

## Binary Particle Swarm Optimization for Variables Selection Optimization in Taguchi's T-Method

<sup>1</sup>N. Harudin\*, <sup>2</sup>K. R. Jamaludin, <sup>3</sup>F. Ramlie, <sup>4</sup>M. N. Muhtazaruddin,  
<sup>5</sup>Che Munira Che Razali and <sup>6</sup>W. Z. A. W. Muhamad

<sup>1,2,3,4</sup>Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia  
Jalan Sultan Yahaya Petra, 54100 Kuala Lumpur

<sup>5</sup>Malaysia-Japan international Institute of Technology, Universiti Teknologi Malaysia  
Jalan Sultan Yahaya Petra, 54100 Kuala Lumpur

<sup>1</sup>Department of Mechanical Engineering, Universiti Tenaga Nasional  
43000 Kajang Selangor, Malaysia

<sup>6</sup>Institute of Engineering Mathematics, Universiti Malaysia Perlis  
Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia

\*Corresponding author: nolia.harudin@gmail.com

### Article history

Received: 2 April 2019

Received in revised form: 13 November 2019

Accepted: 27 January 2020

Published online: 1 April 2020

---

**Abstract** Prediction analysis has drawn significant interest in numerous fields Taguchi's T-Method is a prediction tool that was practically developed to predict even with a limited sample data. It was developed explicitly for multivariate system prediction by relying on historical data as the baseline model and adapting the signal to noise ratio (SNR) as well as zero proportional concepts in strengthening its robustness. Orthogonal array (OA) in T-Method is a variable selection optimization technique in improving the prediction accuracy as well as helping to eliminate variables that may deteriorate the overall performance. However, the limitation of OA in dealing with higher multivariate data restrains the optimization accuracy. The binary particle swarm optimization used in this study helps to cater to the limitation of OA as well as optimizing the variable selection process for better prediction accuracy. A comparison between the T-Method+OA and T-Method+BPSO in four different case studies shows that the T-Method+BPSO performs better with a higher coefficient of determination ( $R^2$ ) value and means relative error (MRE) value compared to the T-Method+OA. The T-Method with the BPSO element as variables screening optimization is able to increase or even maintain the prediction accuracy for cases that are normally distributed, have a high  $R^2$  value, and with low sample data.

**Keywords** Binary Particle Swarm Optimization; Optimum variable selection; Small sample data; Prediction; Taguchi's T-Method

**Mathematics Subject Classification** 90C90

## 1 Introduction

Disruption is a daily challenge in ensuring a reliable system with responsive risk control towards a robust design and a robust operating system. Uncertainty in various areas has led to an unpredictable future industrial system. Most of the current modes of uncertainty have specific causes and are commonly treated as one of the influential factors towards better prediction accuracy. Leveraging and managing it properly with advanced cognitive and predictive modeling and algorithms for a better system has been discussed progressively among researchers and industrial practitioners. Several practitioners in various fields have used various prediction tools for the past few decades and have progressively enhanced them. In the last few decades, machine learning has been found to be the preferable approach in prediction analysis compared to the statistical approach. This view is supported by a recent study carried out by Hussain *et al.* [1] which found that machine learning based approaches are performing better than time series-based approaches in predicting the future quality of services in order to avoid service violation and violation penalty.

Taguchi's T-Method or commonly called the T-Method is one of the prediction tools governed under the Mahalanobis-Taguchi System (MTS) that was developed explicitly for prediction and optimization purposes by relying on historical data. One of the significant contributions of Taguchi's T-Method is its ability to predict even with limited sample data. If multiple regression analysis has a limitation that the sample size must be larger than the number of items, the Taguchi's T-Method does not have this limitation [2–4]. The smallest sample data that have been used in Taguchi's T-Method analysis is the yield prediction [2], which involved seven training data and two test data. Unlike other optimization methods, an element of the signal to noise ratio (SNR) within MTS supports the robustness and insensitiveness to the variation in the multidimensional system. In a multidimensional system, it is essential to preserve only relevant variables rather than redundant variables that may deteriorate performance so that the loss of valuable information will be minimised.

The element of the orthogonal array (OA) within MTS has been debated and is believed to be insufficient as it offers a sub-optimal solution. Abraham and Variyath [5] argued that a fixed combination in OA is not optimal since the results may vary significantly if the column-to-column information is rearranged. Foster *et al.* [6] agreed with Abraham and Variyath [5] after proving the fact with 1000 random variables to column assignment. Another limitation of OA can be seen when there is a need to analyze a large number of variables [5]. Issues in OA have been highlighted as well by [5], [8–9], especially the fact that the OA design has limitations in handling the higher-order interaction between variables, which might lead to an inconsistency in the identification of the crucial variables. Therefore, developing a hybrid methodology for a better accuracy gain is the solution to these concerns.

Up until recently, the element of OA in the MT-Method has been continuously enhanced with various machine-learning algorithm approaches. Unlike the MT-Method, enhancement of the OA element within the T-Method for variables optimization is still at an initial stage but Harudin *et al.* [10] did enhance the OA in the T-Method with an artificial bee colony algorithm (ABC) and the results showed that T-Method+ABC performed better compared to T-Method+OA for the specific case study conducted. There are several other studies done with the same intention which are yet to be published but have been shared in several forums.

Iquebal *et al.* [7] compared MT-ABC, MT-Particle Swarm Optimization (PSO), and MT-

Genetic Algorithm (GA) in their study on several benchmark datasets and MT-PSO was found to be the best in terms of running time with very minimal differences in terms of its accuracy compared to MT-ABC. Several other studies whether within MTS or beyond the MTS nature, also highlighted that Binary PSO converges faster compared to other algorithms. Up until recently, many BPSO algorithms have been progressively enhanced and proven to provide better accuracy [11–15]. In the area of the T-Method, BPSO has not been explored yet, and it serves as a motivation for this research study to be conducted. The objective of this research is, therefore, to integrate binary particle swarm optimization (BPSO) with the T-Method by replacing the OA for variables optimization.

