Semarak Engineering Journal

Journal homepage:
https://semarakilmu.com.my/journals/index.php/sem_eng/index
ISSN: 3036-0145

# Development Of Enhanced Taguchi's T-Method Model in Predicting Energy Demand

Muhammad Nabihan Arif Nasir[1], Nolia Harudin[1,*], Farah Elida Selamat[1], Zulkifli Marlah@Marlan[2]

[1] College of Engineering, Universiti Tenaga Nasional, Jalan Ikram-Uniten, 43000 Kajang, Selangor, Malaysia
[2] Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia

**ABSTRACT**

Energy is important for a country to grow and become a developed nation. Energy management will help a country to supply energy demand without any shortage or excess of energy. To manage energy, the government must know important aspects that lead to energy consumptions, such as growth domestic product (GDP) of the nation, energy supply, energy transformation, and energy consumptions. Here, prediction analysis will come in handy to predict the final energy consumptions. The Taguchi's T-Method is one of the prediction analyses that may be used for this purpose as it is applicable in various fields. It can predict with a limited data sample, thus making it reliable and cost saving. It depends on history data to develop a prediction model by adapting signal-to-noise ratio (SNR) and zero proportional concepts. Orthogonal Array is introduced in the Taguchi's T-Method to optimise the model by eliminating variables that may reduce the accuracy of the prediction model. The Taguchi's T-Method uses dynamic SNR that is not suitable for all types of predictions, and unit space selection in the middle position between the lowest and highest data will reduce the number of raw data. In this study, Ta-Method and larger-the-better signal-to-noise ratio (LTB SNR) were introduced to compare prediction accuracy of the current method with that of the default Taguchi's T-Method. Three models were developed, namely default Taguchi's T-Method, T-Method + LTB SNR, and Ta-Method + LTB SNR for three different case studies. The result showed the Ta-Method + LTB SNR had the most accurate prediction compared to the other models for mean absolute percentage error (MAPE) for predicting energy demand case study.

***Keywords:***
Energy demand predictions; larger-the-better signal-to-noise ratio; prediction; Taguchi's T-Method

## 1. Introduction

Energy demand prediction is vital to meet a nation's energy demand so its economy can grow, and it can reduce excess energy supply to save cost. The government may save some budgets in terms of supplying energy for other sectors. Moreover, energy management could help the government to plan on reducing greenhouse gases emissions (GHG) and prepare for energy supply transition from non-renewable energy to renewable energy. Most of the developing countries, such as Malaysia is depending on natural gas and crude oil for energy supply. An increase in GDP will cause an increase in energy consumption [1]. The dependency on crude oil may be reduced by formulating energy demand policies and preparing for alternative energy supply by the government. To be a developed nation, energy supply must be sufficient and sustainable development for energy sources

* *Corresponding author.*
*E-mail address: Nolia@uniten.edu.my*

must be available [1]. In Malaysia, forecasting of energy consumption mainly involves electricity consumption, not crude oil production. Crude oil cannot be used directly as it must be processed first. Primary energy supply from crude oil is the second highest after natural gas in 2017 [2].

The Mahalanobis-Taguchi System (MTS) is a processing multivariate data for quantitative decision-making. There are a few methods under MTS, where each of them is designed for different purposes and one of them is the Taguchi's T-Method for prediction analysis. One of the Taguchi's T-Method abilities is it can perform prediction with a small sample data [3]. For multiple linear regression, it needs to have more sample data than the number of variables in order to get an accurate prediction result. To develop a prediction model, history data is needed to identify the pattern and variables to compute the prediction result. The Taguchi's T-Method also implements SNR and zero proportional coefficient to develop a prediction model. SNR is applied to measure the quality for each of the variables in multivariate system, while zero proportional coefficient is achieved through normalise process. Once the SNR and proportional coefficient have been computed for each of the variables, the model can be developed. Then, the model is optimised using Orthogonal Array (OA) to maintain only the relevant variables because some variables may deteriorate the performance of the prediction model.

The Mahalanobis-Taguchi System (MTS) is a processing multivariate data for quantitative decision-making. There are a few methods under MTS, where each of them is designed for different purposes and one of them is the Taguchi's T-Method for prediction analysis. One of the Taguchi's T-Method abilities is it can perform prediction with a small sample data [3]. For multiple linear regression, it needs to have more sample data than the number of variables in order to get an accurate prediction result. To develop a prediction model, history data is needed to identify the pattern and variables to compute the prediction result. The Taguchi's T-Method also implements SNR and zero proportional coefficient to develop a prediction model. SNR is applied to measure the quality for each of the variables in multivariate system, while zero proportional coefficient is achieved through normalise process. Once the SNR and proportional coefficient have been computed for each of the variables, the model can be developed. Then, the model is optimised using Orthogonal Array (OA) to maintain only the relevant variables because some variables may deteriorate the performance of the prediction model.

