

# Statistical Analysis for Swarm Intelligence — Simplified

Nor Azlina Ab. Aziz, Marizan Mubin, Zuwairie Ibrahim, and Sophan Wahyudi Nawawi

**Abstract**—Usage of statistical tools is important in reporting valid and unbiased findings of a research. Therefore, this paper aims to discuss the statistical tools commonly used among researchers in the swarm intelligence field. Among the tools discuss is the boxplot. Boxplot is a common data representation method chosen in the field. Nonparametric statistical test for pairwise comparison, namely, Wilcoxon signed rank test, and a test for multiple comparison, namely, Friedman test are also reviewed.

**Index Terms**—Boxplot, Friedman test, Holm procedure, Wilcoxon signed rank test.

## I. INTRODUCTION

The days where mean and standard deviation are sufficient in evaluation of the performance of a swarm intelligence algorithm are gone. Now a common trend observed in this area is the usage of statistical analysis methods to validate any observations and findings regarding algorithms performance. The usage of statistical tools is important in order to make a more concrete and unbiased conclusion on any algorithm.

This paper aims to discuss the statistical methods that can be used in representing and analyzing the results of swarm intelligence algorithm. In Section II, boxplot which is a common data representation method is discussed. The boxplot shows the quality and distribution of the data. The fundamental components used in the statistical analysis are described in Section III. Often the results obtained by swarm intelligence algorithms are not normally distributed [1], hence nonparametric statistical tests need to be used to test the significance of the results. Therefore, in Section IV and Section V, nonparametric statistical tests are discussed. Wilcoxon signed rank test is chosen for pairwise comparison, while Friedman test is picked for multiple comparison. The paper is concluded in Section VI.

Manuscript received December 30, 2014; revised March 5, 2015. This research is funded by the Department of Higher Education of Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2012/TK06/MMU/03/7) and the UM-UMRG Scheme (CG031-2013) and UM-Postgraduate Research Grant (PG097-2013A).

Nor Azlina Ab Aziz is with the Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia and Faculty of Engineering and Technology, Multimedia University, 75450 Melaka, Malaysia (e-mail: azlina.aziz@mmu.edu.my).

Marizan Mubin is with the Department of Electrical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia (e-mail: marizan@um.edu.my).

Zuwairie Ibrahim is with the Faculty of Electrical and Electronic Engineering, University Malaysia Pahang, 26600 Pahang, Malaysia (e-mail: zuwairie@ump.edu.my).

Sophan Wahyudi Nawawi is with the Department of Control and Mechatronic Engineering, Universiti Teknologi Malaysia, 81310 Johor, Malaysia (e-mail: sophan@fke.utm.my).

## II. DATA REPRESENTATION – BOXPLOT

Boxplot is commonly chosen in this area to represent the collected data [2]–[6]. A boxplot represent numerical data using their quartiles [7].

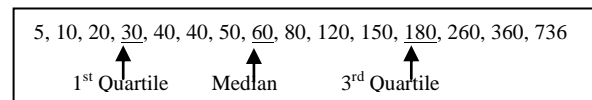


Fig. 1. Data.

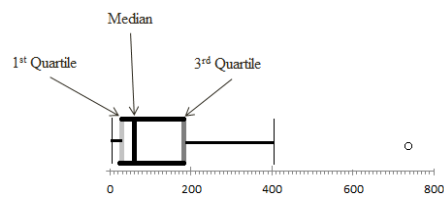


Fig. 2. Usage of quartiles in boxplot.

A boxplot construction starts with sorting the data in ascending order and identification of the median, the first quartile, Q1, and third quartile, Q3 (Fig. 1). These data is then marked and the box is formed between the third quartile and the first quartile (Fig. 2). The minimum whisker is drawn based on the following;

$$\max(\text{smallest data}, Q1 - 1.5 \times IQR) \quad (1)$$

where  $IQR$  is the interquartile range, which can be calculated as follow;

$$IQR = Q3 - Q1 \quad (2)$$

The maximum whisker is drawn based on the following condition;

$$\min(\text{biggest data}, Q3 + 1.5 \times IQR) \quad (3)$$

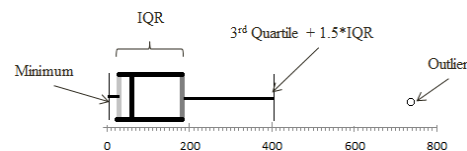


Fig. 3. Outliers.