## 2 Related Studies

The application of OA within the T-Method has similarities with the MT-Method. Due to minimal resources within the T-Method itself, the authors are relying on the progress of OA enhancement within the MT-Method for practical purposes. The most recent study conducted by Mota-Gutiérrez *et al.* [16] which summarised the 18 year progress of the MT-method in various industrial practices was found to be very helpful for this study since the area of dimensional reduction or variables optimization by several researchers are still being progressively enhanced till recently [7], [17–18]. This study provides a conceptual idea in applying the same concept to T-Method variable screening or optimization since very few studies have been conducted within that area. Khanesar [19] introduced the novel binary particle swarm optimization (BPSO) in 2007 and it was adapted by Pal and Maiti [20] for dimensional reduction within the MT-Method. Reséndiz and Rull-Flores [21] implemented the Pal and Maiti [20] work as part of a comparison between binary ant colony optimization (BACO) and Gompertz binary PSO (GBPSO). The results concluded that GBPSO converges faster compared to the other two algorithms.

Since the T-Method has been progressively explored up until recently, most of the cases use the T-Method in real life prediction problems [22–24]. Inoh [25] improved the method to define the unit space theory into  $T_a$  and  $T_b$  methods while [3], [4], [26–29] focused on improving the baseline model accuracy for outliers and abnormal case data. Harudin *et al.* [10] is the only published work outside Japan currently that improved the variables selection optimization in the T-Method using ABC. The findings were concluded to be case-to-case dependent. This research was intentionally conducted to implement the theory of BPSO into the T-Method as part of the improvement approach to enhance the limitation of the orthogonal array within the T-Method as currently practiced. However, this study will not focus on the algorithm convergence time. It is more focused towards examining the accuracy of the predicted data on MRE and  $R^2$  across four different case studies.

## 3 Methodology

This section will explain the theory of the T-Method and Binary Particle Swarm Optimization (BPSO) together with the absorption of BPSO into the T-Method for variables optimization and prediction purposes.

### 3.1 T-Method

Emphasizing on three core elements (sensitivity, linearity, and variability), signal to noise ratio (SNR) is the critical element that has been adapted into the T-Method formulation together with the concept of Mahalanobis Distance (MD) theory as a crucial aspect in defining the reference data. The regression model is formulated linearly and relies on the Zero-proportional concept with the linear regression line passing through zero-point (origin). The establishment of unit space and signal space for the normalization analysis is the core element of MD in the T-Method. The selection of unit space is a crucial predefined decision before making further analysis. It must be as homogenous as possible since high-density population data representation requires particular expertise to accurately define the target region for the unit space selection. Taguchi stated that the selection of unit space is always in the middle position between the lowest and highest data of the selected region if following the rules [2]. Figure 1 illustrates the unit space concept more clearly. Inoh [24] improved the way to define unit space with the theory of  $T_a$  and  $T_b$  methods.  $T_a$  relies on the overall average input value and output as baseline data while  $T_b$  relies on samples that contribute to higher SNR value. We keep the unit space concept definition as per the actual theory for this research. By fulfilling the rule of unit space, reliable prediction accuracy and justification is achievable. We need to put in some effort towards understanding the historical data trend.

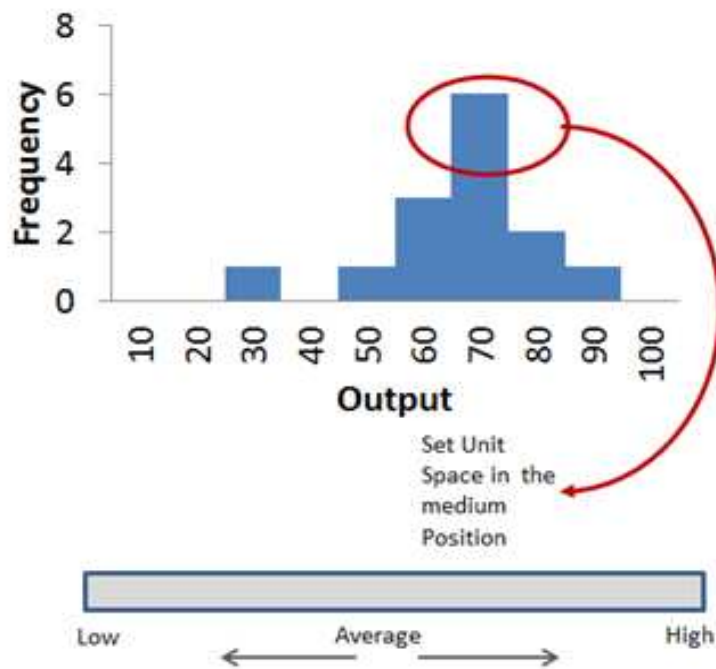


Figure 1: The Concept of Unit Space Selection for T-Method

The main idea of this theory is to fulfill the prediction model, which is the integrated estimate output value ( $\hat{M}_i$ ), as in Equation (1). In order to do that, the proportional coefficient ( $\beta$ ) and SNR ratio ( $\eta$ ) will have to be computed on an item by item basis using normalized data ( $X_{ij}, M_0$ ) calculated by Equation (2). The determination of unit space data is conducted from the raw data taken, which is then excluded and averaged out from the raw data. Whatever data left will be treated as signal data. This is how Equation (2) can be used. Equation (3)

to Equation 9 are the theory for calculating the proportional coefficient  $\beta$ ) and SN ratio ( $\eta$ ) to be filled into Equation (1). If the value of the SN ratio ( $\eta$ ) calculated is negative, it will be assumed to be zero. It is clearly seen that a higher SN ratio of an item will result in a higher degree of contribution to the overall model estimation. SNR here is called duplicate SNR since it represents weightage to the overall equation.