An alternative method for unit space selection has been proposed by Inoh *et al.,* [4] using the Ta-Method and the Tb-Method. The results showed an improvement from the default Taguchi's T-Method [4]. Previous research, such as Matsui *et al.,* [6], Harudin *et al.,* [11], Nakao *et al.,* [5] and Kawada *et al.,* [7] used dynamic SNR to develop the model. The accuracy is acceptable, but some improvements may be applied to increase or maintain the accuracy level. This research proposes to replace the T-Method with the Ta-Method and dynamic SNR with LTB SNR to improve the level of accuracy and enhance the Taguchi's T-Method. The objectives of this research are to predict energy consumption for crude oil and petroleum supply using the Taguchi's T-Method, formulate an enhanced mathematical model of the Taguchi's T-Method for better crude oil and petroleum supply prediction accuracy, and compare the result of enhanced Taguchi's T-Method with that of multiple linear regression.

## 2. Related Studies

A lot of research has been carried out to improve the default Taguchi's T-Method on various procedures from unit space selection to OA optimization. Inoh *et al.,* improved the unit space selection using the Ta-Method and the Tb-Method [4]. In the Ta-Method, all sample data were included as signal data and the average value for each of the variables and output value were calculated. The Normalise process was calculated by subtracting each of the variables and output value with the average value for each variable and average output value, respectively. While for the Tb-Method, it still uses all sample data as the signal data, but the normalised process is different. SNR for each variable needs to be calculated. The normalised process was done by subtracting each of the variables from the sample that had a higher SNR. The normalised process was done differently and independently for each of the variables [4-6]. Kawada *et al.,* used the generalized inverse regression (GIR) by improving the estimated output values for each of the items [7]. GIR was used as linear calibration for newly obtained data. Another research from Harudin *et al.,* [9] used the M-Estimator to increase the T-Method prediction accuracy. The M-Estimator was used to calculate β by replacing dynamic SNR from the default Taguchi's T-Method. Nishino and Suzuki introduced the median-median line (MML) for small training data with outliers [8]. The researchers did not use the least square method for all the variables, and they replaced them with the MML models to reduce the influence of outliers. Normalization was also not done and it was replaced with the MML models because SNR will be zero if the error variance is larger than the SNR when the training data is small [8]. In terms of optimization process, Harudin *et al.,* introduced the Artificial Bee Colony to increase the accuracy of the Taguchi's T-Method [9].

## 3. Methodology

This section explains the theory of the default Taguchi's T-Method, T-Method + LTB SNR, and Ta-Method + LTB SNR. All models were normalise using OA. The results before and after optimization were analysed to check and compare their prediction accuracy. The results were also be compared with multiple linear regression.

### *3.1 Default Taguchi's T-Method*

The default Taguchi's T-Method was the first model. It is important to define the unit space and signal data for the normalisation process. Unit space was chosen from the output value, where it must be near the centre of the population data [3,6]. Unit space selection should be homogenous and in the centre position between the high and low population data [9-11]. The sample data that formed the unit space was excluded from training data and the rest of the data were the signal data. Then, the average value for unit space was calculated. The normalization process was done by subtracting the signal data from the average value of the unit space, as shown in Eq. (1).

$$normalised\ data\ \left(X_{ij}, M_0\right) = \text{ signal data} - \text{average of unit space} \tag{1}$$

The following steps are the important aspects in the default Taguchi's T-Method, which are computation of proportional coefficient β and SN ratios η for item-by-item basis. Eq. (2) to Eq. (9) show the theory for proportional coefficient β and SN ratios η calculations. Proportional coefficient β is based on the least square method, where it helps to achieve the zero-point proportional concept through the normalization process [3]. Eq. (3) is known as the dynamic SNR, which was used to measure the quality for each of the variables. Based on Eq. (3) and Eq. (4), a higher SNR value of an item will give a higher contribution to the overall model. If the SNR value is negative, it will be assumed zero.

$$\text{Proportional coefficient } \beta_1 = \frac{M_1 X_{11} + M_2 X_{21} + \cdots + M_l M_{l1}}{r} \tag{2}$$

SN ratio $\eta_1 = \frac{\frac{1}{r}(S_\beta - V_{e1})}{V_{e1}}$ (when $S_\beta > V_{e1}$) (3)

$\eta_1 = 0$ (when $S_\beta < V_{e1}$) (4)

where

Effective divider $r = M_1^2 + M_2^2 + \cdots + M_l^2$ (5)

Total variation $S_{T1} = X_{11}^2 + X_{21}^2 + \cdots + X_{l1}^2$ (f = *l*) (6)

