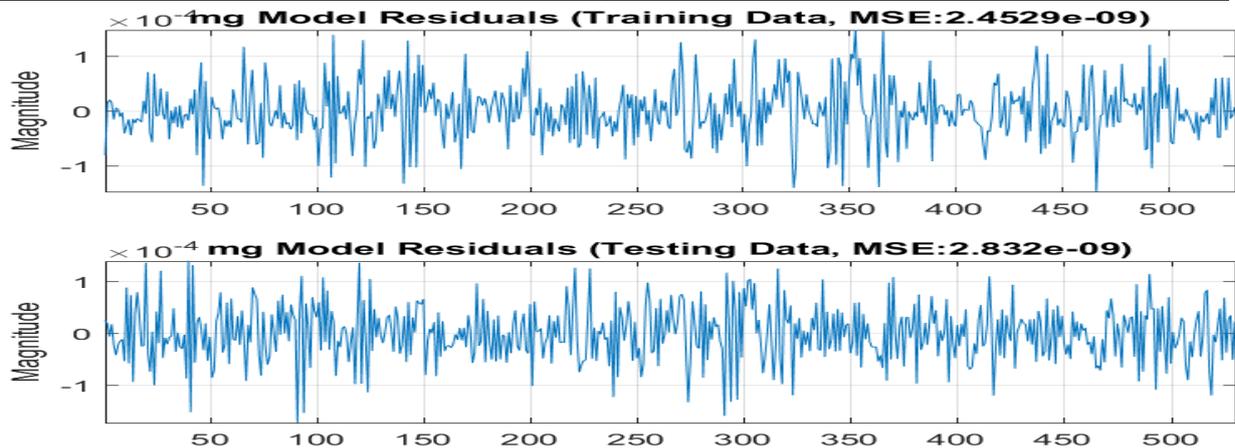
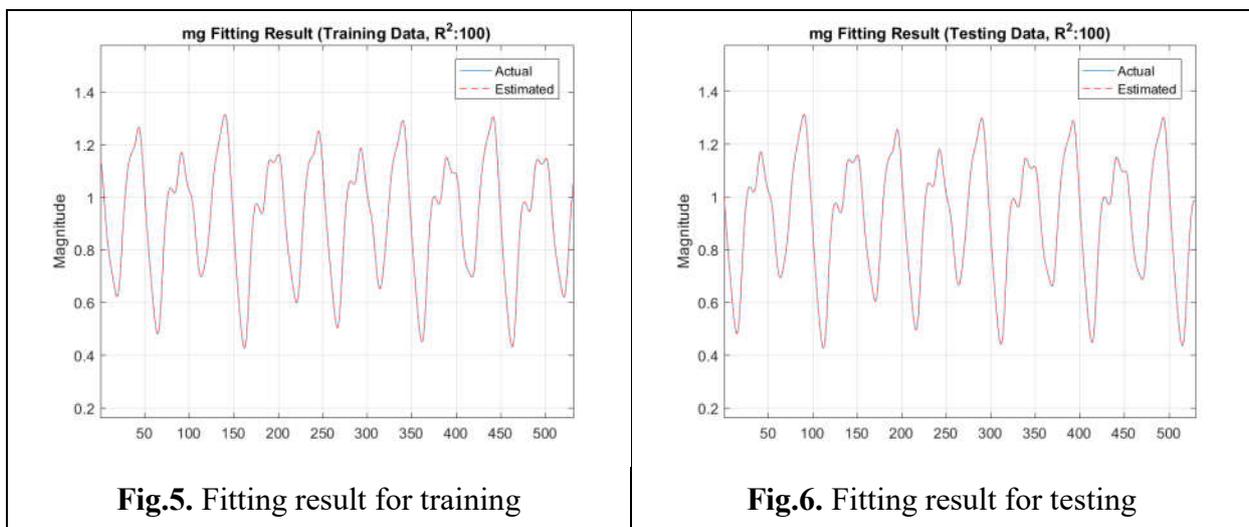


**Fig.4.** Fitness distribution for all possible combination between BPSO and BABC

From this result, the best result achieve from BABC was selected for validation. The optimal results were obtained using BABC with optimization algorithm and AIC as the fitness

criterion with PP1 as the preprocessing method. BABC obtained lowest fitness solution of  $2.43 \times 10^{-9}$ . The training and testing MSE values from the optimal solutions were  $2.45 \times 10^{-9}$  and  $2.83 \times 10^{-9}$  respectively with 113 regressors selected.

