A Novel Single Parent Mating Technique in Genetic Algorithm for Discrete – Time System Identification

Farah Ayiesya Zainuddin[1], Md Fahmi Abd Samad[1,2,*], Hishamuddin Jamaluddin[3], Abul K. M. Azad[4]

1   Fakulti Teknologi dan Kejuruteraan Mekanikal, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
2   Center for Advanced Computing Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
3   Faculty of Engineering and Information Technology, Southern University College, 81300 Johor Bahru, Johor, Malaysia
4   College of Engineering and Engineering Technology, Department of Engineering Technology, Northern Illinois University, Illinois 60115, United States of America

ARTICLE INFO	ABSTRACT
	System identification is concerned with the construction of a mathematical model based on given input and output data to represent the dynamical behaviour of a system. As a step in system identification, model structure selection is a step where a model perceived as adequate system representation is selected. A typical rule is that the model must have a good balance between parsimony and accuracy in estimating a dynamic system. As a popular search method, genetic algorithm (GA) is used for selecting a model structure. However, the optimality of the final model depends much on the optimality of GA. This paper introduces a novel mating technique in GA based on the chromosome structure of the parents such that a single parent is sufficient in achieving mating that demonstrates high exploration capability. In investigating this, four systems of linear and nonlinear classes were simulated to generate discrete-time sets of data i.e. later used for identification. The outcome shows that GA incorporated with the mating technique within 10% - 20% of the population size is able to find optimal models quicker than the traditional GA.

1. Introduction

System identification (SI) is known as a field of study where a mathematical relation between variables and terms of a process is determined. This is done based on observed input-output data with the aim to enable better control of a system [1,2]. Modelling of system can be divided into continuous-time and discrete- time modelling. Many real-world systems, for example, in the fields of mechanics, electricity, chemistry, economics, biology and ecology are dynamic systems [3-5]. Although these systems warrant a continuous-time model rather than a discrete-time model, it is practical to assume that data are interconnected in a discrete-time connections and therefore employing principles of discrete-time identification is favorable.

SI is generally known to encompass 4 distinct steps i.e. data acquisition, model structure selection, parameter estimation and lastly, validation in search of an optimal model that represents the system at-hand [6]. The definition of 'optimal' requires the model to be able to accurately predict the output of the system using limited input and system characteristics yet the model also has to be simple or parsimonious. A parsimonious model is a model where the variables or terms that it carries is few enough to make acceptable prediction [7].

Any implementation of system identification or also called system modelling should be able to be completed within a short time and cost to ensure process efficiency. With the emergence of various meta-heuristic methods today, such as evolutionary computation and genetic algorithm, many researchers took the advantage of incorporating these methods to increase the efficiency in optimization. GA had been applied in image processing, laser technology, aeronautics, robotics and many more practical problem areas successfully [8-12]. Works involving the use of evolutionary computation in system identification are widely available [13-16]. However, the efficiency of using evolutionary computation, more specifically genetic algorithm (GA), in system identification is still insufficient. There is still potential that more optimal models may be found and furthermore, found at a quicker rate. When a simulated (known) model is used in investigation, it is always probable that the original model is the most optimal and therefore, inefficiency is easily detected.

Some researchers proposed strategies to include a mating process in GA. An example includes a technique called correlative tournament selection where it analyses correlation between parents for implementation of tournament selection [17]. Galán *et al.,* [18] introduced a self-adaptive mating based on parent similarity or fitness. Other techniques use assortative mating [19,20] and tabu search [21]. However, some of these approaches require extra calculation such as to check for parent similarity before suitable mating is achieved.

This paper aims to propose a novel mating technique that is targeted to maintain proper balance between exploring for optimal solutions and exploiting heuristic information within GA. The sections are divided as follows: Section 2 explains the methodology of study including the mechanism of the mating proposal. Section 3 contains the raw results followed by a discussion in Section 4. The conclusion of the study is provided in Section 5.

2. Method

When GA starts, a specified number of chromosomes are created at random in a population. In system identification, these chromosomes mean different structures of possible models to represent the system. Their optimality is then evaluated based on a specified objective function (OF). From this population, in the hope of achieving better model, certain manipulations are made to the chromosomes, and every time a cycle of manipulation is completed, the new chromosomes are evaluated. The process continues until the allowed maximum number of cycles, called generation. It is well-known that in traditional GA, the manipulation revolves around three processes - selection, crossover and mutation (Figure 1(a)) [22,23].