Variation of Proportional term $S_{\beta 1} = \frac{(M_1 X_{11} + M_2 X_{21} + \cdots + M_l X_{l1})^2}{r}$ (f = 1) (7)

Error variation $S_{e1} = S_{T1} - S_{1\beta}$ (8)

Error variance $V_{e1} = \frac{S_{e1}}{l-1}$ (9)

Once all proportional coefficient β and SN ratios η were computed, the prediction model was developed, as shown in Eq. (10). Each of the integrated estimate output value ($\hat{M}_i$) was calculated by replacing the variables $X_{ij}$ for each column.

$$\hat{M}_i = \frac{\eta_1 \times \frac{X_{i1}}{\beta_1} + \eta_2 \times \frac{X_{i2}}{\beta_2} + \cdots + \eta_k \times \frac{X_{ik}}{\beta_k}}{\eta_1 + \eta_2 + \cdots + \eta_k} \quad (i = 1, 2, \ldots, l) \qquad (10)$$

For further optimisation of the model, Orthogonal Array can be applied to select variables for better prediction and saving cost. Below are the formulas to calculate SNR (db) for optimisation:

Integrated Estimated SN Ratio $\eta = 10 \log(\frac{\frac{1}{r}(S_\beta - V_{e)}}{V_{e)}})$ (db) (11)

where

Linear equation $L = M_1\hat{M}_1 + M_2\hat{M}_2 + \cdots + M_l\hat{M}_l$ (12)

Effective divider $r = M_1^2 + M_2^2 + \cdots + M_l^2$ (13)

Total variation $S_T = \hat{M}_1^2 + \hat{M}_2^2 + \cdots + \hat{M}_l^2$ (14)

Variation of proportional term $S_\beta = \frac{L^2}{r}$ (f = 1) (15)

Error variation $S_e = S_T - S_\beta$ (f = *l* – 1) (16)

Error variance $V_e = \frac{S_e}{l-1}$ (17)

This research used two-level OA type, which represents either to use the variables or not to use them. There are many types of arrays depending on the total number of variables for the study. Linear equation, L is the sum of multiplication between estimated output value $\hat{M}_i$ for each sample and its actual output value $M_i$. Table 1 shows the OA $L_{12}$ type. One means the variables are used, while 2 means the variables are not being used and treated as zero in Eq. (10). The number of computations depends on the number of samples in the normalised data in Eq. (1).

For this research, 34 data were training data and 5 data were training data. For the default Taguchi's T-Method, 3 samples were excluded for unit space. Each of the simulation had 31 samples. In total, there were 31 estimation output values $\hat{M}_l$ for one simulation. The model simulated for 12 times with different sets of variables being used to calculate the integrated estimate output value ($\hat{M}_i$). Once all values of $\hat{M}_l$ were computed, Eq. (12) to Eq. (17) were used to compute the dynamic SNR for level 1 and level 2. This will analyse the average value of SNR for each of the variables. If it has a higher value of level 1 SNR compared to level 2, it will highly contribute to the prediction model. If the SNR value of the variable has higher level 2 compared to level 1, it will be excluded in the prediction model. Eq. (10) can be used by considering all variables or only the variables that will give optimum prediction result. The integrated estimate output value ($\hat{M}_i$) is in the normalised form, for it to be transformed back to raw data unit, $\hat{y}$ , it must be added by average output value unit space $M_0$ as shown in Eq. (18). The results before and after optimisation were recorded for the first model.

$$\hat{y} = \widehat{M}_i + M_0 \tag{18}$$

**Table 1**
Orthogonal array $L_{12}$

| Exp | Variables | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 |

*3.2 T-Method + LTB SNR*

The second model was the T-Method + LTB SNR. The model used the same calculations as the first model, but the SNR calculation used Eq. (19). Based on Eq. (19), $x_{ij}$ is the variable, while n is the number of sample data. It used larger-the-better SNR to maximise the signal for all variables. Eq. (3), and Eq. (6) to Eq. (9) for dynamic SNR, which were originally used for the default Taguchi's T-Method were replaced. Unit space selection was the same as the first model.

Larger-the-better SNR, $\eta = -10log\frac{1}{n}(\Sigma\frac{1}{{x_{ij}}^2})$ (i = 1,2,3, …, n) (j = 1,2,3, …, n) (19)

Once the proportional coefficient, β and LTB SNR, η were computed for all variables, the model was developed as in Eq. (10). The model was trained using 34 training data and 5 testing data. The results were recorded before optimisation. Then, the optimisation process was done using the OA type $L_{12}$, similar to the first model, to evaluate the importance of all variables. The model was run again by only selecting the optimum variables and the results were recorded after the optimisation process.

