The Performance of Binary Artificial Bee Colony (BABC) in Structure Selection of Polynomial NARX and NARMAX Models

Azlee Zabidi, Nooritawati Md Tahir, Ihsan Mohd Yassin, Zairi Ismael Rizman

Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
E-mail: z_lee_82@yahoo.co.uk, pghrribu@salam.uitm.edu.my, ihsan.yassin@gmail.com

Faculty of Electrical Engineering, Universiti Teknologi MARA, 23000 Dungun, Terengganu, Malaysia
E-mail: zairi576@tganu.uitm.edu.my

Abstract—This paper explores the capability of the Binary Artificial Bee Colony (BABC) algorithm for feature selection of Nonlinear Autoregressive Moving Average with Exogenous Inputs (NARMAX) model, and compares its implementation with the Binary Particle Swarm Optimization (BPSO) algorithm. A binarized modification of the BABC algorithm was used to perform structure selection of the NARMAX model on a Flexible Robot Arm (FRA) dataset. The solution quality and convergence were compared with the BPSO optimization algorithm. Fitting and validation tests were performed using the One-Step Ahead (OSA), correlation and histogram tests. BABC was able to outperform BPSO in terms of convergence consistency with equal solution quality. Additionally, it was discovered that BABC was less prone to converge to local minima while BPSO was able to converge faster. Results from this study showed that BABC was better-suited for structure selection in huge dataset and the convergence has been proven to be more consistent relative to BPSO.

Keywords—system identification; binary artificial bee colony; binary particle swarm optimization; nonlinear auto-regressive moving average with exogenous; flexible robot arm

I. INTRODUCTION

SI is a process of building a mathematical model of a system from observed input and/or output data [1]-[3]. Among its applications is simulation, forecasting, controller design, process planning, monitoring, and modelling as well as an explanation of naturally-occurring phenomena in various fields [4]-[7].

Recently, research in non-linear SI has increased as science increasingly discovers complex systems that require these modelling techniques to understand their behavior and to closely approximate their behaviour [8].

A highly efficient and accurate model for modelling nonlinear systems called the NARMAX model was introduced [9] as a unified and accurate representation for a wide variety of nonlinear systems [10]-[18]. The NARMAX model represents its current output as a combination of its past inputs, outputs, and residuals [19], [20]. The residuals provide a feedback mechanism that may further improve the NARMAX model prediction.

The NARMAX model can be constructed using various methods such as polynomials [21]-[24], Multilayer Perceptrons (MLP) [25]-[28] and Wavelet ANNs (WNN) [29], [30]. From all, the polynomial approach is more attractive and versatile because of the simplicity of the model structure, and it reveals the dynamical properties of the system in a straightforward manner [31]. Other than that, it has been proven to work well in practical application with a smaller number of parameters used compared to the Volterra or Wiener series.

NARMAX identification involves three major processes namely structure selection, parameter estimation, and model validation. Of particular interest in this paper is model structure selection that is concerned with selecting a set of the most representative regressors from a large number of candidates. In addition, the selected regressors must obey the principle of parsimony in which the structure must be as small as possible, yet has the ability to represent the system well [32].

This research proposes a binarized modification of the Artificial Bee Colony (ABC) algorithm to perform structure selection of a polynomial NARMAX model on an FRA dataset. The ABC algorithm is an optimization algorithm that mimics the intelligent behavior of bee colonies in finding the best food source around its perimeter by dividing the responsibilities of the swarm and working together to achieve a common goal [33], [34]. 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 discovering 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 indicates the direction and quality of available food sources [35], [36].

The performance of ABC is compared with another swarm-inspired optimization algorithm, the BPSO [37] with the experiment design focusing on solution quality and convergence consistency. This work is motivated by previous works [38], [39] that suggest that the ABC algorithm is capable of outperforming PSO in terms of local and global optimization problems.

II. MATERIAL AND METHOD

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

![Flowchart of experiments](image1)

A. Dataset Description

The dataset used for the structure selection experiments were the FRA dataset. FRA is a single input and single output (SISO) nonlinear system which torque input is effecting the acceleration in operation. The data have been used as a benchmark data for comparing seven different of identification program. The lag used was reported as 4 [40]. The possible combination of the regressor is $2 \times 10^{44}$ from 44 regressors.