The OSA prediction for BABC and BPSO training and testing sets are shown in Fig.5 and Fig.6 respectively. The prediction results show a close fit between the predicted results and original data. This indicates that the model could approximate the dynamics of the original system. This observation is also confirmed based on the small magnitude of the residuals as shown in Fig.7.

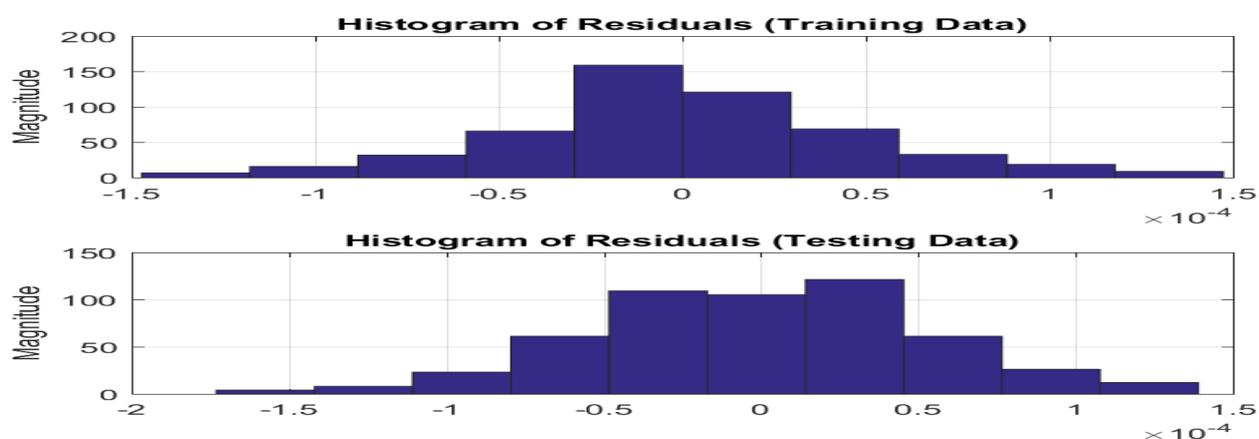


**Fig.7.** Residual between predicted and actual result

Although the residuals and model fit were very good, another important aspect of modeling this result is the model bias. The residuals produced by the model should be random in nature, indicating no bias in predicting the results. To test the whiteness of the residuals, the residuals

were validated using correlation and histogram tests (Fig.8). The correlation test results are shown in Fig.9 to Fig.13.

Based on Fig.9 to Fig.13, a majority of the correlation coefficients reside between the 95% confidence limit while the minority only exceeds the confidence limit by a very small margin except in Fig.8 (5/5). This correlation tests produced 32 training and 32 testing violations. Additionally, the residuals' histogram test results in Fig.8 followed the Gaussian distribution. Both these observations indicate that the residuals are random and uncorrelated, thus exhibiting white noise properties. Based on this, the model is considered as acceptable.



**Fig.8.** Histogram of residuals

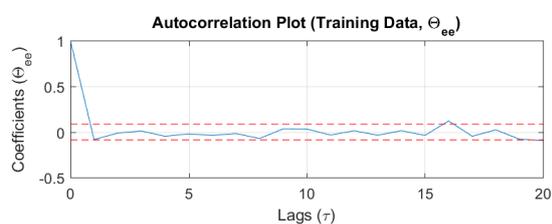
## 5. CONCLUSION

BABC and BPSO could solve feature selection problem in SI. The extended test conducted for BABC and BPSO showed some significant difference between the algorithms.