To select the best individuals and let them pass their genes, the idea of the selection phase is based on their fitness scores. The fitness values determine which chromosomes to survive and generate the population of the next generation [24-26]. The selection stage functions by copying some good chromosomes where some chromosomes are copied more than once into an intermediate population or mating pool where crossover takes place. Traditional GA allows the parents to be taken randomly from this pool without any mating strategy for crossover to be carried out. It has become somewhat a stigma that crossover must be done with more than 1 parent. However, the 2 selected parents from the pool may have the same genes which means the crossover

between these parents may not be able to effectively produce different, or possibly better, chromosome offsprings. Figure 1(b) illustrates the possible offsprings, in a common crossover, specifically uniform crossover. It is seen that without any mating strategy, there can only be 4 possible offsprings.
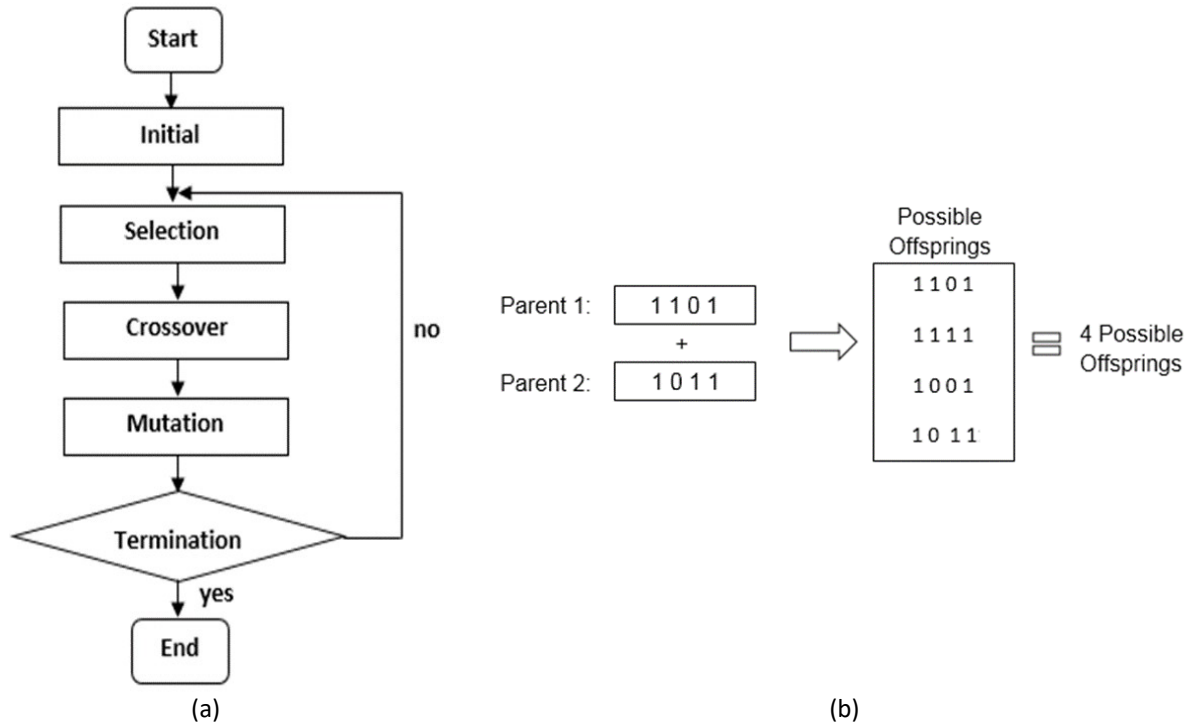