![The FRA dataset](image2)

B. NARMAX Structure Selection Using BABC and BPSO

The polynomial representation of the NARX/NARMA/NARMAX model for a given input-output series is:

$$y(t) = \sum_{m=0}^{n_p} P_m \theta_m + \epsilon(t)$$  \hspace{1cm} (1)

where $n_p$ is the number of terms in the polynomial expansion, $P_m$ is the $m$-th regression term with $P_0 = 1$ 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 Least Squares (LS) problem:

$$P\theta + \epsilon = y$$  \hspace{1cm} (2)

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 \( r \) lagged regressors. \( e \) is the white noise residuals.

For feature selection, suppose that the feature selection problem is defined as:

\[
\mathbf{F}_{\text{reduced}} = \mathbf{F} \tag{3}
\]

where \( \mathbf{F} \) consists of \( n \) columns representing each coefficient. To select a feature subset, \( \mathbf{F}_{\text{reduced}} \) using OA, a binary string of length \( 1 \times n \) is defined so that each column has a bit assigned to it. The initial value of the binary string can be randomly defined during initialization. A value of 1 will be given to the binary string to indicate the column will be considered in the construction of \( \mathbf{F}_{\text{reduced}} \), while the value of 0 is written to show the column is ignored.

In the swarm/colony, each particle carries a vector in solutions \( \mathbf{x} \). This vector contains the “probabilities of change” defined earlier. During optimization, the vector elements will be used as a reference to altering the binary string from its initial state (zero). If the particle vector element is more than 0.5, the binary bit will be changed to 1, otherwise the bit value is maintained.

C. Experiment Setup

Table 1 shows the parameter values for the NARMAX experiments using BABC and BPSO. Four different pre-processing methods were tested in the experiments:

- No magnitude scaling, 50:50 training and testing division ratio using block division method (PP0).
- No magnitude scaling, 50:50 training and testing division ratio using interleaving division method (PP1).
- Magnitude scaling between -1 and 1, 50:50 training and testing division ratio using block division method (PP2).
- Magnitude scaling between -1 and 1, 50:50 training and testing division ratio using interleaving division method (PP3).

| Parameters                | NARMAX |
|---------------------------|--------|
| Fitness criterion         | AIC, FPE, MDL |
| Swarm size                | 10,20,30,40,50 |
| Max Iterations            | 500, 1000, 1500 |
| Initial Seed              | 0, 10 000, 20 000 |
| Limit (BABC)              | 0, 0.25, 0.5, 0.75, 1 |

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

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

III. RESULTS AND DISCUSSION

Because NARMAX is a recursive model, it has to be implemented in stages. Initially, the NARX model is constructed (section A). The residuals from the NARX model was then used to create the Moving Average (MA) model (section B). Finally, the combined NARX and MA model was used to construct the NARMAX model (section C).

A total of 45 and 180 parameter combinations were analyzed for BPSO and BABC, respectively. Additionally, experiments on each of these parameter combinations were repeated for each of the information criteria (AIC, FPE, MDL). BABC had more parameter combinations because of the additional limit parameter, which is not applicable for BPSO.

A. NARX Part

For the NARX model, the optimal results were obtained using FPE as the fitness criterion and PP0 as the pre-processing method. Both BABC and BPSO obtained similar solutions with a fitness of \( 1.60 \times 10^{-5} \). The training and testing MSE values of the optimal solution were \( 2.69 \times 10^{-5} \) and \( 2.72 \times 10^{-5} \) respectively as shown in Table 2.

| Criterion          | FPE          |
|--------------------|--------------|
| Pre-processing     | PP0          |
| Fitness            | 1.60 \times 10^{-5} |
| Training Violation | 45           |
| Testing Violation  | 47           |
| Training MSE       | 2.69 \times 10^{-5} |
| Testing MSE        | 2.72 \times 10^{-5} |

The significant result is shown in Fig. 2 and Fig. 3, which demonstrates the ability of BABC to consistently be able to find the optimal solution many times over the many experiments with different parameter settings. In comparison to BPSO, the solution distribution was inconsistent as the solutions were much more scattered and unfocused.

Fig. 2: BABC and BPSO convergence histogram for the FRA NARX model. BABC managed to consistently discover the best solution \( (1.60 \times 10^{-5}) \) even when given different test parameters. The BPSO results were more distributed, discovering many different model structures resulting in various fitness values.
This observation is further demonstrated in Fig. 3, where it shows the convergence versus the number of iterations for all experiments with different parameter settings with the FPE information criterion and PP0 as the pre-processing method. It can be seen that BPSO has a tendency to converge faster (as demonstrated by the rapid decrease in the fitness function), but fails to improve its fitness anymore after 500 iterations. In comparison to BABC, the reduction in the fitness function is more gradual with the ability to search the solution space in a more in-depth manner which resulting in overall better-quality solutions.

