An Overview of Taguchi' S T-Method as A Prediction Tool for Multivariate Analysis

N Harudin[1], Jamaludin KR[2], Ramlie F[3], M Nabil Muhtazaruddin[4], ZM Marlan[5], WZAW Muhamad[6], NN Jaafar

[1,2,3,4]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur
[1]Department of Mechanical Engineering, Universiti Tenaga Nasional, 43000 Kajang Selangor, Malaysia

nolia.harudin@gmail.com

Article history

Received:
22 Oct 2019

Received in revised form:
7 Nov 2019

Accepted:
4 Dec 2019

Published online:
25 Dec 2019

*Corresponding author:
nolia.harudin@gmail.com

Abstract

Analysis of prediction has attracted considerable interest in various fields. Taguchi's T-Method is a prediction method introduced by Genichi Taguchi in mid-year 2000, among several other Mahalanobis Taguchi system tools. It was explicitly created for the prediction of multivariate data. Taguchi's T-Method has shown that even with limited sample size, making a prediction based on historical data is possible. The key elements that have been adapted in reinforcing Taguchi's T-Method robustness are by introducing the unit-space principle and adaptation of the signal to the noise ratio (SNR) as a weighting as well as a zero-proportional theory, as proposed by Genichi Taguchi in a robust model. Taguchi's T-Method was widely practicing in Japan and began to be practiced by non-Japanese researchers due to its simplicity and simple understanding. Up to recent, various applications of Taguchi's T-Method been applied, which prove to be beneficial to industrial needs. This research paper outlines the T-method procedures by applying it in a few benchmark datasets and compare the accuracy with the existing multiple linear regression method for an overview. The results show that Taguchi's T-Method is better than multiple regression in dealing with limited sample data in which the sample size is smaller than the input variables. Taguchi's T-Method proved to have the ability to predict output with an acceptable range of prediction accuracy.

Keywords: *Taguchi's T-Method; Prediction; Multivariate; Signal to noise ratio (SNR); Small sample data*

1. Introduction

The Mahalanobis Taguchi System (MTS) is a well-known quantitative multivariate decisions tools and data analytics method developed by Genichi Taguchi that incorporates clustering, classification, and prediction analysis. MTS fundamental objective is to make exact expectations in multidimensional frameworks by building a measurement scale. Since early 2000 up to recent, various studies and published work available in the area of MTS, that has created a significant outcome and benefited society [1],[2],[3],[4],[5]. MTS methods can be divided into two categories. The first group is made up of methods that assess the abnormal and normal data points. The MT-Method, the MT-adjoint (MTA), and the Recognition-Taguchi (RT) are the conventional methods in this category. The second group consists of methods that predict future data values; this group is comprised of the Taguchi method (T-Method) and the Taguchi-Schmidt (TS) method[6]. The T-Method, also known as the Taguchi's T-Method, was developed

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

based on the essential elements of the Taguchi Quality Engineering System used to evaluate an overall system of prediction. This technique is widely practiced in Japan yet starts getting attention from researchers outside Japan. One of the benefits of Taguchi's T-Method is its ability to predict even when the sample size is limited. Work well with its very own capacity and contextual analysis lead to the intention of this paper to provide an overview of Taguchi's T-Method as a tool for multivariate analysis prediction. Taguchi's T-Method is a simple tool that the practitioner can easily understand and function well on a case-by-case basis, especially data of limited sample size.