Fig. 1. (a) Traditional GA (without mating) (b) Outcome of possible offsprings

By introducing a suitable mating technique after the selection process, it is possible to maintain high diversity in the outcome of crossover. An example of GA with mating technique is shown in Figure 2(a). In this paper, prior to crossover, some good parents in the mating pool are replicated. The genes of these replicates are converted from 1 to 0 and vice versa using inversion. The original parent is then paired with its inverted self, leading to single parent mating. After then, the parents in the pair go through the customary crossover process, such as single-point crossover. The mating emphasizes how the pair of chromosomes are made, not how their information is exchanged, as information exchange are carried out through crossover. Figure 2(b) illustrates that, with single parent mating, there are 2^4 possible offsprings, assuming uniform crossover (number and positions of crossover points are random). It can be seen that many more possibilities may be found and the population may contain suitable diversity of solution choices.
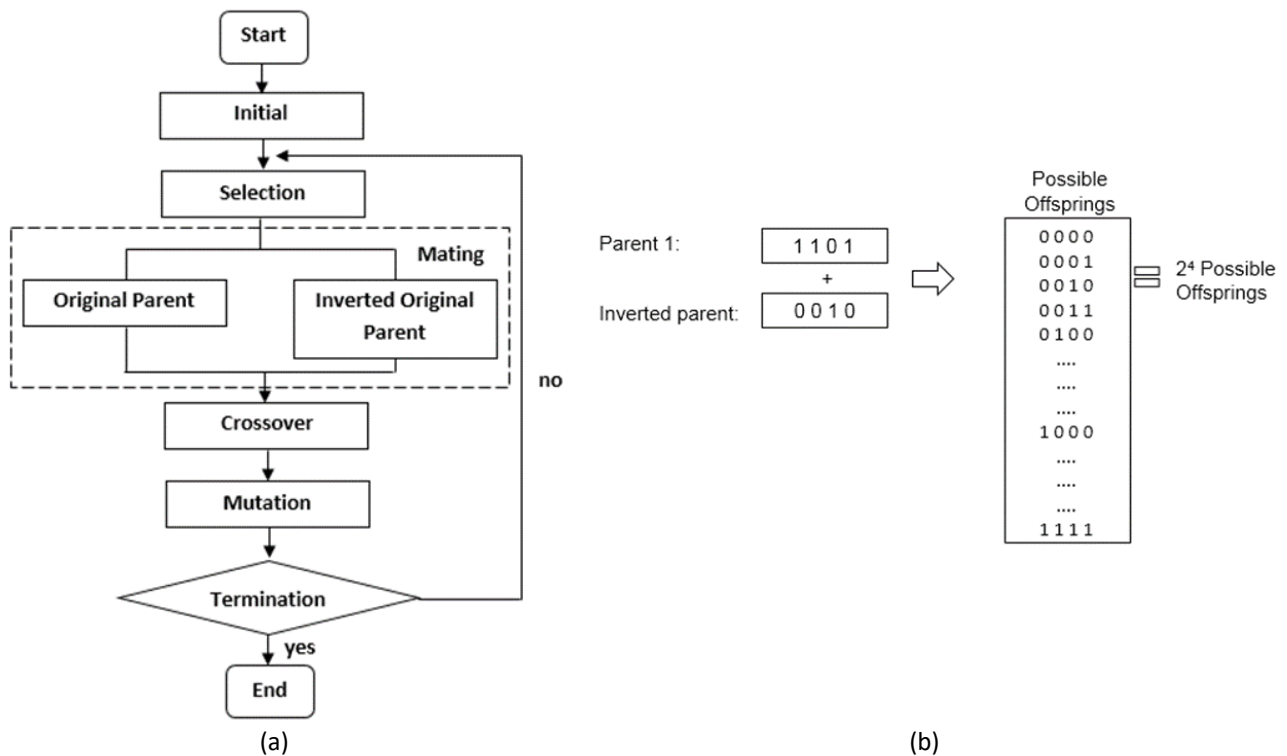
Fig. 2. (a) GA with single parent mating (b) Outcome of possible offsprings

The effectiveness of the proposal was made by testing on 4 simulated models. These models were discrete-time models known as linear or nonlinear autoregressive with exogenous input (ARX or NARX) models. Following are the models while the model characteristics and search space specifications are given in Table 1.