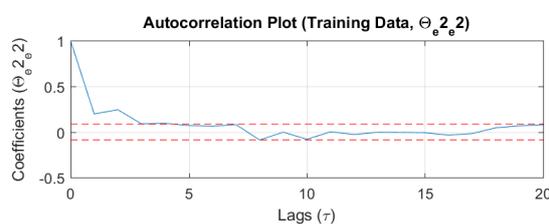
Based on the results, it seems that for a massive solution space, BABC algorithm was able to converge consistently better compared to BPSO if the number of iteration is acceptable. This has been proven with the experiments conducted. By using BABC as feature selection algorithm, the optimal solution achieved was  $2.13 \times 10^{-9}$  by using AIC as fitness criterion with PP1 preprocessing method.

The solution had passed all the necessary tests for it to be considered a valid model. BABC significantly outperformed BPSO by 9 from 11 possible test and BABC still finding the possible lowest fitness value in time BPSO started to stall at certain iteration. This finding

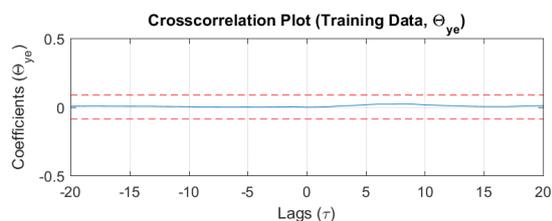
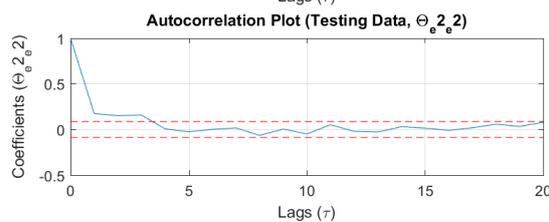
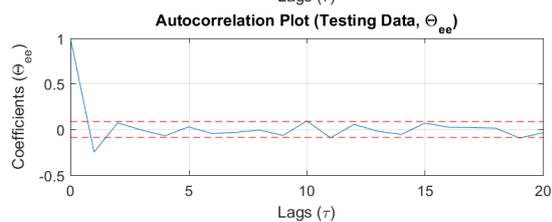
conclude BABC [66] is more consistent in finding optimal solution compared to BPSO in solving feature selection problem for SI.



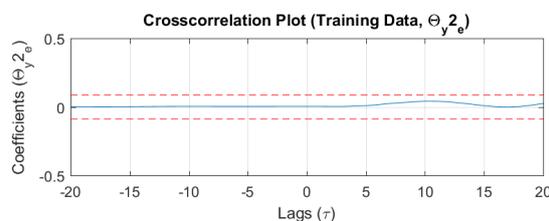
**Fig.9. Auto-correlation plot (1/5)**



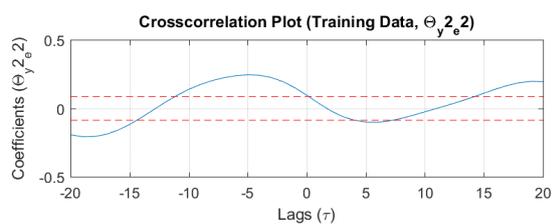
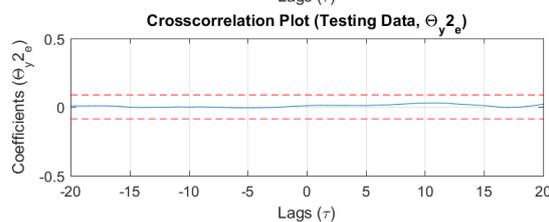
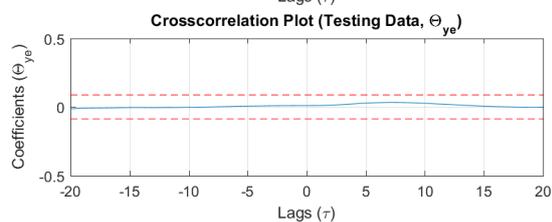
**Fig.10. Auto-correlation plot (2/5)**