### *3.3 Ta-Method + LTB SNR*

The third model developed was the Ta-Method + LTB SNR. The unit space selection was different from that of the first and second models. This model applied the Ta-Method, which was introduced by Inoh *et al.,* [4]. The Ta-Method used all data as the signal data and no data will be excluded for unit space since it will not be defined in the Ta-Method. The average value was calculated for each of the variable and output value [3,7,11]. Eq. (20) shows the calculation to find the normalised value for each variable $\bar{x}_j$, while Eq. (21) shows the calculations to normalise the value for output value $M_0$.

$$\bar{x}_j = x_j - \frac{1}{n}\,(x_{1j} + x_{2j} + \cdots + \, x_{nj}) \tag{20}$$

$$M_0 = y_j - \frac{1}{n}(\,y_1 + \, y_2 + \cdots + y_n) \tag{21}$$

The proportional coefficient, β is still the same as in the default Taguchi's T-Method in Eq. (2). The LTB SNR was used in the third model by replacing the dynamic SNR. The equation for η is the same as in Eq. (19) for the second model. The model was developed as in Eq. (10) after the proportional coefficient, β and SN ratios η were computed. The model was then optimised using the OA type $L_{12}$, similar to the default Taguchi's T-Method. The results before and after optimisation were recorded.

## **4. Results Analysis**

The results before and after optimization for all the three models were recorded. The results were analysed based on the error between the actual and predicted values. Eq. (22) shows the error calculation for each sample. Then, the errors were analysed using the mean absolute percentage error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE), as shown in Eq. (23), Eq. (24), and Eq. (25), respectively. The equations denote n as the number of sample data, $y_i$ as the predicted value, and $x_i$ as the actual value.

$$Error = Actual\ value - Predicted\ value \tag{22}$$

$$Mean\ Absolute\ Error, MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{23}$$

$$Root\ Mean\ Square\ Error, RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \tag{24}$$

$$Mean\ Absolute\ Percentage\ Error, MAPE = \frac{\sum_{i=1}^{n}\left|\frac{y_i - x_i}{x_i}\right|}{n} x100 \tag{25}$$

Sample data was used in multiple linear regression to compare the results with all three models. This will show whether or not the three models can perform better than the multiple linear regression.

*4.1 Case Study Data Collection*

The three models were applied to energy consumption for crude oil and petroleum in Malaysia. The data were obtained from the Malaysia Energy Information Hub (MEIH) under the Energy Commission of Malaysia. The energy demand was based on GDP, primary energy supply, and energy transformation. The data were available from year 1980 to 2018 in a total of 39 sample data. The obtained data were rechecked with the National Energy Balance Table (NEB) for year 2010 to 2017, while the rest were not rechecked due to absence of their NEB [2]. The data were divided into two groups, namely 34 sample data for training data and latest 5 sample data for testing data.

The other two case studies were yield prediction for manufacturing production and prediction for tensile strength of a product from the mixing ratio of raw materials, which were taken from the Quality Recognition and Prediction e-book [12]. For these two case studies, all three models were not optimised where all variables will be used it in Eq. (10). All variables for the three case studies are shown in Table 2.

**Table 2**
List of variables

| Case study: Energy consumptions | Case study: Yield prediction | Case study: Tensile strength |
|---|---|---|
| GDP at 2015 Prices (RM Million) | B temp (degree) | Raw material 1 |
| GDP at Current Prices (RM Million) | C temp (degree) | Raw material 2 |
| Crude Oil | P1 | Raw material 3 |
| Petroleum Products | P2 | Raw material 4 |
| Losses and Own Use | Pre-ht time | Raw material 5 |
| Natural Gas conversion | Manuf time | Additive 1 |
| Refinery Input | | Additive 2 |
| Total Petroleum Product (refinery output) | | |

The first model, which was the default Taguchi's T-Method was developed using MT System All-Purpose Software MTRT-AddIns. The second model, namely the T-Method + LTB SNR and the third model, namely the Ta-Method + LTB SNR were developed using Microsoft Excel. All results were analysed using error calculations to find the difference between the actual and predicted values. Then, the errors were used in MAE, RMSE, and MAPE to measure the accuracy of the prediction model. Lower errors indicate a better and more accurate prediction model. Multiple linear regression was used to compare the results for all models and case studies. All results were compared between four types of models as shown below:

i. Actual value
ii. The first model: Default Taguchi's T-Method
iii. The second model: T-Method + LTB SNR
iv. The third model: Ta-Method + LTB SNR
v. Multiple Linear Regression

Figure 1 shows the result for energy consumptions for actual value, default Taguchi's T-Method, T-Method + LTB SNR, Ta-Method + LTB SNR, and multiple linear regression. The results were not optimised. For the training data, the first model showed a better result compared to the second and third models. For testing data, the second model showed better results compared to the first and second models. However, multiple linear regression was more accurate compared to the three models.