Simulated Model 1 (Sim 1):

$$y(t) = 0.5y(t-1) - 0.2y(t-4) + 0.5y(t-8) + 0.6u(t-2) - 0.2u(t-9) + e(t) \tag{1}$$

Simulated Model 2 (Sim 2):

$$y(t) = 0.4y(t-1) + 0.4u(t-1) - 0.6u(t-3) - 0.7y(t-1)u(t-1) - 0.2y^2(t-2) + 0.2u^2(t-3) + e(t) \tag{2}$$

Simulated Model 3 (Sim 3):

$$y(t) = 0.4y(t-3) + 0.3u(t-1) + 0.7y(t-2)u(t-1) + 0.1y(t-3)u(t-2) - 0.5u(t-1)u(t-2) - 0.4u(t-1)u(t-3) + e(t) \tag{3}$$

Simulated Model 4 (Sim 4):

$$y(t) = 0.8u(t-1) + 0.5y(t-1)y(t-2) + 0.1y(t-1)u^2(t-1) - 0.1y(t-2)u(t-1)u(t-2) - 0.4u^2(t-1)u(t-2) + 0.2u(t-1)u^2(t-2) + e(t) \tag{4}$$

Table 1
Model Characteristic and Search Space Specification

		Sim 1	Sim 2	Sim 3	Sim 4
Model characteristics	Number of correct regressor	5	6	6	6
	Nonlinearity	1	2	2	3
	Max output lag order	10	2	3	2
Search space specification	Max input lag order	10	3	3	2
	Time delay	1	1	1	1
	Number of possible regressor	20	20	27	34
	Number of possible model	1048575	1048575	134217727	17179869183

The input $u(t)$ was generated randomly from a uniform distribution to represent white signal. The values were selected from the interval [-1, 1]. The noise $e(t)$ was also generated randomly from a uniform distribution to represent white noise and the values were from the interval [-0.01, 0.01]. Using the models, 500 data were generated to be used in identification.

In GA, the model structures were represented by binary chromosome where 1 means presence and 0 means absence of variable / term [6]. The specifications of GA were population size = 200, maximum generation = 100, mutation probability = 0.01 and crossover probability = 0.6. This paper used roulette-wheel selection, bit-flip mutation and single-point crossover. The elitism strategy was also used where the best chromosome found so far was kept from generation to generation. The parameter estimation was done by the least squares method.

The objective function used was parameter magnitude-based information criterion 2 [27]:

$$\text{PMIC2} = \sum_n ((y(t) - \hat{y}(t))^2 + \sum_j \frac{1}{\theta}; \ \theta \neq 0 \tag{5}$$

where $y(t)$ was the original output and $\hat{y}(t)$ was the predicted output. The variable θ was the value of parameter, n was the number of data and j was the maximum number of parameters in the model.

Another measurement used in this study was the error index (EI), written as follows:

$$\text{EI} = \sqrt{\frac{\sum (y(t) - \hat{y}(t))^2}{\sum y^2(t)}} \tag{6}$$

During the GA run, the mating was carried out on specific percentages of the parents in the mating pool where the size of the mating pool was the same as the population once selection was completed. The percentages tested were 0%, 10%, 20%, 30%, 40% and 50%. For instance, if the percentage was 30%, this much percentage of parents from the pool was copied then inverted, making another 30% of (inverted) parents. Mating was done between the parents and its inverted selves (comprising now of 60% of population). The remaining 40% came randomly from the initial mating pool. With this setting, it may be noted that 0% represents GA without mating. For each percentage, 5 runs were made.

3. Results

Figure 3 shows the average results of 0% to 50% mating techniques. The graphs are labelled single 0.0 to single 0.5 representing 0% mating to 50% mating using single-point crossover. Based on Figure 3(a), in the final generation, 30% mating (single 0.3) has the lowest OF value. Based on all runs, analysis of the final generation shows that 0% mating (single 0.0) has the lowest value of OF of 16.6213 in 1 run out of 5 runs. Out of 5 correct regressors, the run selected 3 regressors and has an