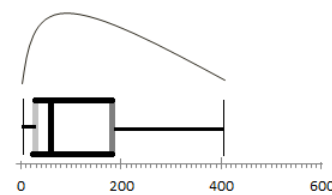


Fig. 4. Data normality.

Any values beyond the whiskers are outliers, which are out of ordinary observations (Fig. 3). The boxplot shown here is drawn horizontally, however it can also be drawn vertically.

The boxplot can be used to approximate the normality of

the data. Based on this information suitable statistical test can be properly selected. A boxplot with longer tail towards a direction show the skewedness of the data distribution towards that particular direction (Fig. 4). Therefore, the data is not normally distributed, thus, nonparametric statistical test need to be chosen in analyzing the data.

Other than analyzing normality of the data distribution, boxplot can also be used to indicate the stability of the algorithm performance. A smaller box indicate smaller variance in the data hence a stable performance.

The location of the box plot also indicates the quality of the algorithm. In minimization function optimization, a lower location of the box indicates better quality of solutions. It is important for researchers to take note on the presence of outliers, as they can be indicators whether the proposed algorithm is problematic or not.

III. STATISTICAL TEST — PRELIMINARIES

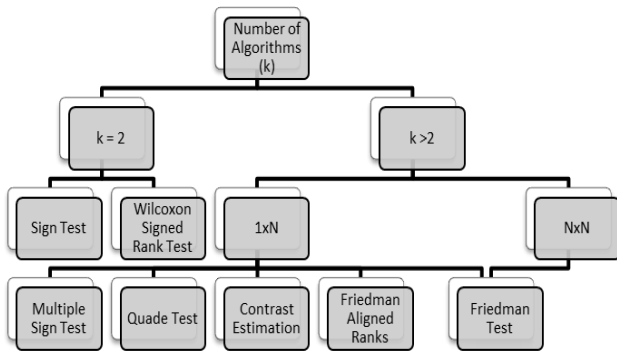


Fig. 5. Options of statistical test procedures based on categories.

Selection on the type of statistical test to be used depends on number of algorithms to be compared,  $k$ . It can be either comparison of two algorithms – pairwise comparison – or more than two algorithms. If more than two algorithms are compared, the choice of test depends on whether the comparisons are  $1 \times N$  – control method vs other algorithms, or  $N \times N$  comparisons, where all algorithms tested are compared with each other. Proposed algorithm is often chosen as the control method in  $1 \times N$  comparison. Multiple choices exist for each of the categories [1]. Among the statistical procedures commonly chosen are shown in Fig. 5.

The null hypotheses of the tests state that the algorithms compared are statistically equivalent, with no significant difference. For statistical test involving multiple algorithms, a post hoc procedure should be conducted if a significant difference is detected. A post hoc procedure helps to identify which algorithms caused the rejection of the null hypothesis.

Once a suitable test is identified the significant level,  $\alpha$  needs to be determined. This value indicates the significant difference of the algorithms. The common value chosen for  $\alpha$  is 0.05, other values are 0.1 and 0.01. A bigger  $\alpha$  has a higher chance of claiming significant difference, while a lower  $\alpha$  is more rigid in detecting significant difference.

An algorithm performance is commonly studied using a set of  $n$  benchmark functions over several runs. Hence, each of the functions will have set of data for every algorithm. The mean values of these data are used for the statistical test.

IV. PAIRWISE NONPARAMETRIC STATISTICAL TEST: WILCOXON SIGNED RANK TEST

In this paper, the Wilcoxon signed rank test is chosen as a pairwise statistical analysis tool. The test starts with finding the differences between the performances of the two algorithms in each test functions,  $d_i$ . The next step is to rank  $d_i$ 's based on their absolute value. In event of ties between two  $d_i$  values, the rank is evenly assigned for each fitness functions. After that, the sum of ranks where the first algorithm outperforms the second,  $R^+$ , and the sum of ranks where the second algorithm is the best,  $R^-$ , are calculated as follow;

$$R^+ = \sum_{d_i > 0} rank(d_i) + \frac{1}{2} \sum_{d_i = 0} rank(d_i) \tag{4}$$

$$R^- = \sum_{d_i < 0} rank(d_i) + \frac{1}{2} \sum_{d_i = 0} rank(d_i) \tag{5}$$

The smallest value between  $R^+$  and  $R^-$  is then chosen as the test statistic,  $T$ .

$$T = \min(R^+, R^-) \tag{6}$$

The null hypothesis is rejected if the test statistic value is lesser than critical value,  $T_0$ . These critical values are listed in Table I below. The application of Wilcoxon signed rank test is shown in example 1.