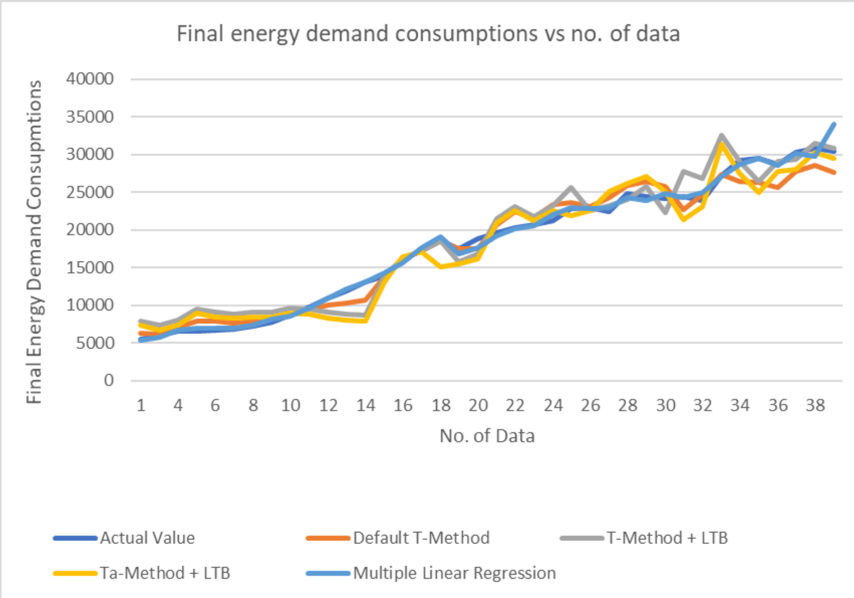


**Fig. 1.** Before optimisation

Figure 2 shows the results after optimisation. Only variables that gave the optimum results were used in the model. The graphs shown are smoother compared to the results before optimization. The third model showed better results for both training and testing data compared to the first and third models. However, multiple linear regression still had the overall best results.

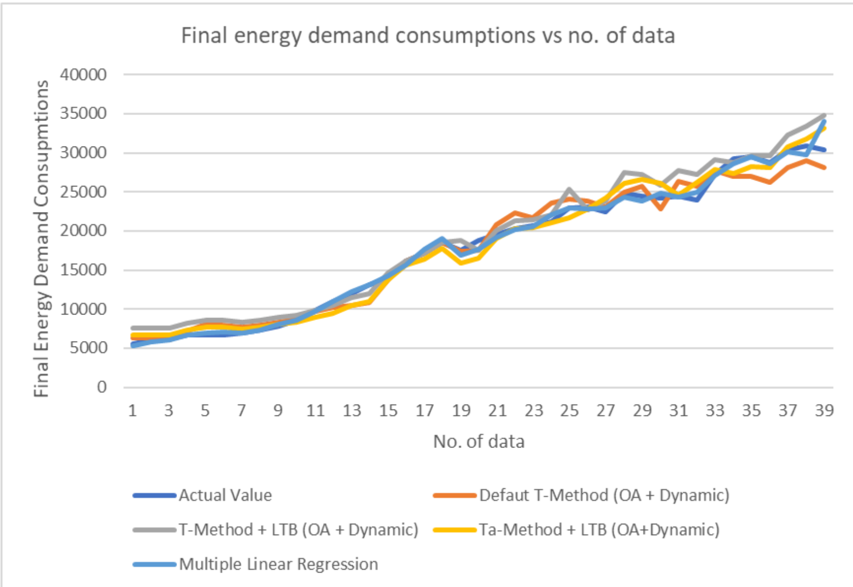


**Fig. 2.** After optimisation

Table 3 shows the analysis for all models and multiple linear regression before optimization. Comparisons were done in terms of MAPE because the values of MAE and RMSE were large. The errors were quite large for MAE and RMSE due to the value of the raw data, which started from 5000 ktoe up to 30000 ktoe. Energy itself is large and the unit has the prefix of kilo. Lu *et al.,* and Zhao and

Lifeng also recorded high value of error between the actual and predicted values when predicting energy demand [13,14]. For training data, the first model was more accurate with 7.419% MAPE, followed by the third and second models with 12.534% and 15.323%, respectively. For testing data, the second model had a more accurate result with 5.626% MAPE, followed by the third and first models with 6.183% and 9.283%, respectively. The first model was considered as overfitting because the testing data error was larger than the training data. The second and third model showed improvements from the default Taguchi's T-Method.