EI of 0.2317. The lowest OF throughout all runs, is however found with 10% mating in 2 runs, 20% mating in 3 runs and 30% mating in 2 runs which is 14.5075. The runs collected 4 correct regressors and has an EI of 0.1476. This shows that GA with mating is able to find model of lower OF, gathers more correct regressors and has lower EI.
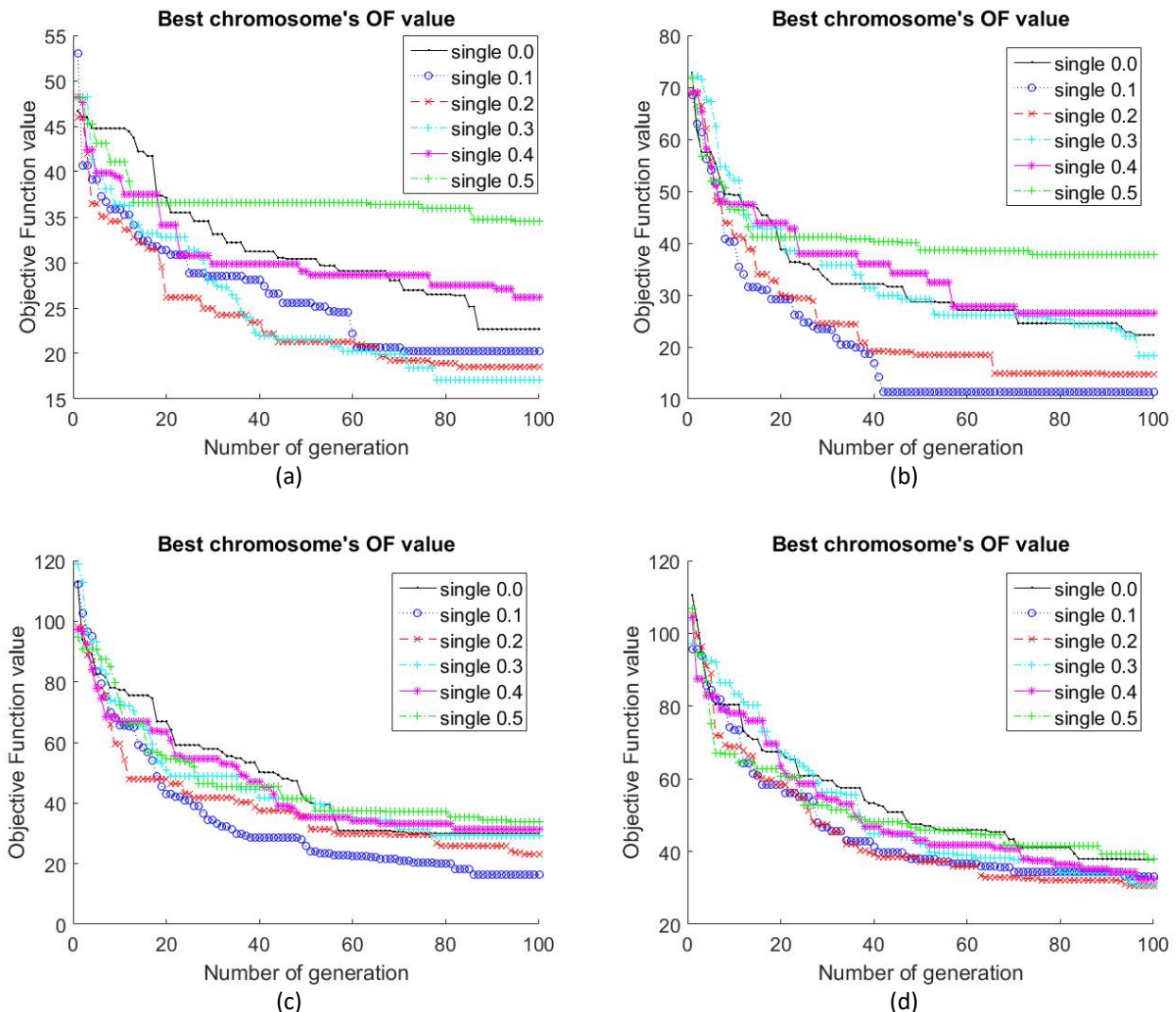

Fig. 3. Best chromosome's OF value (a) Sim 1 (b) Sim 2 (c) Sim 3 (d) Sim 4

Figure 3(b) shows that, in the final generation, 10% mating (single 0.1) has the lowest average OF value compared to the other percentages of mating. Analysis show that the lowest OF value which is 11.3023 is found in 0% mating (1 run), 10% mating (5 runs), 20% mating (4 runs), 30% mating (3 runs) and 40% mating (1 run). These runs collected 4 exact regressors and has an EI of 0.1655. A trend of increasing frequency of finding the model then decreasing with the rise of mating percentage is easily noted. The model is found in all runs when using 10% mating and in almost all runs when using 20% mating.