![Fig. 3 BABC and BPSO convergence across iterations over various parameter combinations. BABC solutions are more gradual compared to BPSO, but BABC has the ability to explore in a more in-depth manner resulting in a much better fitness relative to BPSO.](image)

**B. MA Part**

Being a recursive model, NARMAX introduces the residuals back into the NARX model in an attempt to improve its prediction ability.

**TABLE III**

| Technique       | BPSO         | BABC         |
|-----------------|--------------|--------------|
| Criterion       | FPE          | FPE          |
| Pre-processing  | PP0          | PP0          |
| Fitness         | 7.07×10^{-7} | 7.05×10^{-7} |
| Training Violation | 43          | 42           |
| Testing Violation | 52          | 51           |
| Training MSE    | 1.35×10^{-6} | 1.35×10^{-6} |
| Testing MSE     | 1.53×10^{-6} | 1.54×10^{-6} |

Table 3 shows the optimal MA result achieved using BABC and BPSO. The FPE fitness and MSE values were comparable. However, with the BABC algorithm, the number of correlation violations was lower compared to BPSO. This indicates that the MA residuals were more randomly distributed compared to BPSO. Furthermore, the BABC algorithm showed much more consistency in finding the best fitness as indicated in Fig. 4. This observation is similar to the experiment performed in section A.

![Fig. 4 BABC and BPSO convergence histogram for the MA NARX model. BABC managed to consistently discover the best solution (7.05×10^{-7}) even when given different test parameters. The BPSO results were more distributed, discovering many different model structures resulting in various fitness values.](image)

**C. NARMAX Part**

The NARMAX model is a combination of the optimal NARX (section A) and MA (section B) models. The results of the NARMAX model are shown in Table 4.

Comparing Table 4 and Table 2, it appears that the addition of MA terms to the NARX model did not add a significant advantage to the NARX model as the MSE values were slightly increased. Because of this, NARX was selected as the final model for comparison between BABC and BPSO.

**TABLE IV**

| Technique       | BPSO         | BABC         |
|-----------------|--------------|--------------|
| Criterion       | FPE          | FPE          |
| Pre-processing  | PP0          | PP0          |
| Training Violation | 47          | 47           |
| Testing Violation | 53          | 53           |
| Training MSE    | 6.2465×10^{-5} | 6.2469×10^{-5} |
| Testing MSE     | 6.2894×10^{-5} | 6.2923×10^{-5} |

**D. Model Validation**

The validation results for the NARX model is presented in this section. Validation measures two aspects, which are the model fit (using the OSA test) and the whiteness of the residuals (residual plot, correlation tests, and residual histogram analysis).

The OSA results for BABC and BPSO training and testing sets are shown in Fig. 5 and Fig. 6 respectively. The model prediction results (red dotted line) closely follow the actual system output (solid blue line). This indicates that the model was able to approximate the actual system well. This observation is further supported by the R2 value nearing 100% for both training and testing sets.

For the model to be accepted, the residuals of the model need to exhibit properties similar to white noise. This can be validated using residual plots (Fig. 7), correlation (Fig. 8(a) to Fig. 8(e)) and histogram tests (Fig. 9).
The correlation tests measure the similarity between two signals [41] across different lags. For the residuals to appear as white noise, all correlation coefficients (except for lag zero in autocorrelation tests) should be within the 95% confidence limits (dotted red line). However, in real-life cases, a perfect result is difficult to achieve, and a sufficiently good result is acceptable. Based on the results in Fig. 8(a) to 8(e), the total number of correlation violations is 45 from 160 lags tested (28%) which leaves the remaining 72% correlation coefficients within the 95% confidence limits. The error histogram in Fig. 9 also supports the observation that the residuals are random. This is because the distribution of residuals appears as a bell-shaped curve indicating the residuals are randomly distributed. Because of this, we consider the NARX model to be valid and acceptable.
The optimal NARX solution similarly achieved by both BPSO and BABC (FPE criterion equal to 2.69×10^{-5} and 2.72×10^{-5} for training and testing sets respectively with PP0 pre-processing). Both algorithms managed to find the optimal solution from 2×10^{44} possible model structures. The optimal solution had passed all the necessary tests for it to be considered a valid and acceptable model.