TABLE I: CRITICAL VALUES OF  $T_0$  IN WILCOXON SIGNED RANK TEST [7]

$\alpha$	0.10	0.05	0.01	$\alpha$	0.10	0.05	0.01
$n = 5$	1			$n = 28$	130	117	92
$n = 6$	2	1		$n = 29$	141	127	100
$n = 7$	4	2		$n = 30$	152	137	109
$n = 8$	6	4	0	$n = 31$	163	148	118
$n = 9$	8	6	2	$n = 32$	175	159	128
$n = 10$	11	8	3	$n = 33$	188	171	138
$n = 11$	14	11	5	$n = 34$	201	183	149
$n = 12$	17	14	7	$n = 35$	214	195	160
$n = 13$	21	17	10	$n = 36$	228	208	171
$n = 14$	26	21	13	$n = 37$	242	222	183
$n = 15$	30	25	16	$n = 38$	256	235	195
$n = 16$	36	30	19	$n = 39$	271	250	208
$n = 17$	41	35	23	$n = 40$	287	264	221
$n = 18$	47	40	28	$n = 41$	303	279	234
$n = 19$	54	46	32	$n = 42$	319	295	248
$n = 20$	60	52	37	$n = 43$	336	311	262
$n = 21$	68	59	43	$n = 44$	353	327	277
$n = 22$	75	66	49	$n = 45$	371	344	292
$n = 23$	83	73	55	$n = 46$	389	361	307
$n = 24$	92	81	61	$n = 47$	408	379	323
$n = 25$	101	90	68	$n = 48$	427	397	339
$n = 26$	110	98	76	$n = 49$	446	415	356
$n = 27$	120	107	84	$n = 50$	466	434	373

1) Example 1

Two algorithms are tested here ( $k=2$ ), asynchronous particle swarm optimization (A-PSO) and synchronous-asynchronous PSO (SA-PSO). The particles in A-PSO are updated asynchronously where a particle immediately moves to new location once its performance is measured [8]. This is done without the need to wait for the entire swarm to finish their performance evaluation (Fig. 6). The SA-PSO algorithm is hybrid of the standard PSO with A-PSO (Fig. 7) [9]. The swarm in SA-PSO is divided into groups of particles. All particles within a group are evaluated first before they move to new position. However, these particles movement is independent of particles from other groups. The performance of the algorithms is compared using set of nine benchmark functions ( $n=9$ ). The functions tested are minimization functions consisting of unimodal and multimodal functions. The functions are – quadric, quartic, hyperellipsoid, Rosenbrock, sphere, Ackley, Griewank, Rastrigin and Salomon functions (Table II).

```

while not stopping condition
  for each particle
    evaluate performance
    identify best values
    move //update velocity & position
  end
end
    
```

Fig. 6. A-PSO algorithm.

```

while not stopping condition
  for each group
    for each member
      evaluate performance
    end
    identify best values
    for each member
      move //update velocity &
      position
    end
  end
end
end
    
```

Fig. 7. SA-PSO algorithm.