Figure 3(c) again shows that in the final generation, 10% mating (single 0.1) has the lowest average OF value. Comparison shows that 0% mating (1 run) has the lowest value of OF of 21.6643. It collected 4 regressors where 2 are correct and has an EI of 0.5106. Meanwhile, 10% mating (1 run) has the lowest overall OF value which is 11.9666. The EI is 0.0602. It collected 5 correct regressors – the regressor $y(t - 3)u(t - 2)$ was not collected. It is worth mentioning that this regressor has a

small parameter of 0.1 and therefore not easily picked up by any GA. Conclusively, in 10% mating, there are more correct regressors, lower OF and far lower EI.

In Figure 3(d), 30% mating (single 0.3) has the lowest average OF value. The differences with other percentages are noted small from the graph especially when looking at 20% and 30% mating. Zero percent mating (1 run) has the lowest value of OF of 16.5421 with 12 regressors where $u(t - 1)$ is the only correct regressor. The EI is 0.1408. Individually, 20% mating (1 run) has the lowest OF value equal to 19.3740 with EI of 0.1985. It collected 8 regressors where $u(t - 1)$ and $u(t - 1)u^2 (t - 2)$ are the two correct regressors. Despite a slightly higher EI, the run selected a more parsimonious model with more correct regressors. For this model, the 20% and 30% mating shows better potential.

4. Discussion

In this section, the average of best chromosomes OF value from the final generation are shown in Table 2. The data is graphed in Figure 4. The ones boldfaced in the table are the smallest OF. From the table, 10% mating and 30% mating was found to have the lowest OF for two simulated models, each. The average lowest OF value based on all 4 models is found using 10% mating. The graph shows a trough in the middle area between 0% and 40% mating.

Table 2
Average Final Best Chromosome's Objective Function

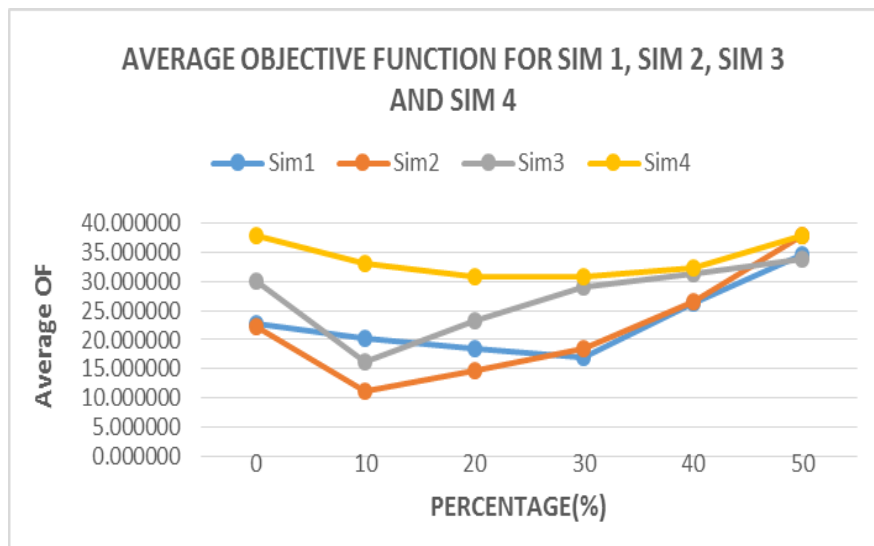|	0% Mating	10% Mating	20% Mating	30% Mating	40% Mating	50% Mating
Sim1	22.6850	20.2957	18.5549	**17.0856**	26.1471	34.5842
Sim2	22.3254	**11.3023**	14.8166	18.3864	26.4402	37.7246
Sim3	29.9479	**16.2996**	23.2071	29.1054	31.2604	33.8902
Sim4	37.8153	33.0298	30.6942	**30.6725**	32.3268	37.9068
Average	28.1934	**20.2319**	21.8182	23.8125	29.0436	36.0265

Fig. 4. Best chromosome's OF value (a) Sim 1 (b) Sim 2 (c) Sim 3 (d) Sim 4

Table 3 shows the result of best mating percentage based on curve fitting of the data from Table 2 using polynomial of 2 degrees to maximum possible degree (5), using intervals of 2.5%. Average A is computed from the average OF data in Table 2 whilst average B is calculated from the percentage data of Sim 1 to Sim 4 in Table 3. Both averages show that the best is between 10% and 20%, inclusively.