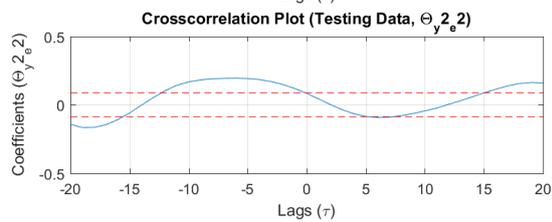
**Fig.11. Cross-correlation plot (3/5)**



**Fig.12. Cross-correlation plot (4/5)**



**Fig.13. Cross-correlation plot (5/5)**



## 6. ACKNOWLEDGEMENTS

The authors would like to graciously acknowledge the Ministry of Higher Education and Universiti Teknologi Mara for supporting this research work through Grant No: FRGS/1/2016/TK04/UITM/03/5.

## 7. REFERENCES

- [1] Fu L, Li P. The research survey of system identification method. In 5th IEEE International Conference on Intelligent Human-Machine Systems and Cybernetics, 2013, pp. 397-401
- [2] Ling TG, Rahmat MF, Husain AR. System identification and control of an electro-hydraulic actuator system. In 8th IEEE International Colloquium on Signal Processing and its Applications, 2012, pp. 85-88.
- [3] Xiuqin X System identification based on particle swarm optimization algorithm. In IEEE International Conference on Computational Intelligence and Security, 2009, pp. 259-263
- [4] Khalil B, Yesildirek A. System identification of UAV under an autopilot trajectory using ARX and Hammerstein-Wiener methods. In 7th IEEE International Symposium on Mechatronics and its Applications, 2010, pp. 1-5
- [5] Ren Z, Zhu GG. Multirate closed-loop system identification of a variable valve timing actuator for an internal combustion engine. In IEEE American Control Conference, 2010, pp. 664-669
- [6] J. Tchorzewski J. Model development of the electrical power system shaped by forecast values of the electricity market using example data of the IEEE RTS test system. In 11th IEEE International Conference on the European Energy Market, 2014, pp. 1-5
- [7] Noshadi A, Shi J, Lee WS, Shi P, Kalam A. Genetic algorithm-based system identification of active magnetic bearing system: A frequency-domain approach. In 11th IEEE International Conference on Control and Automation, 2014, pp. 1281-1286
- [8] Gudupudi LK, Beaugeant C, Evans N. Characterisation and modelling of non-linear loudspeakers. In 14th IEEE International Workshop on Acoustic Signal Enhancement, 2014, pp. 134-138
- [9] Salimifard M, Jafari M, Dehghani M. Hammerstein model identification of multivariable nonlinear systems in the presence of colored noises. In 2nd IEEE International Conference on

---

Control, Instrumentation and Automation, 2011, 2011, pp. 1206-1210

[10] He H, Chen Y, Yang Z, Deng H. Fault diagnosis method of non-linear analog circuits based on Volterra series and SVM. In 26th IEEE Chinese Control and Decision Conference, 2014, pp. 3217-3222

[11] Loghmanian SM, Yusof R, Khalid M. Nonlinear dynamic system identification using Volterra series: Multi-objective optimization approach. In 4th IEEE International Conference on Modeling, Simulation and Applied Optimization, 2011, pp. 1-5

[12] Mendez EM, Billings SA. An alternative solution to the model structure selection problem. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2001, 31(6):597-608

[13] Chiras N, Evans C, Rees D. Nonlinear gas turbine modeling using NARMAX structures. *IEEE Transactions on Instrumentation and Measurement*, 2001, 50(4):893-898