2. An overview of Taguchi's T-Method

As described in the previous section, Taguchi's T-method is one of the methods built primarily for predictive purposes within MTS. It started to be introduced in mid-year 2000 by late Genichi Taguchi. SNR (η) is the crucial component that has been adapted as the weighting factor in the T-Method formulation. The regression model is developed linearly based on the zero-proportional principle with the linear regression line through zero-point (origin). The overall purpose of Taguchi's T-Method is to fulfill equation 2.1 below in calculating the integrated estimate output value ($\widehat{\mathrm{M}}$). There are four core fundamentals elements in Taguchi's T-Method, which are unit-space, zero-proportional, and weightage SNR concept. These elements are explained in brief in next sub-section for better understanding.

$$\textit{Integrated estimated output value, } \hat{M}_i = \frac{\eta_1 \times \left(\frac{X_{i1}}{\beta_1}\right) + \eta_2 \times \left(\frac{X_{i2}}{\beta_2}\right) + \cdots + \eta_j \times \left(\frac{X_{ij}}{\beta_j}\right)}{\eta_1 + \eta_2 + \cdots + \eta_j} \tag{2.1}$$

2.1. Unit-Space Concept

The unit space and signal space are defined for the normalization stage before the prediction model is constructed. Unit-space selection is one of the most critical processes to be identified in the early stages of the analysis. Unlike regular practice, which takes the average of individual independent and dependent variables as a reference point, the selection of unit-spaces is based on highly dense population data. Unit space should be as homogeneous as possible, and the option of unit space between low and high data of the chosen population will always be in the middle position. Figure 1 clarified this definition of unit space in a clear picture. Sample data selected as unit-space is excluded from the original data, and the remaining data is called signal data. The selected unit-space data is then being averaged for normalization purposes, as shown in equation 2.2.

$$\textit{normalized data } (X_{ij}, M_0) = \text{signal data} - \text{average of unit space} \tag{2.2}$$

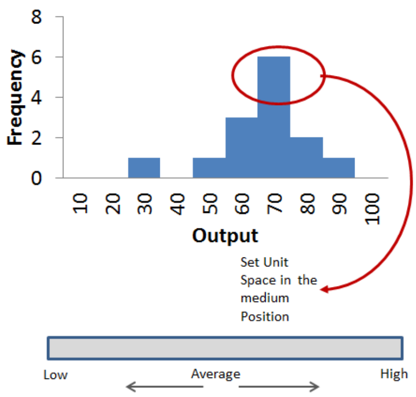
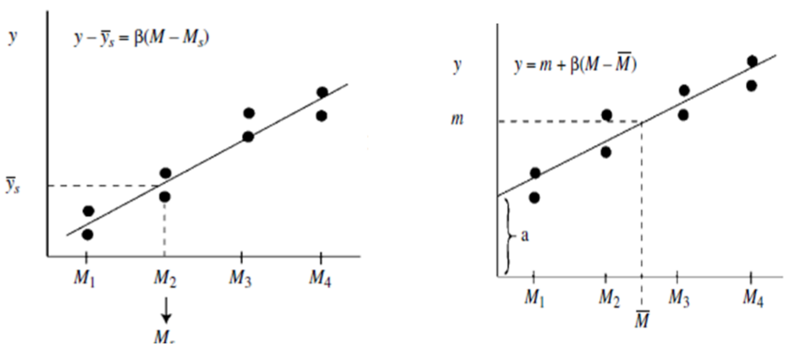


Figure 1. Unit-space selection concept [6]

2.2. The zero-proportional theory in Taguchi's T-Method

In determining the integrated estimate output value ($\widehat{\mathrm{M}}$) in equation 2.1, it should be noticed that $\left(\frac{X_{ij}}{\beta_j}\right)$ is defined for individual variables and been added up together without considering any interception point to be added as the typical linear regression model. The X_{ij} is referring to the normalized input while β_j is the probability coefficient of each data variable. This scenario is following the zero-point proportional equation. Since not all cases are ideally intercepted at zero-point, the use of reference-point, which in Taguchi's T-Method case is known as average unit-space help to achieve the zero-point proportional concept through the normalization process conducted. The zero intercept ensures that additional error due to the variation in the intercept is avoided. Optimizing the system close to the intercept ensures that the system can be adjusted down to the minimal value of *M* and *X*.

The created linear regression line will pass the zero points (origin) of the graph by following the definition of the proportional equation reference point [7]. Figures 2a and 2b illustrate the reference-point equation and typical linear equation difference for better illustration.