Table 3
Best Mating Percentage for Four Simulated Models

	Polynomial 2 (%)	Polynomial 3 (%)	Polynomial 4 (%)	Polynomial 5 (%)
Sim1	20	22.5	22.5	27.5
Sim2	17.5	15	12.5	7.5
Sim3	15	12.5	10	7.5
Sim4	25	27.5	27.5	25
Average A	20	17.5	12.5	10
Average B	20	20	17.5	17.5

5. Conclusion

GA speeds up the identification of discrete-time model without the need to evaluate all possible models. In some of the runs, the correct model is almost able to be identified using PMIC2 as the OF. In most cases, a more parsimonious model is found with small EI. This is important in achieving a model which is both accurate and parsimonious. Incorporation of single parent mating in GA is shown to be able to find a lower OF model than without mating. Even though using half of the population for mating does not perform well, the best percentage is found to be between 10% and 20%, inclusively. With proper mating, the "marriage" of these parents is able to explore new search space of solution, producing more varied offspring (hence, system model) that cannot be achieved by traditional procedure of GA.

References
[1] Keesman, Karel J. *System identification: an introduction*. Springer Science & Business Media, 2011. https://doi.org/10.1007/978-0-85729-522-4_1
[2] Duong, Vu, and Allen R. Stubberud. "System identification by genetic algorithm." In *Proceedings, IEEE Aerospace Conference*, vol. 5, pp. 5-5. IEEE, 2002.
[3] Saidon, Mohd Saifizi, Nasrul Amri Mohd Amin, Aqilah Che Sulaiman, Mohd Rizal Manan, Siti Marhainis Othman, Wan Azani Mustafa, and Faiz Arith. "Identification of ARX Model for Thermoelectric Cooling on Glass Windows." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 96, no. 2 (2022): 144-157. https://doi.org/10.37934/arfmts.96.2.144157
[4] Stappenbelt, Brad. "Prediction of the heave response of a floating oscillating water column wave energy converter." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 5, no. 1 (2015): 8-23.
[5] Oleolo, Ibrahim, Hayati Abdullah, Maziah Mohamad, Mohammad Nazri Mohd Jaafar, and Akmal Baharain. "Model Selection for the Control of Temperature in a Centralized Air Conditioning System." *Journal of Advanced Research in Applied Mechanics* 74, no. 1 (2020): 10-20.
[6] Abd Samad, Md Fahmi, Hishamuddin Jamaluddin, Robiah Ahmad, And Mohd Shafiek. "Deterministic Mutation-Based Algorithm for Model Structure Selection in Discrete-Time System Identification." *International Journal of Intelligent Control and Systems* 16, no. 3 (2011): 182-190.
[7] Samad, M. F. A., and A. R. M. Nasir. "Parameter Magnitude-Based Information Criterion for Identification of Linear Discrete-Time Model." *Proceedings of Innovative Research and Industrial Dialogue '16* (2017): 129-130.
[8] AL-Smadi, Yahia M., Yusra Aburmaileh, Kevin Russel, and Rajpal Sodhi. "Genetic Algorithm for Sensitivity Analysis of Automobile Hood Four-bar Mechanism Synthesis Using Motion Generation." *International Journal of Automotive and Mechanical Engineering* 19, no. 1 (2022): 9362-9378. https://doi.org/10.15282/ijame.19.1.2022.01.0720
[9] Hassanat, Ahmad B. A., and Esra'A. Alkafaween. "On enhancing genetic algorithms using new crossovers." *International Journal of Computer Applications in Technology* 55, no. 3 (2017): 202-212. https://doi.org/10.1504/IJCAT.2017.084774

[10] Zainuddin, Farah Ayiesya, and Md Fahmi Abd Samad, "A review of crossover methods and problem representation of genetic algorithm in recent engineering applications." *International Journal of Advanced Science and Technology* 29, no. 6s (2020): 759-769.