[14] Vallverdu M, Korenberg MJ, Caminal P. Model identification of the neural control of the cardiovascular system using NARMAX models. In *Computers in Cardiology Proceedings*, 1991, pp. 585-588

[15] Kukreja SL, Galiana HL, Kearney RE. NARMAX representation and identification of ankle dynamics. *IEEE Transactions on Biomedical Engineering*, 2003, 50(1):70-81

[16] Kukreja SL, Kearney RE, Galiana HL. A least-squares parameter estimation algorithm for switched Hammerstein systems with applications to the VOR. *IEEE Transactions on Biomedical Engineering*, 2005, 52(3):431-444

[17] Biao L, Lide W, Ping S, Gang L. NDEKF neural network applied to electronically controlled fuel injection system. In 2nd IEEE Conference on Industrial Electronics and Applications, 2007, pp. 351-354

[18] Bai Y, Zhu Y, Jiang Y. A new nonlinear system identification method using gene expression programming. In *IEEE International Conference on Mechatronics and Automation*, 2007, pp. 2951-2956

[19] Li H, Ji G, Ma Z. A nonlinear predictive model based on multilayer perceptron network. In *IEEE International Conference on Automation and Logistics*, 2007, pp. 2686-2690

[20] Sun D, Ye F, Gu X. A fuzzy identification algorithm for NARMAX model and its

---

application to the estimation of oxygen concentration. In 5th IEEE World Congress on Intelligent Control and Automation, 2004, pp. 3455-3457

[21] Billings SA, Coca D. Identification of NARMAX and related models. Research report, South Yorkshire:University of Sheffield, 2001

[22] Yu P, Li J, Peng H. A least square method for parameter estimation of RSC sub-codes of turbo codes. IEEE Communications Letters, 2014, 18(4):644-647

[23] Liu J, Fang J, Wu J, Kang Z, Ning X. Fast non-linearly constrained least square joint estimation of position and velocity for X-ray pulsar-based navigation. IET Radar, Sonar and Navigation, 2014, 8(9):1154-1163

[24] Raphan M, Simoncelli EP. Least squares estimation without priors or supervision. Neural Computation, 2011, 23(2):374-420

[25] Zheng Z, Hao C, Yang X. Least squares channel estimation with noise suppression for OFDM systems. Electronics Letters, 2016, 52(1):37-39

[26] Abdollahi A, Matinfar F. Frequency estimation: A least-squares new approach. IEEE Transactions on Power Delivery, 2011, 26(2):790-798

[27] Yassin IM, Taib MN, Adnan R. Extended analysis of BPSO structure selection of nonlinear auto-regressive model with exogenous inputs (NARX) of direct current motor. Songklanakarin Journal of Science and Technology, 2014, 36(6):683-699

[28] Abdullah SM, Yassin AI, Tahir NM. Particle swarm optimization and least squares estimation of NARMAX. ARPN Journal of Engineering and Applied Sciences, 2015, 10(22): 17139- 17145

[29] Yassin IM. Nonlinear auto-regressive model structure selection using binary particle swarm optimization algorithm. Phd thesis, Selangor: Universiti Teknologi MARA, 2014

[30] Yassin IM, Taib MN, Rahim NA, Salleh MK, Abidin HZ. Particle swarm optimization for NARX structure selection-Application on DC motor model. In IEEE Symposium on Industrial Electronics and Applications, 2010, pp. 456-462

[31] Ahmad RO, Jamaluddin HI. Orthogonal least square algorithm and its application for modelling suspension system. Jurnal Teknologi, 2001, 34(A):71-84

[32] Amisigo BA, Van de Giesen N, Rogers C, Andah WE, Friesen J. Monthly streamflow

---

prediction in the Volta Basin of West Africa: A SISO NARMAX polynomial modelling. *Physics and Chemistry of the Earth, Parts A/B/C*, 2008, 33(1):141-150