$$\hat{M}_i = \frac{\eta_1 \times \left(\frac{X_{i1}}{\beta_1}\right) + \eta_2 \times \left(\frac{X_{i2}}{\beta_2}\right) + \dots + \eta_j \times \left(\frac{X_{ij}}{\beta_j}\right)}{\eta_1 + \eta_2 + \dots + \eta_j} \tag{1}$$

$$\text{normalised data } (X_{ij}, M_0) = \text{signal data} - \text{average of unit space} \tag{2}$$

$$\text{Effective Divider, } r = M_1^2 + M_2^2 + \dots + M_i^2 \tag{3}$$

$$\text{Total Variation, } S_T = X_1^2 + X_2^2 + \dots + X_i^2 \tag{4}$$

$$\text{Variation of proportional term, } S_\beta = \frac{(M_1 X_{11} + M_2 X_{21} + \dots + M_l X_{l1})^2}{r} \tag{5}$$

$$\text{Error variation, } S_e = S_T - S_\beta \tag{6}$$

$$\text{Error Variance, } V_e = \frac{S_e}{l - 1} \tag{7}$$

Duplicate SN ratio,

$$\eta = \left(\frac{(S_\beta - V_e)}{r V_e}\right) \tag{8}$$

$$\beta_M = \frac{M_1 X_{11} + M_2 X_{21} + \dots + M_l X_{l1}}{r} \tag{9}$$

By fulfilling Equation (1), the overall prediction can be achieved by considering all variables. In predicting the unknown future output, the known input can be used in Equation (1) by maintaining the value of the SNR, proportional coefficient, and replacing the known input data. Considering the need for optimizing the variables selection for better prediction and cost-saving purposes as mentioned in the previous section, the orthogonal array is the current method used within the T-Method. Below are the formulas to calculate the SN ratio,  $\eta$  (B) for optimization on feature selection:

$$\text{Linear equation, } L = M_1 \hat{M}_1 + M_2 \hat{M}_2 + \dots + M_i \hat{M}_i \tag{10}$$

$$\text{Effective Divider, } r = M_1^2 + M_2^2 + \dots + \tag{11}$$

$$\text{Total Variation, } S_T = \hat{M}_1^2 + \hat{M}_2^2 + \dots + \hat{M}_i^2 \tag{12}$$

$$\text{Variation of proportional term, } S_\beta = \frac{L^2}{r} \tag{13}$$

$$\text{Error variation, } S_e = S_T - S_\beta \tag{14}$$

$$\text{Error Variance, } V_e = \frac{S_e}{n - 1} \tag{15}$$

$$\text{SN ratio, } \eta = 10 \log \left(\frac{(S_\beta - V_e)}{r V_e}\right). \tag{16}$$

Taguchi specifically used a two-level OA type within the MTS theory, which represents either to use the variable or not to use it. The type of two-level array selected depends on the total number of variables within the study. The linear equation, L, is the sum of multiplication between estimated output ( $\hat{M}_i$ ) for each sample calculated using Equation (1) with its actual output value ( $M_i$ ) as in Equation (10). Table 1 below shows an example of  $L_{12}$  combination. One means the variable for that particular level (row) and combination (column) is included as  $X_{ij}$  in Equation (1), while 2 means the value for that combination and level will be treated as zero (excluded) for  $X_{ij}$  in Equation (1). Each level of computation will rely on the number of samples in the normalized data in Equation (2). If the normalized data have five samples with five values of the  $\hat{M}_i$ , a total of 5 estimation output ( $\hat{M}_i$ ) is generated then for each level computation. Once the values of ( $\hat{M}_i$ ) and ( $M_i$ ) are specified, Equation (11) until Equation (16) are used to compute the dynamic SNR of each level computation. For the final optimization result, if the average value of SNR for 1 (used) is significantly higher compared to 2 (excluded) in each variables analysis, it shows that the contribution of that variable is relatively significant. The average unit space value of the output called ( $M_0$ ) will need to be added with the integrated estimate output value ( $\hat{M}_i$ ) in order to compute the integrated estimate value (Y) for unknown data as in Equation (17). At this stage, ( $\hat{M}_i$ ) is either considering all variables or only referring to the optimum variable selection. We can repeat the same procedure accordingly for the remaining prediction of unknown data.

$$\hat{Y}_1 = \hat{M}_1 + M_0. \tag{17}$$

Table 1:  $L_{12}$  Orthogonal Array Combination

	Control Factors											
Sample no.	A	B	C	D	E	F	G	H	I	J	K	SNR (db)
1	1	1	1	1	1	1	1	1	1	1	1	SNR1
2	1	1	1	1	1	2	2	2	2	2	2	SNR2
3	1	1	2	2	2	1	1	1	2	2	2	SNR3
4	1	2	1	2	2	1	2	2	1	1	2	SNR4
5	1	2	2	1	2	2	1	2	1	2	1	SNR5
6	1	2	2	2	1	2	2	1	2	1	1	SNR6
7	2	1	2	2	1	1	2	2	1	2	1	SNR7
8	2	1	2	1	2	2	2	1	1	1	2	SNR8
9	2	1	1	2	2	2	1	2	2	1	1	SNR9
10	2	2	2	1	1	1	1	2	2	1	2	SNR10
11	2	2	1	2	1	2	1	1	1	2	2	SNR11
12	2	2	1	1	2	1	2	1	2	2	1	SNR12
(Note: 1 “Item is used,” and 2 “Item is <i>not</i> used.”)												