Comparing the two optimization algorithms, BPSO converged faster than BABC. However, in terms of convergence consistency, BABC was proven to significantly outperform BPSO by a huge percentage margin (BABC found the optimal solution 82.2% of the time, while BPSO managed 6.67% for the NARX model). This proves that BABC was the better method for feature selection.

ACKNOWLEDGMENT
The authors would like to express their gratitude to Universiti Teknologi MARA, Malaysia for equipment and infrastructure support for the research.

REFERENCES
[1] F. Li and L. Pengfei, "The research survey of system identification method," in Proc. IEEE ICHIMSC’13, 2013, p. 397.
[2] T. G. Ling, M. F. Rahmat, and A. R. Husain, "System identification and control of an electro-hydraulic actuator system," in Proc. IEEE ICSPA'12, 2012, p. 85.
[3] D. Xiaog, "System identification based on particle swarm optimization algorithm," in Proc. ICCIS'09, 2009, p. 259.
[4] B. Khalil and A. Yesildirek, "System identification of UAV under an autopilot trajectory using ARX and Hammerstein-Wiener methods," in Proc. IEEE ISMA'10, 2010, p. 1.
[5] R. Zhen and G. G. Zhu, "Multirate closed-loop system identification of a variable valve timing actuator for an internal combustion engine," in Proc. ACC’10, 2010, p. 664.
[6] J. Tchorzewski, "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 Proc. ICEEM’14, 2014, p. 1.
[7] A. Noshadi, J. Shi, W. S. Lee, P. Shi, and A. Kalam, "Genetic algorithm-based system identification of active magnetic bearing system: A frequency-domain approach," in Proc. IEEE ICAA’14, 2014, p. 1281.
[8] L. Guangjun, X. Xiaoping, and W. Feng, "Identification of a kind of nonlinear system," in Proc. IJCNN’11, 2011, p. 1730.
[9] I. J. Leontaris and S. A. Billings, "Input-output parametric models for nonlinear systems," International Journal of Control, vol. 41, pp. 311-341, 1985.
and Measurement using NARMAX structures," IEEE Transactions on Instrumentation and Measurement, vol. 50, pp. 893-898, Aug. 2001.

[11] N. Chiras, C. Evans, and D. Rees, "Nonlinear gas turbine modeling using NARMAX structures," IEEE Transactions on Instrumentation and Measurement, vol. 50, pp. 597-608, 2001.

[12] M. Valivere, M. J. Korenberg, and P. Caminal, "Model identification of the neural control of the cardiovascular system using NARMAX models," in Proc. CVC'91, 1991, p. 585.

[13] S. L. Kukreja, H. L. Galiana, and R. E. Kearney, "NARMAX representation and identification of ankle dynamics," IEEE Transactions on Biomedical Engineering, vol. 50, pp. 70-81, Jan. 2003.

[14] S. L. Kukreja, R. E. Kearney, and H. L. Galiana, "A least-squares parameter estimation algorithm for switched Hammerstein systems with applications to the VOR," IEEE Transactions on Biomedical Engineering, vol. 52, pp. 431-444, Mar. 2005.

[15] B. Liu, L. Wang, P. Shen, and G. Lv, "NDEKF neural network applied to electronically controlled fuel injection system," in Proc. IEEE CIEA'07, 2007, p. 351.

[16] B. Yan, Z. Yaочun, and J. Yiheng, "A new nonlinear system identification method using gene expression programming," in Proc. IEEE ICCMA'07, 2007, p. 2951.

[17] L. Huijun, J. Gang, and M. Zengliang, "A nonlinear predictive model based on multilayer perceptron network," in Proc. IEEE ICAL'07, 2007, p. 2686.

[18] I. M. Yassin, A. Zabidi, M. S. A. M. Ali, N. M. Tahir, H. A. Hassan, H. Z. Abidin, and Z. I. Rizman, "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, vol. 6, pp. 630-637, Oct. 2016.

[19] O. M. Vall and R. M'hiri, "An approach to polynomial NARX/NARMAX systems identification in a closed-loop with variable structure control," International Journal of Automation and Computing, vol. 5, pp. 313-318, Jul. 2008.

[20] S. A. Billings, Identification of NARMAX and Related Models, ser. Control System, Robotics, and Automation. France: Encyclopedia of Life Support Systems, 2001, vol. 6.

[21] S. R. Anderson, N. F. Lepora, J. Porrill, and P. Dean, "Nonlinear dynamic modeling of isometric force production in primate eye muscle," IEEE Transactions on Biomedical Engineering, vol. 57, pp. 1554-1567, Jul. 2010.