(a) **(b)**

Figure 2. Two different case of equation: a) reference-point proportional equation b) linear equation *[Note: the figure is used for illustration only since X represents the actual Y-axis in Taguchi's T-Method].*

2.3. Weightage Signal to Noise ratio, SNR (η) in Taguchi's T-Method

Signal to noise ratio (SNR) is a measure used to assess the quality of the measurement system. Emphasizing on three core elements (sensitivity, linearity, and variability), SNR is one of the critical components been adapted into Taguchi's T-Method formulation. The linearity defined as the correlation between input and output described by a straight line. The sensitivity is the magnitude of the slope to the line, and the variability defines the deviation from the line causes by noise factors[7]. By referring to equation 2.1, the SNR value is defined as the weightage component for individual input variables and is finally divided by the overall weight (total SNR). As shown in equation 2.3, the general SNR formula shows that the higher the variance, the lower the SNR is. The SNR weighing in equation 2.1 implicitly reflects the contribution of each variable to the overall prediction model. Variables with high variation will be given lower weightage to the overall prediction, while variables with low variation are vice versa.

$$SNR = \frac{sensitivity}{varibility} = \frac{(slope)^2}{variability} \tag{2.3}$$

3. Methodology

This section will discuss the procedures of calculating the Taguchi's T-Method as well as the case study involved and the tools used in analyzing the data.

3.1 Taguchi's T-Method Procedures

The aim of Taguchi's T-method, as stated earlier in section 2, is to predict the $\widehat{M}$ as highlighted in equation 2.1. In order to do this, it is appropriate to measure the proportional coefficient (β) and SN ratio (η) by item basis using normalized data *(Xij)* determined by equation 2.2. The SN ratio(η) theory used in Taguchi's T-method is the dynamic type. Therefore, equation 2.4 until equation 2.9 describes the procedures for measuring the dynamic SNR. The probability coefficient (β_M) in this study is following the least square theory similar to the classical linear regression model. Equation 2.10 shows how the β_M is calculated in Taguchi's T-Method theory. If the value of the SN ratio is defined as negative, it should be regarded as zero. It is seen that the higher SN ratio of an object will result in a higher degree of contribution to the model's overall estimate.

$$\text{Effective Divider, r} = \mathrm{M_1}^2 + \mathrm{M_2}^2 + \cdots + \mathrm{M}_l^2 \tag{2.4}$$

$$\text{Total Variation}, S_T = \mathrm{X_1}^2 + \mathrm{X_2}^2 + \cdots + \mathrm{X}_l^2 \tag{2.5}$$

$$\text{Variation of proportional term}, S_\beta = \frac{(M_1X_{11} + M_2X_{21} + \cdots + M_lX_{l1})^2}{r} \tag{2.6}$$

$$\text{Error variation}, S_e = S_T - S_\beta \tag{2.7}$$

$$\text{Error Variance}, V_e = \frac{S_e}{l-1} \tag{2.8}$$

$$\text{Duplicate SN ratio}, \eta = \left(\frac{(S_\beta - V_e)}{r\, V_e}\right) \tag{2.9}$$

$$\beta_M = \frac{M_1 X_{11} + M_2 X_{21} + \cdots . + M_l X_{l1}}{r} \tag{2.10}$$

Once the η and β of each variable been defined, it is easy to predict the unknown future output ($\hat{M}_{new,i}$). By having new values on input variables, each value will need to follow the procedure as in equation 2.2 in normalizing the input variables. The new values are deducted with the defined unit-space value for respective variables. Once the input value been normalized, the prediction for the new unknown output ($\hat{M}_{new,i}$) can be easily calculated, which follow the equation 2.1. The new output value is calculated sample by sample basis. For the other unknown output estimation, the same process must be repeated. In order to reflect the actual value as the raw data before the normalization is done, the new output value will have to be added back with the average unit-space output value known as M_0. The $\hat{Y}_i$ is the finalized actual predicted value.

$$\hat{Y}_i = \hat{M}_{new,i} + M_0 \tag{2.11}$$

3.2. Case Study use in this analysis

The first and second case studies were taken from historical data recorded by Nuclear Agency Malaysia team to predict power consumption for a thermal energy storage system (TES). The **first case study** is conducted to predict power consumption based on chiller performance. It involved 10 days of the data for training and four days of data for testing with 44 variables. TES serves as a battery to the area's air-conditioning system, helping to reduce energy consumption by having high consumption during the off-peak hour as charging happens during the night shift.