### 3.2 Particle Swarm Optimization

Russell Eberhart and James Kennedy invented the Particle Swarm Optimization algorithm (PSO) in 1995 which was inspired by the flocking and schooling patterns of birds and fish [30–31]. This algorithm was initially proposed to graphically simulate the uniqueness of the flying pattern of the flocking birds. Each bird represents a particle that has its own position and velocity updated from time to time. The fitness function of all swarms are evaluated for each iteration. The velocity for each particle is updated based on its current position, the best position it has explored ( $P_{\text{best}}$ ) and the global best position ( $G_{\text{best}}$ ) discovered by the swarm. The best position between these three elements will finally be updated or maintained as the best  $G_{\text{best}}$  position in each iteration.

Below are the equations that show how a particle's velocity and position are updated.

$$v_i(t+1) = wv_i(t) + c_1 * rand_1(p_i - a_i(t)) + c_2 * rand_2(p_g - a_i(t)) \quad (18)$$

$$a_i(t+1) = a_i(t) + v_i(t+1) \quad (19)$$

Here  $rand_1$  and  $rand_2$  are random numbers within  $[0, 1]$  updated each time they occur, and  $w$  is known as 'inertia weight' that can control the impact of the previous velocity on any particle to its current one. If  $w > 1$  the particle favoured exploration overexploitation, else if  $w < 1$  the particle gave more importance to the current best positions. The  $c_1$  and  $c_2$  are positive constants, called 'acceleration coefficients'. From a psychological perspective,  $c_1$  is the cognitive component that measures the degree of self-confidence of a particle and measures the degree at which it trusts its performance. Meanwhile,  $c_2$  is the social component that relies on the capability of the swarm to find better candidate solutions [32]. The PSO algorithm is run through below pseudocode iteratively until the termination criterion is satisfied.

**Begin**

*Initialize swarm position  $a(0)$  and velocities  $v(0)$*

*Set iteration counter,  $t=0$*

**Repeat**

*Compute fitness function for each of the swarm*

**Begin** (perform PSO operation)

*Compute  $v(t+1)$  from Eq. (18)*

*Compute  $a(t+1)$  from Eq. (19)*

**End**

*Set  $t=t+1$*

**Until** *termination criteria satisfied*

**End**

Figure 2: Pseudocode of Particle Swarm Optimization

The original PSO is mainly applicable to continuous cases. However, for discrete cases, there might be a need for extra things to ponder. Due to the limitation of current PSO on discrete cases, Kennedy and Eberhart [30] has proposed a discrete binary version to tackle the discrete problem called binary particle swarm optimization (BPSO) which defines the position of the birds in terms of  $[0,1]$ . A detailed description of the PSO algorithm is given in [33]. The same goes for BPSO which is also presented in detail by [30] as well as other enhanced versions

of BPSO by [32], [34–36]. However, for this study, we apply the primary BPSO as explained in the next sub-section.

### 3.2.1 Binary Particle Swarm Optimization

The BPSO initializes the positions and velocities of the particle swarm randomly by using:

$$a_i = \begin{cases} 1, & \text{if } rand() > 0.5 \\ 0, & \text{Otherwise} \end{cases} \quad (20)$$

$$v_{ij} = -v_{max} + 2 * rand() * v_{max}. \quad (21)$$

The position  $a_i$  for each variable is calculated below:

$$a_i = \begin{cases} 1, & \text{if } S(v_i) > rand() \\ 0, & \text{Otherwise} \end{cases} \quad (22)$$

Here  $S(v_i)$  represents the logistic function, and it serves as the probability distribution for position  $a_i$ .

$$S(v_i) = \frac{1}{1 + \exp(-v_i)}. \quad (23)$$

The velocities  $v_i$  are then iteratively updated by Equation (21). The population size (N) used in this study is 20 with minimum and maximum inertia weight value of 0.4 and 0.9 respectively. The acceleration factor ( $C_1, C_2$ ) is two with 100 iterations and 20 runs.

### 3.2.2 Binary Particle Swarm Optimization as Variables Selection Optimization in T-Method

As stated earlier, fulfilling Equation (1) is enough to make any predictions considering all variables as necessary. However, if an optimization on variables is a requirement or if studying the effect of all variables to the overall prediction may deteriorate the overall performance, then optimization should be done. Knowing the limitation of OA in dealing with high-dimensional data, this is where the BPSO is considered into the loop of the T-Method replacing the OA. The objective function of this study is to maximize the SNR value in Equation (16). This study purposely maintains the SNR as the objective function, due to its insensitiveness to variation. The following paragraph gives a brief overview of how BPSO has been applied within this study.

The BPSO algorithm will generate an initial position randomly based on the population size, and a total number of variables are identified until the initial fitness value (SNR) is determined. The fitness function is calculated using Equation (10) to Equation (16). The Pbest and Gbest values are initialized, and then the BPSO concept shown in Equation (18) to Equation (21) is applied to define and randomize the new position and velocity updates. The fitness function is then recalculated using the new position and velocity updates, as well as the Pbest and Gbest. The process is repeated and updated until it reaches the termination criterion and converges to the same value of fitness function. In this analysis, as shown in Figure 3 below, the best position within each run (total of 20) consisting of 100 iterations in each run will be stored. The crucial choice of optimum variables depends on how many times the variables are selected from the total of 20 runs. The example in Figure 3 shows that out of six variables (# column), only four variables are vital to be considered.