[22] B. A. Amisigo, N. V. D. Giesen, C. Rogers, W. E. I. Andah, and J. Friesen, "Monthly streamflow prediction in the Volta Basin of West Africa: A SISO NARMAX polynomial modelling," Physics and Chemistry of the Earth, vol. 33, pp. 141-150, Dec. 2008.

[23] Y. Guo, L. Z. Guo, S. A. Billings, and H. L. Wei, "Identification of nonlinear systems with non-persistent excitation using an iterative forward orthogonal least squares regression algorithm," International Journal of Modelling, Identification and Control, vol. 23, pp. 1-7, Jan. 2015.

[24] S. A. Billings and S. Chen, "Extended model set, global data and threshold model identification of severely non-linear systems," International Journal of Control, vol. 50, pp. 1897-1923, Nov. 1989.

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

[26] N. A. Rahim, M. N. Taib, and M. I. Yusof, "Nonlinear system identification for a DC motor using NARMAX approach," in Proc. IEEE ACS'03, 2003, p. 305.

[27] M. H. F. Rahman, "System identification of essential oil extraction system," Phd thesis, Universiti Teknologi MARA, Selangor, Malaysia, 2008.

[28] M. Norgaard, O. Ravn, N. K. Poulsen, and L. K. Hansen, Neural Networks for Modeling and Control of Dynamic Systems: A Practitioner's Handbook (Advanced Textbooks in Control and Signal Processing), London, England: Springer, 2013.

[29] H. Wang, Y. Shen, T. Huang and Z. Zeng, Nonlinear System Identification Based on Recurrent Wavelet Neural Network, ser. Advances in Soft Computing, London, England: Springer, 2009, vol. 56.

[30] H. Louisif, K. Nouri, and N. B. Braiekh, "A new efficient hybrid intelligent method for nonlinear dynamical systems identification: The wavelet kernel fuzzy neural network," Communications in Nonlinear Science and Numerical Simulation, vol. 32, pp. 10-30, Mar. 2016.

[31] G. Kerschen, K. Worden, and J. C. Golinval, "Past, present and future of nonlinear system identification in structural dynamics," Mechanical Systems and Signal Processing, vol. 20, pp. 505-592, Apr. 2006.

[32] H. L. Wei and S. A. Billings, "Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information," International Journal of Modelling, Identification and Control, vol. 3, pp. 341-356, Jan. 2008.

[33] Y. Xue, J. Jiang, T. Ma, and C. Li, "The performance research of artificial bee colony algorithm on the large scale global optimisation problems," International Journal of Wireless and Mobile Computing, vol. 9, pp. 300-305, 2015.

[34] V. Kaliapan and M. Thathan, "Enhanced ABC based PID controller for nonlinear control systems," Indian Journal of Science and Technology, vol. 8, pp. 48-56, Apr. 2015.

[35] C. Ozturk, E. Hancer and D. Karaboga, "A novel artificial bee colony algorithm based on genetic operators," Information Sciences, vol. 297, pp. 154-170, Mar. 2015.

[36] K. S. Kaswan, S. Choudhary, and K. Sharma, "Applications of artificial bee colony optimization technique: Survey," in Proc. IEEE ICCSGD'15, 2015, p. 1660.

[37] Z. L. Yang, A. Wu, and H. Q. Min, "Deployment of wireless sensor networks for oilfield monitoring by multiobjective discrete binary particle swarm optimization," Journal of Sensors, 2016, pp. 1-15, Apr. 2016.

[38] M. Konar and A. Bagis, "Performance comparison of particle swarm optimization, differential evolution and artificial bee colony algorithms for fuzzy modelling of nonlinear systems," Elektronika ir Elektrotechnika, vol. 22, pp. 8-13, Mar. 2016.

[39] B. Akay, "A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding," Applied Soft Computing, vol. 13, pp. 3066-3091, Jun. 2013.

[40] I. Markovsky, J. C. Willems, and B. De Moor, "Comparison of identification algorithms on the database for system identification DAVIS," in Proc. BMTNS'06, 2006, p. 2838.

[41] M. N. M. Nor, R. Jailani, N. M. Tahir, I. M. Yassin, Z. I. Rizman, and R. Hidayat, "EMG signals analysis of BF and RF muscles in autism spectrum disorder (ASD) during walking," International Journal on Advanced Science, Engineering and Information Technology, vol. 6, pp. 793-798, Oct. 2016.