**Table 3**
Results before optimization

| Data type | Training data | | | Testing data | | |
|---|---|---|---|---|---|---|
| Model type | Default Taguchi's T-method | T-Method + LTB SNR | Ta-Method + LTB SNR | Default Taguchi's T-method | T-Method + LTB SNR | Ta-Method + LTB SNR |
| No. of data | 31 | 31 | 34 | 5 | 5 | 5 |
| MAE | 1063.325 | 1867.405 | 1677.153 | 2773.650 | 1658.592 | 1842.012 |
| RMSE | 1295.648 | 2247.691 | 2075.402 | 2789.203 | 2024.700 | 2351.051 |
| MAPE | 7.419% | 15.323% | 12.534% | 9.283% | 5.626% | 6.183% |

Table 4 shows the analysis for all models and multiple linear regression after optimization. All three models were optimised using the OA type $L_{12}$. The errors were improved. For training data, multiple linear regression had the lowest MAPE with 2.381%, followed by the third model with 7.210%. The first model had MAPE of 7.523%, which was not much different from that of the third model. The second model had the largest MAPE with 11.398%. For testing data, multiple linear regression still had the lowest MAPE with 3.375%, followed by the third model with 3.99%. The first and second models had MAPE of 7.854% and 6.430%, respectively. Without considering multiple linear regression, the Ta-Method + LTB SNR outperformed the default Taguchi's T-Method and T-Method + LTB SNR as it had more accurate results for both training and testing data.

After optimisation to select the optimum variables for the third model, three variables were excluded from the model because they had higher level 2. The three variables were petroleum products, losses and own use, and total petroleum product (refinery output). Most petroleum products are exported and may not contribute much for energy consumptions. For losses an own use, some of the energy is lost to the surroundings and being used for the refinery. The values are small and mostly not available, thus making them quite difficult to be analysed. For petroleum product (refinery output), the values are almost constant with refinery input. Therefore, petroleum product variable and refinery output variable can be considered as the same variables. The model for Ta-Method + LTB SNR model is shown below:

$$\widehat{M}_i = \frac{\begin{array}{c}(138.411\times\frac{X_{i1}}{32.721}+139.843\times\frac{X_{i2}}{36.658}+108.722\times\frac{X_{i3}}{1.076}+0\times\frac{X_{i4}}{(-0.129)}+\\ 0\times\frac{X_{i5}}{0.0067}+95.022\times\frac{X_{i6}}{0.198}+108.679\times\frac{X_{i7}}{1.069}+0\times\frac{X_{i8}}{0.964})\end{array}}{138.411+139.843+108.722+0+0+95.022+108.679+0} \tag{26}$$

**Table 4**
Results after optimization

| Data type | Training data | | | | Testing data | | | |
|---|---|---|---|---|---|---|---|---|
| Model type | Default Taguchi's T-method | T-Method+ LTB SNR | Ta-Method+ LTB SNR | Multiple Linear Regression | Default Taguchi's T-method | T-Method +LTB SNR | Ta-Method+ LTB SNR | Multiple Linear Regression |
| No. of data | 31 | 31 | 34 | 34 | 5 | 5 | 5 | 5 |
| MAE | 1058.538 | 1363.230 | 969.400 | 355.031 | 2347.508 | 1950.514 | 1202.491 | 1028.268 |
| RMSE | 1252.785 | 1646.023 | 1182.981 | 455.502 | 2362.412 | 2441.815 | 1477.774 | 1723.772 |
| MAPE | 7.523% | 11.398% | 7.210% | 2.381% | 7.854% | 6.430% | 3.99% | 3.375 |

For the second model, the T-Method + LTB SNR was not included in the case studies. Figure 3 shows the results for actual value, default Taguchi's T-Method, Ta-Method + LTB SNR, and multiple linear regression for the percentage yield case study. Percentage yield had 7 training data and 1 testing data. In Table 5, the default Taguchi's T-Method recorded 1.441% MAPE for training data, while Ta-Method + LTB SNR had MAPE of 2.183%. Multiple linear regression was not consistent. Therefore, the testing data could not be tested. The data were not suitable to develop multiple linear regression. For testing data, the default Taguchi's T-Method had the lowest MAPE with 2.436%, while the Ta-Method + LTB SNR had MAPE with 8.242%. The default Taguchi's T-Method is the best model when the sample data is too small to develop a prediction model, and it suits its purpose. Multiple linear regression needs bigger sample data than input variables to develop its model and produce an accurate result [3].