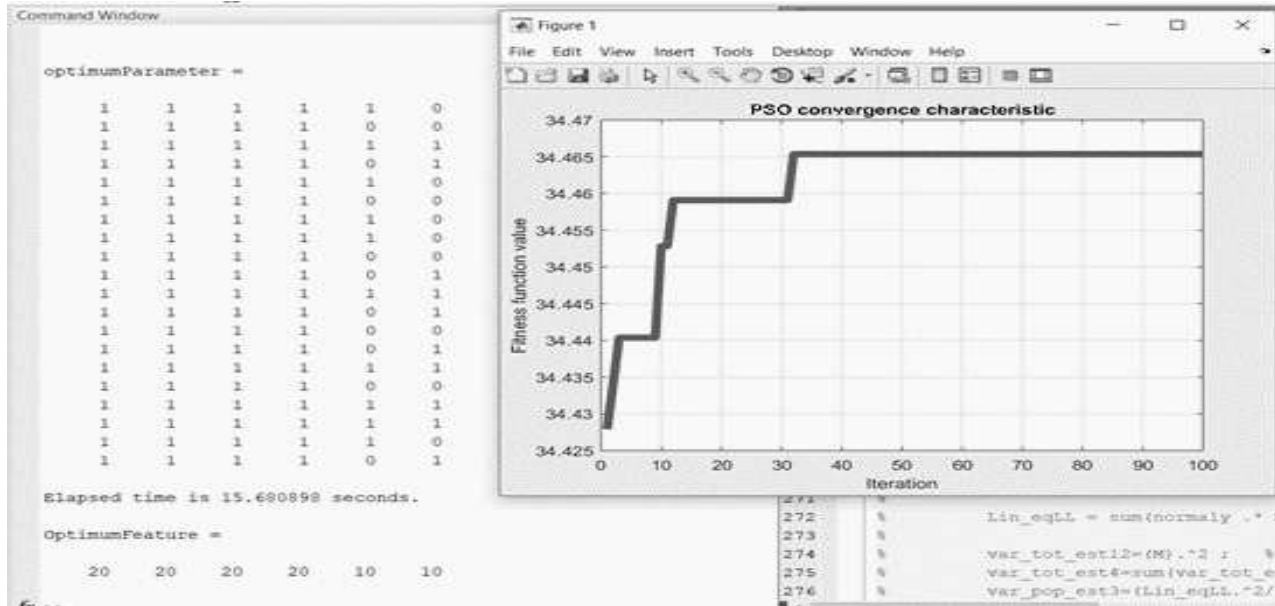


Figure 3: Result and Convergence Characteristics

## 4 Case Study Data Collection

The above discussed the hybrid methodology applied to several case studies which are estimations of manufacturing yield and energy performance of a residential building as well as a prediction on calories burned during ice skating activity and power consumption for a thermal energy storage system. The manufacturing yield prediction is an example case study taken from [2], which involved six relevant variables with seven sample data and one unknown to be estimated. The unknown data to be estimated across all case studies taken was initiated with known output value; however, to test the accuracy of the method, it was purposely reanalyzed and compared to measure the accuracy in terms of error calculation.

The dataset in estimating the energy performance of residential buildings is taken from the UCI machine learning database. It estimates the highest sample data and unknown data with 768 samples and eight variables. This case study provides a different overview of the T-Method itself since it involves a higher sample data. The outcome of this case study will be compared to the study conducted by [37] in comparing the same dataset with random forest (RF) and iterative reweighted least square (IRLS) methods.

The third case study of predicting calorie burn (kcal) in ice skating activity is a real case study conducted at the icescape ice rink at the IOI City Mall. A pedometer was given to each respondent who participated in this study to be worn before skating. A simple questionnaire was given to them as well to acquire some useful information for use in the analysis. Overall, 17 respondents were chosen as the sample with four respondents as unknown data to be estimated and 15 variables taken into consideration. Figure 4 below shows some samples of the questions asked in the questionnaire. The value stated in the bracket on the right side of each answer selection represents the rate used for the T-Method analysis.

The final case study was also taken from real life data collection involving the prediction of power consumption for a thermal energy storage system (TES). Data on 14 days charging

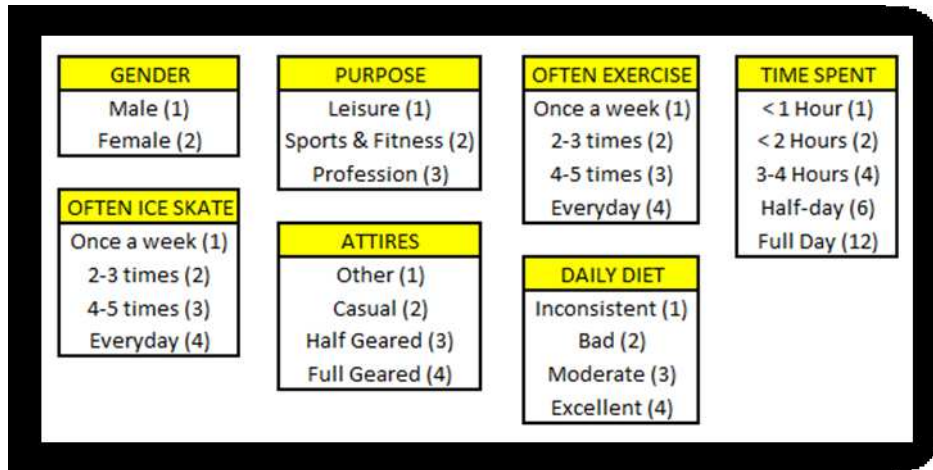


Figure 4: Sample of Questions Included in the Questionnaire

state condition is used with 13 days of data as the unknown to be estimated with a total of 15 variables. The TES acts like a battery to the air-conditioning system within the area which helps in reducing energy consumption by having high consumption during off-peak hours since charging is happening during the night shift. Detailed parameters of all case studies are shown in Table 2.