[33] Mohamad MS, Yassin IM, Zabidi A, Taib MN, Adnan R. Comparison between PSO and OLS for NARX parameter estimation of a DC motor. In *IEEE Symposium on Industrial Electronics and Applications*, 2013, pp. 27-32

[34] Wei HL, Billings SA. Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information. *International Journal of Modelling, Identification and Control*, 2008, 3(4):341-356

[35] Karaboga D, Akay B. A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*, 2009, 214(1):108-132

[36] Karaboga D, Basturk B. On the performance of artificial bee colony (ABC) algorithm. *Applied Soft Computing*, 2008, 8(1):687-697

[37] Balasubramani K, Marcus K. A comprehensive review of artificial bee colony algorithm. *International Journal of Computers and Technology*, 2013, 5(1):15-28

[38] Javadi MR, Mazlumi K, Jalilvand A. Application of GA, PSO and ABC in optimal design of a stand-alone hybrid system for north-west of Iran. In *7th IEEE International Conference on Electrical and Electronics Engineering*, 2011, pp. 203-210

[39] Dixit GP, Dubey HM, Pandit M, Panigrahi BK. Artificial bee colony optimization for combined economic load and emission dispatch. In *International Conference on Sustainable Energy and Intelligent Systems*, 2011, pp. 340-345

[40] Samuel GG, Rajan CC. Hybrid Particle swarm optimization-Genetic algorithm and particle swarm optimization-evolutionary programming for long-term generation maintenance scheduling. In *IEEE International Conference on Renewable Energy and Sustainable Energy*, 2013, pp. 227-232

[41] Karasek J, Burget R, Povoda L, Dutta MK, Singh A. Genetic programming operators for work-flow optimization in logistic distribution centers. In *IEEE International Conference on Medical Imaging, m-Health and Emerging Communication Systems*, 2014, pp. 105-109

[42] Reinauer V, Magele C, Scheiblich C, Stermecki A, Banucu R, Albert J, Jandl M, Rucker WM. Object-oriented development of an optimization software in Java using evolution

---

strategies. IEEE Transactions on Magnetics, 2012, 48(2):603-606

[43] Regis RG. Evolutionary programming for high-dimensional constrained expensive black-box optimization using radial basis functions. IEEE Transactions on Evolutionary Computation, 2014, 18(3):326-347

[44] Liao T, Socha K, de Oca MA, Stützle T, Dorigo M. Ant colony optimization for mixed-variable optimization problems. IEEE Transactions on Evolutionary Computation, 2014, 18(4):503-518

[45] Juang CF, Hung CW, Hsu CH. Rule-based cooperative continuous ant colony optimization to improve the accuracy of fuzzy system design. IEEE Transactions on Fuzzy Systems, 2014, 22(4):723-735

[46] Anderson SR, Lepora NF, Porrill J, Dean P. Nonlinear dynamic modeling of isometric force production in primate eye muscle. IEEE Transactions on Biomedical Engineering, 2010, 57(7):1554-1567

[47] Amisigo BA, Van de Giesen N, Rogers C, Andah WE, Friesen J. Monthly streamflow prediction in the Volta Basin of West Africa: A SISO NARMAX polynomial modelling. Physics and Chemistry of the Earth, 2007, 33(1):141-150

[48] Chen S, Billings SA, Luo W. Orthogonal least squares methods and their application to non-linear system identification. International Journal of Control, 1989, 50(5):1873-1896

[49] Billings SA, Chen S. Extended model set, global data and threshold model identification of severely non-linear systems. International Journal of Control, 1989, 50(5):1897-1923

[50] Rahim N A. The design of a non-linear autoregressive moving average with exogenous input (NARMAX) for a DC motor. Master thesis, Selangor: Universiti Teknologi MARA, 2004

[51] Rahim NA, Taib MN, Yusof MI. Nonlinear system identification for a DC motor using NARMAX approach. In IEEE Asian Conference on Sensors, 2003, pp. 305-311