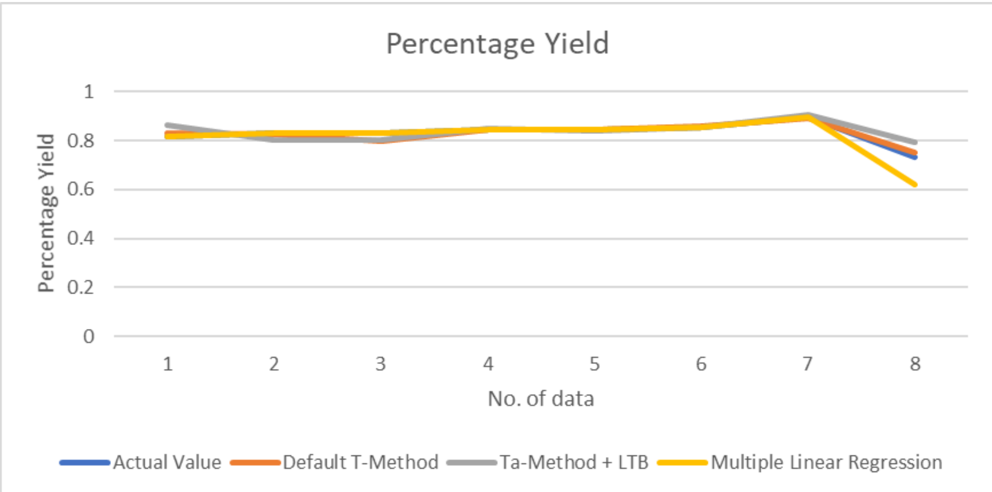


**Fig. 3.** Yield prediction for manufacturing production

**Table 5**
Yield prediction

| Data type | Training data | | | Testing data | | |
|---|---|---|---|---|---|---|
| Model type | Default Taguchi's T-method | Ta-Method+ LTB SNR | Multiple Linear Regression | Default Taguchi's T-method | Ta-Method+ LTB SNR | Multiple Linear Regression |
| No. of data | 5 | 7 | 7 | 1 | 1 | 1 |
| MAE | 0.012 | 0.018 | Not Consistent | 0.018 | 0.060 | - |
| RMSE | 0.016 | 0.024 | Not Consistent | 0.018 | 0.060 | - |
| MAPE | 1.441% | 2.183% | Not Consistent | 2.436% | 8.242% | - |

The next case study was on tensile strength of raw materials. It had 10 training data and 2 testing data. Based on Figure 4, the Ta-Method + LTB SNR has the largest error. The graph is inconsistent and fluctuated. For the default Taguchi's T-Method and multiple linear regression, the graphs are more consistent and have less error. As shown in Table 6, the Ta-Method + LTB SNR has MAPE of 13.186% for training data and 15.625% for testing data. The error is larger than that of the default Taguchi's T-Method with 3.832% MAPE for training data and 4.348% for testing data. Multiple linear regression has the lowest MAPE of 0.726% for training data and 1.9% for testing data, thus making it the most accurate model. For this case study, the Ta-Method + LTB SNR could not perform well as in energy consumptions.

The LTB SNR is suitable for data that have infinity output and not for output value that has a certain range. It is used to maximize the SNR for the output until infinity. The issues for LTB SNR are its ability to reduce variability and adjust the average value to the target. As explained for the percentage yield and raw materials for tensile strength case studies, the default Taguchi T-Method is more accurate than the Ta-Method + LTB SNR as it uses dynamic SNR. The dynamic SNR is used to get the prediction value near to a certain target range.