Table 2: List of Parameters

Case study: Estimation of the energy performance of residential building	Case study: estimation of Calories burned (kcal) for ice skating activity	Case study: Estimation of power consumption for thermal energy storage
Relative Compactness	Hear rate (bpm)	DP speed (%)
Surface Area ((m)	Step	Diff press (Bar)
Wall Area (m)	Age	STL-Tempsupply (deg)
Roof Area (m)	Gender	STL-Tempreturn (deg)
Overall Height (m)	Weight (kg)	STL-PHE supply (de)
Orientation	Height (m)	CH1-Temp supply (deg)
Glazing Area (%)	Often skating	CH1-Temp return (deg)
Glazing Area Distribution (%)	Main purpose	CH2-Temp supply (deg)
	Attire	CH2-Temp return (deg)
Case study: Yield prediction	Time spent (min)	CH2-PHE supply (deg)
B temp (deg)	Ofte exercise	TS 1(Deg)
C temp (deg)	Diet rate	TR1 (Deg)
P1	BM	KW(r) / KW (e)
P2	Temperature Before (deg)	Chiller 1 (CH1) (KW)
Preheat time	Temperature After (deg)	Chiller 2 (CH2) (KW)
Manuf time		

## 5 Results and Discussion

The algorithms of the T-Method, T-Method + OA and T-Method + BPSO were generated using the Matlab R2017a application software [38]. To determine the accuracy measure in this study, we rely on the formula of mean relative error (MRE) as described in Equation (24) where  $s$  denotes the number of samples,  $y_i$  is the actual output and  $\hat{y}_i$  is the predicted output.

$$MRE = 100 \cdot \frac{1}{s} \sum_{i=1}^s \frac{|y_i - \hat{y}_i|}{y_i}. \tag{24}$$

One of the accuracy measures used in this study was using  $R^2$  value. It was applied in 2 different case studies, as stated in Table 3. The  $R^2$  represents how good the correlation between actual and predicted historical data is. The higher the value of  $R^2$ , the higher the chances of getting higher accuracy in predicting the unknown output data in the future. In Table 3, both case studies show that the values of  $R^2$  for both models and predicted unknown data do not vary much for the T-Method, namely 0.97 versus 0.98 for prediction of calorie burn and 0.77 versus 0.7551 for power consumption prediction.

Table 3: Summary of Case Study Prediction Analysis for Prediction of Calorie Burn and Power Consumption

Dataset	No. of observations/no. of unknown data	Method	Model correlation Coefficient ( $R^2$ )	Predicted Value correlation Coefficient ( $R^2$ )	Variable count	Reduced no. of variables
Prediction of calorie burn (kcal) for ice skating activity	17 (4)	T-Method	0.97	0.98	15	none
		T-Method + OA		0.8		9
		T-Method + BPSO		0.74		5
Power consumption prediction for thermal energy storage system	14 (13)	T-Method	0.77	0.7551	15	none
		T-Method + OA		0.7995		8
		T-Method + BPSO		0.803		10

As for the prediction of calorie burn case study, the results show better prediction accuracy compared to the other two optimum methods when using all 15 variables. Referring to Table 2 above, the list of variables used within this study involved some attributes and discrete types of data, which might not be possible to be excluded from the analysis since it will result in bias and may deteriorate the prediction accuracy. It might be the reason why the other two optimization methods have lower performance compared to the original T-Method alone. Table 4 also represents the normality test study of the historical output ( $M_i$ ) data for 10 different test types, which were run using MATLAB mainly for sample size  $<1000$  [39]. The calorie burn and power consumption prediction case studies are normally distributed across all tests.

The T-Method seems to be the best approach for the calorie burn prediction type of study, which involves many attribute types of variables. The prediction of power consumption case study shows that the  $R^2$  value for predicting the unknown data is highest for the T-Method+BPSO method, with only 10 useful variables out of 15 variables.

Table 4: Normality Test for Four Different Case Studies

Case Study	Yield Prediction		Energy Efficiency		Power Consumption		Calorie burn prediction	
	Test Name	p-value	Normality	p-value	Normality	p-value	Normality	p-value
<b>KS Limiting Form</b>	0.9276	1.0000	0.0000	0.0000	0.7352	1.0000	0.9813	1.0000
<b>KS Stephens Modification</b>	0.1500	1.0000	0.0100	0.0000	0.1500	1.0000	0.1500	1.0000
<b>KS Marsaglia Method</b>	0.8742	1.0000	0.0000	0.0000	0.6700	1.0000	0.9638	1.0000
<b>KS Lilliefors Modification</b>	0.2000	1.0000	0.0000	0.0000	0.2000	1.0000	0.2000	1.0000
<b>Anderson-Darling Test</b>	0.2808	1.0000	0.0000	0.0000	0.2050	1.0000	0.7487	1.0000
<b>Cramer- Von Mises Test</b>	0.3148	1.0000	####	1.0000	0.4316	1.0000	0.8087	1.0000
<b>Shapiro-Wilk Test</b>	0.3852	1.0000	0.0000	0.0000	0.0715	1.0000	0.5681	1.0000
<b>Shapiro-Francia Test</b>	0.2151	1.0000	0.0000	0.0000	0.0480	0.0000	0.6688	1.0000
<b>Jarque-Bera Test</b>	0.5826	1.0000	0.0000	0.0000	0.1103	1.0000	0.6335	1.0000

(1: Normal,0: Not Normal)

The next two case studies discuss the effects of optimum variables in terms of the mean relative error prediction (MRE) and  $R^2$  value. Table 5 shows the summary of MRE between the two case studies, comparing the three methods. The  $R^2$  value between actual and predicted historical data of yield prediction and energy performance cases is 0.75, which shows a considerably high correlation between the predicted and actual values. Yield data involved only a minimum sample size, and the T-Method is able to predict unknown data with considerably low error percentages. T-Method+BPSO also helps to further optimize the number of variables from six to only four optimum variables with the same error percentages.

However, in the estimation of energy performance case study, the results show consistently high MRE across the three methods. Compared with the analysis done by [37] which used random forest (RF) and iterative reweighted least square (IRLS) methods, the results show significant differences compared to the three methods discussed in this study. The variable optimization in T-method is mainly relying on the accuracy of the model baseline, as in Equation (1). The normality test results in Table 5 show that the estimation of energy performance case is not normally distributed. It generally reflects that abnormal data analyzed using the T-Method might carry a risk in assessing model accuracy. The Random Forest (RF) is a powerful nonlinear approach that often produces excellent results [37]. The existing T-Method is already showing low error accuracy, and it is not improved when optimization takes place. Further investigation of the T-Method model baseline needs to be done to increase this accuracy level. This is one of the disadvantages of the T-Method that been foreseen now and been further explored by several researchers [3], [4], [28].