TABLE II: TEST FUNCTIONS

Function Type	Function Name	Equation
Unimodal	Quadric	$f_1(x) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j\right)^2$ where $-100 \leq x_j \leq 100$
	Quartic	$f_2(x) = \sum_{i=1}^n i x_i^4$ where $-1.28 \leq x_i \leq 1.28$
	Rosenbrock	$f_3(x) = \sum_{i=1}^{n-1} \left[100 \left(x_i^2 - x_{i+1}\right)^2 + \left(x_i - 1\right)^2\right]$ where $-2.048 \leq x_i \leq 2.048$
	Spherical/ De Jong's	$f_4(x) = \sum_{i=1}^n x_i^2$ where $-5.12 \leq x_i \leq 5.12$
	HyperEllipsoid	$f_5(x) = \sum_{i=1}^n i x_i^2$ where $-5.12 \leq x_i \leq 5.12$
Multimodal	Ackley	$f_6(x) = 20 + e - 20 \exp \left[ -0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right] - \exp \left[ \frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i) \right]$ where $-32.768 \leq x_i \leq 32.768$
	Griewank	$f_7(x) = 1 + \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos \left( \frac{x_i}{\sqrt{i}} \right)$ where $-600 \leq x_i \leq 600$
	Rastrigin	$f_8(x) = 10n + \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i)]$ where $-5.12 \leq x_i \leq 5.12$
	Salomon	$f_9(x) = 1 - \cos \left( 2\pi \sqrt{\sum_{i=1}^n x_i^2} \right) + 0.1 \sqrt{\sum_{i=1}^n x_i^2}$ where $-600 \leq x_i \leq 600$

The algorithms are tested using each function for 500 runs. The averages of the results are recorded in Table III. The significance level,  $\alpha$ , is set to 0.05.

The statistic value,  $T$ , of Wilcoxon signed rank test obtained is compared with  $T_0$  from table of critical values of  $T_0$  in Wilcoxon signed rank test, which can be found in [7]. It is found that  $T < T_0$ , therefore, significance difference exist,

indicating SA-PSO performs better than A-PSO.

V. MULTIPLE ALGORITHMS NONPARAMETRIC STATISTICAL TEST: FRIEDMAN TEST AND ITS POST HOC

Friedman test can be used for both  $1 \times N$  and  $N \times N$  multiple algorithms comparison. The test is done by ranking the

algorithms based on the performance for each benchmark function. The ranks of each algorithm are summed and averaged,  $R_j$ . The rank is then used to calculate the Friedman statistic value,  $F_f$ .

$$F_f = \frac{12n}{k(k+1)} \left[ \sum_j (R_j)^2 - \frac{k(k+1)}{4} \right] \quad (7)$$

This statistic value is then compared with  $\chi^2$  value. The  $\chi^2$  value is taken using  $(k-1)$  degree of freedom and  $\alpha$  degree of rejection. The null hypothesis is rejected if  $F_f \geq \chi^2$ .

TABLE III: A-PSO vs SA-PSO

	A-PSO	SA-PSO	$d_i$	rank( $d_i$ )
Quadric	131.4246	0.0537	-131.371	8
Quartic	3.0793	0	-3.0793	5
Rosenbrock	71.5899	37.9161	-33.6738	7
Spherical	0.1628	0	-0.1628	3
HyperEllipsoid	2.5037	0	-2.5037	4
Ackley	0.0898	0	-0.0898	2
Griewank	371.7447	0.0071	-371.738	9
Rastrigin	42.2694	36.1207	-6.1487	6
Salomon	0.3211	0.3263	0.0052	1

The rejection of null hypothesis indicates that significant difference exist between at least two of the algorithms. However Friedman test alone cannot identify where is the difference. Therefore post hoc procedure of pairwise comparison needs to be conducted to identify the difference.

1) Example 2

Three algorithms are tested here ( $k=3$ ), synchronous PSO (S-PSO), A-PSO and SA-PSO. S-PSO is the standard PSO [10]. The whole particles in S-PSO are evaluated first before moving to their respective new position (Fig. 8). The algorithms are tested using the same benchmark functions as example 1.

```

while not stopping condition
  for each particle
    evaluate performance
  end
  identify best values
  for each particle
    move //update velocity & position
  end
end
    
```

Fig. 8. S-PSO algorithm.

The performances of the algorithms for all test functions are ranked based on their mean value. The average performance of the algorithms for each test function and their rank are tabulated in Table IV. According to the Friedman test's statistic value, significant difference exists between the algorithms. Therefore, post hoc procedure needs to be conducted to identify the difference.

A. Holm Post Hoc Analysis

Post hoc procedure to identify the differences must be chosen based on the main statistical test used. For Friedman test there are selections of post hoc procedures (Fig. 9). Holm procedure is among the suitable post hoc procedures for

Friedman test. It can be used for both  $1 \times N$  comparison and  $N \times N$  comparison. However, there is slight difference between Holm procedure for  $1 \times N$  comparison and  $N \times N$  comparison.