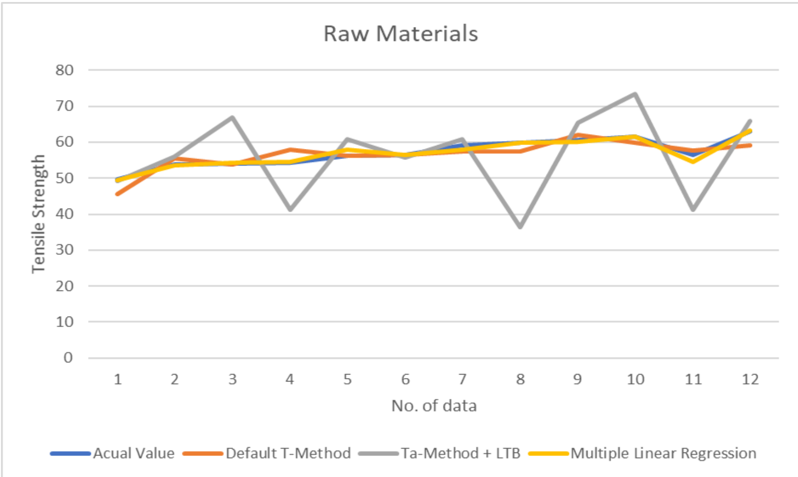


**Fig. 4.** Prediction for tensile strength of a product from the mixing ratio of raw materials

**Table 6**
Raw material for tensile strength prediction

| Data type | Training data | | | Testing data | | |
|---|---|---|---|---|---|---|
| Model type | Default Taguchi's T-method | Ta-Method+ LTB SNR | Multiple Linear Regression | Default Taguchi's T-method | Ta-Method+ LTB SNR | Multiple Linear Regression |
| No. of data | 8 | 10 | 10 | 2 | 2 | 2 |
| MAE | 2.119 | 7.588 | 0.413 | 2.662 | 8.964 | 1.081 |
| RMSE | 2.404 | 10.367 | 0.677 | 2.957 | 10.854 | 1.405 |
| MAPE | 3.832% | 13.186% | 0.726% | 4.348% | 15.625% | 1.900% |

## 5. Conclusion

In the energy consumption case study, the Ta-Method + LTB SNR performed better compared to the default Taguchi's T-Method and the T-Method + LTB SNR after being optimised. However, the Ta-Method + LTB SNR could not perform well in percentage yield and tensile strength of raw materials. Since energy consumption did not have the output range, the LTB could give maximum SNR for the output value. Improving the variability and adjusting the average to the target value will improve the SNR. Recommendations for future research are replacement of the Ta-Method with the Tb-Method, use of different types of SNR, and replacement of the Orthogonal Array for optimization process.

**Acknowledgement**
This research was not funded by any grant

**References**

[1] Tan, C. S., K. Maragatham, and Y. P. Leong. "Electricity energy outlook in Malaysia." In *IOP Conference series: earth and environmental science*, vol. 16, no. 1, p. 012126. IOP Publishing, 2013.

[2] National Energy Balance 2017,Energy Commision, Putrajaya, WP Putrajaya, Malaysia, 0128 - 6323, Oct 13, 2019.

[3] Harudin, Nolia, Khairul Rijal Jamaludin, F. Ramli, M. Nabil Muhtazaruddin, Z. M. Marlan, W. Z. A. W. Muhamad, and N. N. Jaafar. "An overview of taguchi'S T-method as A prediction tool for multivariate analysis." *Open International Journal of Informatics* 7, no. Special Issue 1 (2019): 158-166.

[4] Inoh, Junki, Yasushi Nagata, Keisuke Horita, and Arisa Mori. "Prediction accuracies of improved Taguchi's T methods compared to those of multiple regression analysis." *Journal of the Japanese Society for Quality Control* 42, no. 2 (2012): 265-277.

[5] Nakao, Yuto, and Yasushi Nagata. "Proposal for a new single imputation method using Taguchi's T-method." *Total Quality Science* 5, no. 3 (2020): 102-110.

[6] Matsui, Shushiro, and Yasushi Nagata. "Proposal for a univariate time series analysis method based on Taguchi's T-method." *Total Quality Science* 5, no. 1 (2019): 1-10.

[7] Kawada, Hiroshi, and Yasushi Nagata. "An application of a generalized inverse regression estimator to Taguchi's T-Method." *Total Quality Science* 1, no. 1 (2015): 12-21.

[8] Nishino, Keisuke, and Arata Suzuki. "Taguchi's T-method using median-median line for small sample with outliers." *Electronics and Communications in Japan* 102, no. 1 (2019): 49-56.

[9] Harudin, N., M. N. Muhtazaruddin, F. Ramlie, and WM Wan Zuki Azman. "Artificial bee colony for features selection optimization in increasing T-method accuracy." *International Journal of Engineering & Technology* 7, no. 4 (2018): 885-891.

[10] Harudin, Nolia, K. R. Jamaludin, M. Nabil Muhtazaruddin, F. Ramlie, S. H. Ismail, Wan Zuki Azman Wan Muhamad, and N. N. Jaafar. "Increasing T-method accuracy through application of robust M-estimatior." (2018).

[11] Marlan, Z. M., K. R. Jamaludin, F. Ramlie, N. Harudin, and N. N. Jaafar. "Determination of optimal unit space data for Taguchi's T-method based on homogeneity of output." *Open International Journal of Informatics* 7, no. Special Issue 1 (2019): 167-179.

[12] Teshima, Shoichi. *Quality recognition & prediction: Smarter pattern technology with the Mahalanobis-Taguchi system*. Momentum Press, 2012.

[13] Zhao, Hongying, and Wu Lifeng. "Forecasting the non-renewable energy consumption by an adjacent accumulation grey model." *Journal of Cleaner Production* 275 (2020): 124113.

[14] Lu, Hongfang, Feifei Cheng, Xin Ma, and Gang Hu. "Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: A case study of an intake tower." *Energy* 203 (2020): 117756.