Table 5: Summary of Case Study Prediction Analysis for Yield Prediction and Estimation of Energy Performance

Dataset	No. of observations/no. of unknown data	Methodology	Model Correlation Coefficient ( $R^2$ )	Prediction Accuracy (%)	Variable count	Reduced no. of variables
Yield prediction for manufacturing-production	7 (1)	T-Method	0.7551	2.5	6	none
		T-Method + OA		2.93		2
		T-Method + BPSO		2.5		4
Estimation of energy performance of residential buildings	768(768)	T-Method	0.75	15	8	none
		T-Method + OA		19		2
		T-Method + BPSO		18		3
Source: UCI Machine Learning , (Tsanas,2012- for Random forest (RF) and iteratively reweighted least squares (IRLS))		IRLS		10.09		3
		RF		2.18		3

It appears that the T-Method is more reliable for data with a lower sample size, as shown by the case studies shared, except for the estimation of building energy performance case study. Similar findings have also been reported by [4], [23–24], [27–28]. Cases that deal with high sample data might be at risk since linear regression analysis tends to be biased towards nonlinear and outliers that lead to non-normality issues.

Dealing with a higher sample size is also one of the disadvantages of the T-Method. The main intention of this study (as highlighted in the background study) is to optimize the variable selection to improve or at least maintain the prediction accuracy with much lesser cost. In order to fulfill this, the predefined data (historical data) should be well analyzed with a normality study and  $R^2$  value for model creation which is as high as possible. If both of these requirements can be met, the optimization process in reducing the number of variables would be much more reliable and accurate. A comparison between T-Method+OA and T-Method+BPSO analysis results shows that T-Method+BPSO performs better with higher  $R^2$  and MRE values in terms of variables optimization purposes.

## 6 Conclusion

In both the power consumption prediction and yield prediction case studies, T-Method + BPSO performed well compared to the other two prediction cases. Normal data with higher  $R^2$  value will contribute to higher future prediction accuracy. Dealing with attributes and discrete types of data is among the crucial things to be adequately considered within this method. Reducing the attribute and discrete type variables that are normally distributed and highly correlated might deteriorate the overall prediction performance. The T-Method in this study proved to be useful for predicting cases that are dealing with lower sample sizes. Enhancing the T-Method with BPSO element as variables screening optimization was able to increase or even maintain the prediction accuracy for cases that are normally distributed, with high  $R^2$  value, using low sample size.

## Acknowledgment

This study was financially supported by the Universiti Teknologi Malaysia under the Fundamental Research Grant Scheme FRGS/1/2019/TK08/UTM/02/4 and Research Management Centre (RMC), Universiti Teknologi Malaysia.

## References

- [1] Hussain, W., Hussain, F. K., Saberi, M., Hussain, O. K. and Chang, E. Comparing time series with machine learning-based prediction approaches for violation management in cloud SLAs, *Futur. Gener. Comput. Syst.* 2018. 89: 464–477.
- [2] Teshima, S., Hasegawa, Y. and Tatebayashi, K. *Quality Recognition and Prediction: Smarter Pattern Technology with the Mahalanobis-Taguchi System*. New York: Momentum Press, 2012.
- [3] Nishino, K. and Suzuki, A. Taguchi's T-method using median-median line for small sample with outliers. *Electron. Commun. Japan.* 2018. 49–56,
- [4] Negishi, S., Morimoto, Y., Takayama, S. and Ishigame, A. Daily Peak Load Forecasting by Taguchi's T Method. *Electr. Eng. Japan (English Transl. Denki Gakkai Ronbunshi)*. 2017. 201(1): 57–65,
- [5] Abraham, B. and Variyath, A. M. Discussion. *Technometrics*. 2003. 45(1). 22–24.
- [6] Foster, C. R., Jugulum, R. and Frey, D. D. Evaluating an adaptive One-Factor-At-a-Time search procedure within the Mahalanobis-Taguchi System. *Internatinal J. Ind. Syst. Eng.* 2009. 4(6): 600–614.
- [7] Iquebal, A. S. and Pal, A. Artificial bee colony optimisation-based enhanced Mahalanobis Taguchi system for classification. *Int. J. Intell. Eng. Informatics*. 2014. 2(23): 181–194,
- [8] Hawkins, D. M. Discussion. *Technometrics*. 2003. 45(1): 25–29.
- [9] Tsui, K. , Sukchotrat, T. and Chen, V. C. P. A comparison study and discussion of the Mahalanobis-Taguchi System Seoung Bum Kim. *Int. J. Ind. Syst. Eng.* 2009. 4(6): 631–644.
- [10] Harudin, N., Jamaludin, K. R., Muhtazaruddin, M. N., Ramlie, F. and Muhamad, W. Z. A. W. Artificial bee colony for features selection optimization in increasing T-Method accuracy. *Int. J. Eng. Technol.* 2018. 7(4): 885–891.
- [11] Reséndiz-Flores, E. O., Navarro-Acosta, J. A., Mota-Gutiérrez, C. G. and Reyes-Carlos, Y. I. Fault detection and optimal feature selection in automobile motor-head machining process. *Int. J. Adv. Manuf. Technol.* 2018. 94(5): 2613–2622.
- [12] Jain, I., Jain, V. K. and Jaina, R. Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification. *Appl. Soft Comput. J.* 2018. 62: 203–215,
- [13] Rajamohana, S. P. and Umamaheswari, K. Hybrid approach of improved binary particle swarm optimization and shuffled frog leaping for feature selection. *Comput. Electr. Eng.* 2018. 67: 497–508.