[52] Rahiman MH. System identification of essential oil extraction system. Phd thesis, Selangor: Universiti Teknologi MARA, 2008

[53] Nørgård PM, Ravn O, Poulsen NK, Hansen LK. Neural networks for modeling and control of dynamic systems: A practitioner's handbook. London: Springer, 2000

- 
- [54] Zhao F, Hu L, Li Z. Nonlinear system identification based on recurrent wavelet neural network. In 6th International Symposium on Neural Networks, 2009, pp. 517-525
- [55] Billings SA, Wei HL. A new class of wavelet networks for nonlinear system identification. *IEEE Transactions on Neural Networks*, 2005, 16(4):862-874
- [56] Yi Y, He R. A novel artificial bee colony algorithm. In 6th IEEE International Conference on Intelligent Human-Machine Systems and Cybernetics, 2014, pp. 271-274
- [57] Kaswan KS, Choudhary S, Sharma K. Applications of artificial bee colony optimization technique: Survey. In 2nd IEEE International Conference on Computing for Sustainable Global Development, 2015, pp. 1660-1664
- [58] Qian Q, Cai J, Zhang R. Intrusion detection based on neural networks and artificial bee colony algorithm. In 13th IEEE/ACIS International Conference on Computer and Information Science, 2014, pp. 257-262
- [59] Tankasala GR, Reddy G. Artificial bee colony optimization for economic load dispatch of a modern power system. *International Journal of Scientific and Engineering Research*, 2012, 3(1):1-6
- [60] Kennedy J, Eberhart RC. A discrete binary version of the particle swarm algorithm. In IEEE International Conference on Systems, Man, and Cybernetics, 1997, pp. 4104-4108
- [61] Er MJ, Liu F, Li MB. Self-constructing fuzzy neural networks with extended Kalman filter. *International Journal of Fuzzy Systems*, 2010, 12(1):66-72
- [62] Gómez G P, Ramírez C JM, Hernández S E, Alarcón A V. A neural network scheme for long-term forecasting of chaotic time series. *Neural Processing Letters*, 2011, 33(3):215-233
- [63] Yilmaz S, Oysal Y. Fuzzy wavelet neural network models for prediction and identification of dynamical systems. *IEEE Transactions on Neural Networks*, 2010, 21(10):1599-1609
- [64] MathWorks Inc., Wang WC. *Fuzzy logic toolbox for use with MATLAB: User's guide*. Massachusetts: Mathworks Incorporated, 1998
- [65] Mendez EM, Billings SA. An alternative solution to the model structure selection problem. *IEEE Trans. Systems, Man and Cybernetics-Part A: Systems and Humans*, 2001, 31(6):597-608.

- [66] Azlee Z, Nooritawati M T, Ihsan M Y, Zairi I R. The performance of binary artificial bee colony (BABC) in structure selection of polynomial NARX and NARMAX models. *International Journal on Advanced Science, Engineering and Information Technology*, 2017, 7(2):373-379
- [67] Ihsan M Y, Azlee Z, Rozita J, Megat S A M A, Rahimi B, Abu H A H, Zairi I R. Comparison between cascade forward and multi-layer perceptron neural networks for narx functional electrical stimulation (FES)-based muscle model. *International Journal on Advanced Science, Engineering and Information Technology*, 2017, 7(1):215-221
- [68] Ihsan M Y, Azlee Z, Megat S A M A, Nooritawati M T, Hasliza A H, Husna Z A, Zairi I R. Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm. *International Journal on Advanced Science, Engineering and Information Technology*, 2016, 6(5):630-637

**How to cite this article:**

Zabidi A, Yassin I M, Tahir N M, Rizman Z I, Karbasi M. Comparison between binary particles swarm optimization (b PSO) and binary artificial bee colony (BABC) for nonlinear autoregressive model structure selection of chaotic data. *J. Fundam. Appl. Sci.*, 2017, 9(3S), 730-754