TABLE IV: FRIEDMAN TEST OF S-PSO, A-PSO AND SA-PSO

		S-PSO	A-PSO	SA-PSO
Quadric	Mean	305.4320	131.4246	0.0537
	Friedman Rank	3	2	1
Quartic	Mean	1.9524	3.0793	0.0000
	Friedman Rank	2	3	1
Rosenbrock	Mean	58.7646	71.5899	37.9161
	Friedman Rank	2	3	1
Spherical	Mean	0.0963	0.1628	0.0000
	Friedman Rank	2	3	1
HyperEllipsoid	Mean	0.4151	2.5037	0.0000
	Friedman Rank	2	3	1
Ackley	Mean	0.0941	0.0898	0.0000
	Friedman Rank	3	2	1
Griewank	Mean	317.8628	371.7447	0.0071
	Friedman Rank	2	3	1
Rastrigin	Mean	38.6035	42.2694	36.1207
	Friedman Rank	2	3	1
Salomon	Mean	0.3227	0.3211	0.3263
	Friedman Rank	2	1	3
Average Friedman Rank, $R_j$		2.22	2.56	1.22

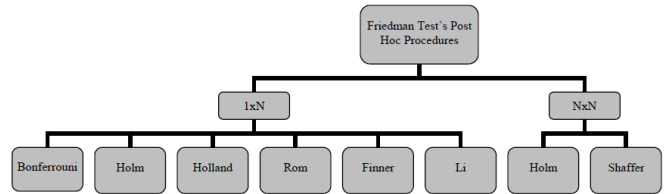


Fig. 9. Friedman test's post hoc procedure.

The statistic value for Holm procedure is calculated as follow;

$$z = \frac{R_i - R_j}{\sqrt{k(k+1)/6n}} \quad (8)$$

$R_i$  and  $R_j$  in the equation are Friedman rank for algorithm  $i^{th}$  and  $j^{th}$ . This statistic value is then used to get the p-value from two tails normal distribution table.

The p-values are then sorted in ascending order such that  $p_1 < p_2 < p_3 < \dots < p_l < \dots < p_k$ . Each of the p-value corresponds to their own null hypothesis,  $H_1, H_2, H_3, \dots, H_l, \dots, H_k$ , which state that the algorithms are statistically identical.

In Holm of  $1 \times N$  comparison the null hypotheses from  $H_1$  to  $H_l$  are rejected if;

$$p_l \geq \frac{\alpha}{k-l} \quad (9)$$

While for Holm of  $N \times N$  comparison the null hypotheses from  $H_1$  to  $H_l$  are rejected if;

$$p_l \geq \frac{\alpha}{m-l+1} \quad (10)$$

where

$$m = \frac{k(k-1)}{2} \tag{11}$$

2) Example 3

The Friedman test in example 2 shows that significant difference exist between the three algorithms. Thus, post hoc procedure is conducted to identify the difference.

Holm procedure for  $1 \times N$  comparison is conducted first. The SA-PSO is chosen as the control method. Table V shows the data of Holm  $1 \times N$  procedure. The result couldn't identify where does the significant difference occurs. This is not uncommon. Therefore Holm post hoc for  $N \times N$  comparison is done (Table VI).

The Holm post hoc for  $N \times N$  comparison successfully detects the significant difference. The null hypothesis,  $H_1$  and  $H_2$  are rejected. Significant difference exists between the performance of SA-PSO and S-PSO and also A-PSO, where SA-PSO has a better performance than both algorithms.

TABLE V: HOLM POST HOC FOR  $1 \times N$  COMPARISON

<i>i</i>	DataSet	<i>z</i>	<i>p</i>	Holm	Observation
1	A-PSO	2.8284	0.0047	0.0250	$p_1 < Holm$
2	S-PSO	2.1213	0.0340	0.0500	$p_2 < Holm$

TABLE VI: HOLM POST HOC FOR  $N \times N$  COMPARISON

<i>i</i>	DataSet	<i>z</i>	<i>p</i>	Holm	Observation
1	A-PSO vs. SA-PSO	2.8284	0.0047	0.0167	$p_1 < Holm$
2	S-PSO vs. SA-PSO	2.1213	0.0339	0.0250	$p_2 > Holm$
3	S-PSO vs. A-PSO	0.7071	0.4795	0.0500	

VI. CONCLUSION

Proper usage of statistical tools in data analysis is important in strengthening and justifying findings of a research. Hence, essential statistical tools to analyze performance of swarm intelligence algorithms are discussed in this paper. The paper starts with description of boxplot which is a common method chosen in data representation for swarm intelligence algorithms. The information represented by the boxplot is also discussed. The Wilcoxon signed rank test for pairwise comparison and Friedman test together with Holm post hoc procedure for multiple algorithms comparison are presented.