- [14] Too, J., Abdullah, A. R. and Saad, N. M. A new co-evolution binary particle swarm optimization with multiple inertia weight strategy for feature selection. *Informatics*. 2019. 6(2): 1–14
- [15] Yuping, G., Longsheng, C. and Zhipeng, C. Classification of imbalanced data based on MTS-CBPSO method: a case study of financial distress prediction. *JIPS(Journal Inf. Process. Syst.* 2019. 15(3): 682–693.
- [16] Mota-Gutiérrez, C. G., Reséndiz-Flores, E. O. and Reyes-Carlos, Y. I. Mahalanobis-Taguchi system: state of the art. *Int. J. Qual. Reliab. Manag.* 2018. 35(3): 596–613.
- [17] Muhamad, W. Z. A. W., Jamaludin, K. R., Yahya, Z. R. and Ramlie, F. A hybrid methodology for the mahalanobis-taguchi system using random binary search-based feature selection. *Far East J. Math. Sci.* 2017. 101(2): 2663–2675.
- [18] Ramlie, F., Jamaludin, K. R., Dolah, R. and Muhamad, W. Z. A. W. Optimal feature selection of Taguchi character recognition in the Mahalanobis-Taguchi system using bees algorithm. *Glob. J. Pure Appl. Math.* , 2016. 12(3): 2651–2671.
- [19] Mojtaba Ahmadiéh, K., Mohammad, T. and Mahdi Aliyari, S. A novel binary particle swarm optimization. *Mediterranean conference on control and automation*. 2007 18(2): 117–122.
- [20] Pal, A. and Maiti, J. Development of a hybrid methodology for dimensionality reduction in Mahalanobis-Taguchi system using Mahalanobis distance and binary particle swarm optimization. *Expert Syst. Appl.* 2010. 37(2): 1286–1293.
- [21] Reséndiz, E. and Rull-Flores, C. A. Mahalanobis-Taguchi system applied to variable selection in automotive pedals components using Gompertz binary particle swarm optimization. *Expert Syst. Appl.* 2013. 40(7): 2361–2365.
- [22] Dasneogi, P., Cudney, E. A. and Ph, D. Comparing the Predictive Ability of T-Method and Linear Regression Method. In *Proceedings of the 2009 Industrial Engineering Research Conference Comparing*. 2009. 2176–2182.
- [23] Cudney, E. A. and Shah, P. A. Predicting annual precipitation using the T-Method. In *Proceedings of the 2010 Industrial Engineering Research Conference*. 2010. 1–6.
- [24] Inoh, J., Nagata, Y., Horita, K. and Arisa, M. Prediction accuracies of improved Taguchi’s T. methods compared to those of multiple regression analysis. *J. japanese Soc. Qual. Control.* 2012. 42(2): 265–277.
- [25] Jalil, I. E. A., Shamsuddin, S. M., Muda, A. K., Azmi, M. S. and Hashim, U. R. Predictive based hybrid ranker to yield significant features in writer identification. *Int. J. Adv. Soft Comput. its Appl.* 2018. 10(1): 1–23.
- [26] Kawada, H. and Nagata, Y. An application of a generalized inverse regression estimator to Taguchi’s T-Method. *J. Japanese Soc. Qual. Control.* 2015. 1985: 3–4.
- [27] Harudin, N., Jamaludin, K. R., Nabil Muhtazaruddin, M., Ramlie, F. and Muhamad, W. Z. A. W. A feasibility study in adapting Shamos Bickel and Hodges Lehman estimator into T-Method for normalization. In *IOP Conference Series: Materials Science and Engineering*. 2018. 319(1): 6
- [28] Harudin, N. et al. Increasing T-Method accuracy through Application of robust M-estimator,” *J. Adv. Reseach Dyn. Control Syst.* 2018. 09-special issue: 44–48.

- [29] Nakao, Y. and Nagata, Y. Analysis of data including missing values in the Taguchi's T method. *Total Qual. Sci.* 2018. 4(2): 53–64.
- [30] Kennedy, J. and Eberhart, R. C. A discrete binary version of the particle swarm algorithm. *1997 IEEE Int. Conf. Syst. Man, Cybern. Comput. Cybern. Simul.* 1997. 5: 4104–4108.
- [31] Eberhart, R. and Kennedy, J. A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science (Nagoya, Japan), IEEE Service Center, Piscataway, NJ.* 1995. 39–43.
- [32] Vieira, S. M., Mendonça, L. F., Farinha, G. J. and Sousa, J. M. C. Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients.' *Appl. Soft Comput.* 2013. 13(8): 3494–3504.
- [33] Kennedy J. et al. Particle swarm optimization. 1995. In *Proceedings. IEEE Int. Conf. on Neural networks.* 1995. 4: 1942–1948.
- [34] Xu, J. and Huiyou, C. The discrete binary version of the improved particle swarm optimization algorithm. In *Proceedings - International Conference on Management and Service Science, MASS 2009.* 2009.
- [35] Zhang, Y., Wang, S., Phillips, P. and Ji, G. Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. *Knowledge-Based Syst.* 2014. 64: 22–31.
- [36] Menhas, M. I., Wang, L., Fei, M. and Pan, H. Comparative performance analysis of various binary coded PSO algorithms in multivariable PID controller design. *Expert Syst. Appl.* 2012. 39(4): 4390–4401.
- [37] Tsanas, A. and Xifara, A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy Build.* 2012. 49: 560–567.
- [38] Alam, M. Codes in MATLAB for Particle Swarm Optimization. 2016. [Source code].<https://www.researchgate.net/publication/296636431>.
- [39] Öner, M. and Deveci Kocakoç, I. JMASM 49: a compilation of some popular goodness of fit tests for normal distribution: their algorithms and MATLAB codes (MATLAB),” *J. Mod. Appl. Stat. Methods.* 2017. 16(2): 547–575.