[11] Bahdad, Ali Ahmed Salem, and Sharifah Fairuz Syed Fadzil. "Design Optimization for Light-Shelves with Regard to Daylighting Performance Improvements in The Tropics." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 100, no. 3 (2022): 35-50. https://doi.org/10.37934/arfmts.100.3.3550

[12] Das, Ranjan. "Estimation of radius ratio in a fin using inverse CFD model." *CFD Letters* 3, no. 1 (2011): 40-47.

[13] Nowaková, Jana, and Miroslav Pokorný. "System identification using genetic algorithms." In *Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA 2014*, pp. 413-418. Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-08156-4_41

[14] Shin, Gang-Wook, Young-Joo Song, Tae-Bong Lee, and Hong-Kyoo Choi. "Genetic algorithm for identification of time delay systems from step responses." *International Journal of Control Automation and Systems* 5, no. 1 (2007): 79-85.

[15] Zhang, Ridong, and Jili Tao. "A nonlinear fuzzy neural network modeling approach using an improved genetic algorithm." *IEEE Transactions on Industrial Electronics* 65, no. 7 (2017): 5882-5892. https://doi.org/10.1109/TIE.2017.2777415

[16] Huang, Yuhao, Liang Gao, Zhang Yi, Kang Tai, Pankaj Kalita, Paweena Prapainainar, and Akhil Garg. "An application of evolutionary system identification algorithm in modelling of energy production system." *Measurement* 114 (2018): 122-131. https://doi.org/10.1016/j.measurement.2017.09.009

[17] Matsui, K. "New selection method to improve the population diversity in genetic algorithms." In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)*, vol. 1, pp. 625-630. IEEE, 1999.

[18] Galán, Severino F., Ole J. Mengshoel, and Rafael Pinter. "A novel mating approach for genetic algorithms." *Evolutionary computation* 21, no. 2 (2013): 197-229. https://doi.org/10.1162/EVCO_a_00067

[19] Fernandes, Carlos, Rui Tavares, Cristian Munteanu, and Agostinho Rosa. "Using assortative mating in genetic algorithms for vector quantization problems." In *Proceedings of the 2001 ACM symposium on Applied computing*, pp. 361-365. 2001. https://doi.org/10.1145/372202.372367

[20] Huang, Chien-Feng. "An analysis of mate selection in genetic algorithms." In *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, pp. 766-766. 2001.

[21] Ting, C-K., and H. K. Buning. "A mating strategy for multi-parent genetic algorithms by integrating tabu search." In *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.*, vol. 2, pp. 1259-1266. IEEE, 2003.

[22] Sivanandam, S. N., and S. N. Deepa. *Genetic algorithms*. Springer Berlin Heidelberg, 2008.

[23] Mirjalili, Seyedali, Jin Song Dong, Ali Safa Sadiq, and Hossam Faris. "Genetic algorithm: Theory, literature review, and application in image reconstruction." *Nature-Inspired Optimizers: Theories, Literature Reviews and Applications* (2020): 69-85. https://doi.org/10.1007/978-3-030-12127-3_5

[24] Akter, Shamima, Nazmun Nahar, Mohammad Shahadat Hossain, and Karl Andersson. "A new crossover technique to improve genetic algorithm and its application to TSP." In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1-6. IEEE, 2019. https://doi.org/10.1109/ECACE.2019.8679367

[25] Katarya, Rahul, and Ashima Yadav. "A comparative study of genetic algorithm in sentiment analysis." In *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, pp. 136-141. IEEE, 2018. https://doi.org/10.1109/ICISC.2018.8399051

[26] Keshanchi, Bahman, Alireza Souri, and Nima Jafari Navimipour. "An improved genetic algorithm for task scheduling in the cloud environments using the priority queues: formal verification, simulation, and statistical testing." *Journal of Systems and Software* 124 (2017): 1-21. https://doi.org/10.1016/j.jss.2016.07.006

[27] Abd Samad, Md Fahmi, and Abdul Rahman Mohd Nasir. "Parameter magnitude-based information criterion in identification of discrete-time dynamic system." *Journal of Mechanical Engineering (JMechE)* SI4, no. 1 (2017): 119-128.