ACKNOWLEDGMENT

This research is funded by the Department of Higher Education of Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2012/TK06/MMU/03/7) and the MMU Mini Fund (MMU/RMC/MiniFund/FET/2014-2015/01) and UM-Postgraduate Research Grant (PG097-2013A). The authors would also like to thank the reviewers for their invaluable feedback and also our friends and colleagues for their help and ideas in the process of completing this paper.

REFERENCES

- [1] J. Derrac, S. Garcia, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3-18, Mar. 2011.
- [2] J. Rada-Vilela, M. Zhang, and W. Seah, "A performance study on the effects of noise and evaporation in Particle Swarm Optimization," in *Proc. 2012 IEEE Congress of Evolutionary Computation*, 2012, pp. 1-8.
- [3] J. Rada-Vilela, M. Zhang, and W. Seah, "A performance study on synchronicity and neighborhood size in particle swarm optimization," *Soft Computing*, vol. 17, no. 6, pp. 1019-1030, Feb. 2013.
- [4] J. Rada-Vilela, M. Zhang, and W. Seah, "A performance study on synchronous and asynchronous updates in particle swarm optimization," in *Proc. 13th Annual Conference on Genetic and Evolutionary Computation, Ireland*, 2011, pp. 21-28.
- [5] J. Rada-Vilela, M. Zhang, and W. Seah, "Random asynchronous PSO," in *Proc. 5th International Conference on Automation, Robotics and Applications*, 2011, pp. 220-225.
- [6] A. Campos, A. Pozo, and E. Duarte, "Evaluation of asynchronous multi-swarm particle optimization on several topologies," *Concurrency and Computation Practice and Experience*, vol. 25, no. 8, pp. 1057-1071, July 2012.
- [7] W. Mendenhall and T. Sincich, *Statistics for Engineering and The Sciences*, 5th ed., United States of America, Pearson, 2007, p. 926.
- [8] A. Carlisle and G. Dozier, "An off-the-shelf PSO," in *Proc Workshop on Particle Swarm Optimization*, 2001, pp. 1-6.
- [9] N. A. A. Aziz, M. Mubin, M. S. Mohamad, and K. A. Aziz, "A synchronous-asynchronous particle swarm optimisation algorithm," *The Scientific World Journal*, p.17, July 2014.
- [10] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Proc. IEEE World Congress of Computational Intelligence*, May 1998, pp. 69-73.



Nor Azlina Ab Aziz holds a master and bachelor's degrees in engineering. She is currently pursuing her PhD in the Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Malaysia. Her research interest includes swarm intelligence and its application in solving communication optimization problems.



Marizan Mubin graduated with an honors degree in telecommunication engineering, from University of Malaya in 2000 and she had continued to post-graduate studies in University of Newcastle Upon Tyne, UK (the MSc in communication and signal processing). In 2003, she was awarded a scholarship by Japan International Cooperation Agency (JICA) to pursue a doctoral degree in Tokai University, Japan. She is currently attached to the Department of Electrical Engineering, University of Malaya as a senior lecturer. Her main research interest is in non-linear control systems.



Zuwairie Ibrahim received his B.Eng. and M.Eng. degrees from Universiti Teknologi Malaysia, in 2000 and 2002, respectively. In 2006, he has been awarded a Ph.D. degree from Meiji University, Japan. He is currently an associate professor in the Faculty of Electrical and Electronic Engineering, Universiti Malaysia Pahang. His research interests include computational intelligence, image processing, and unconventional computation such as molecular or DNA computing.



Sopnan Wahyudi Nawawi joined Universiti Teknologi Malaysia since 1999. He received his PhD degree from the Universiti Teknologi Malaysia in 2010. In 2006, he collaborates with control research group at HKUST as a research scholar and attached to HKUST spin off company Googol Technology (HK) ltd. He was promoted to a senior lecturer in 2010.