The JD power case study data is provided by one of the experienced Taguchi pioneer, Mr. Shin Taguchi. The data is a prediction of the satisfaction index among several car models with regards to several attributes and technical specifications. It involves 44 variables with 15 training dataset and six testing dataset. The third case study is a real case study conducted in predicting calories burn (kcal) for ice-skating activity at the icescape ice rink, IOI city mall. Every respondent participating in this study was given a pedometer to be wear in their wrist prior skating activity. They also obtained a basic questionnaire to obtain some useful information to be used in the study. In total, 17 respondents were selected as the training dataset with six respondents as the predicted test data and 15 variables taken into consideration. The final case study is taken from [8], which contains underwater measuring percentage figures of body fat and different body circumference measurements. Bodyfat data used in this study is split into training and testing which 176 for training and 76 for the test with 14 variables. Bodyfat dataset has a normally distributed output data with considered stable and an excellent example in setting the model prediction for nominal cases.

3.3. Analysis Tools used for analyzing results.

The analysis for Taguchi's T-Method is conducted using the software provided by [6]. In this study, Taguchi's T-method results are compared to the multiple regression method, which is analyzed using Matlab code as equation 2.11 and equation 2.12 below. lmModel is a prediction model generated by training data, while ypred is the prediction made for the test dataset.

$$\mathrm{lmModel} = \mathrm{fitlm(dataTraining\ (X),\ dataTraining\ (M))} \tag{2.12}$$

$$\mathrm{ypred} = \mathrm{predict(lmModel,\ dataTesting(X))} \tag{2.13}$$

The accuracy of the result is measured in terms of root mean squared (RMSE) and coefficient of determination (R^2) as stated in equation 2.13 and 2.14 below

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(\text{actual-predicted})^2}{n}} \tag{2.14}$$

$$R^2=1-\left(\frac{\sum_i\left(M_i-\widehat{M_i}\right)^2}{\sum_i(M_i-\overline{M})^2}\right) \tag{2.15}$$

4. Results and Discussions

The main idea of this paper is to provide an overview of Taguchi's T-Method as a relatively new prediction tool, especially to the practitioner outside Japan. One of the significant contributions of Taguchi's T-method is its ability to predict with low sample data. Unlike multiple regression analysis which 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 [6], [9].

Table 1 below shows the compilation results between Taguchi's T-method and Multiple linear regression methods for three case studies with low sample data. It shows that Taguchi's T-method performing better than multiple linear regression for all cases. Since these three case studies involve very minimum sample data, another benchmark dataset is taken, which is intended to see how both methods are performing with the increasing number of training data and larger sample sizes. The body-fat dataset analysis results are shown in Table 2.2.

Table 1. Prediction accuracy between Taguchi's T-Method and Multiple Linear Regression Method for low sample data case studies

Dataset	no. of training data/no. of test data	no. of variables	T-Method		Multiple-linear regression
			R^2	RMSE	RMSE
Power consumption	10(4)	15	0.7039	**61.3838**	>100
Prediction based on Chillers performance	10(4)	44	0.7488	**4.4507**	38.5279
JD power	15(6)	44	0.9138	**0.6935**	3.6608
Colories burn	17(6)	14	0.9742	0.0057	2.7708
Prediction based on air-conditioning performance	6(2)	17	0.9757	1.98	37.1201

Table 2. Prediction accuracy between Taguchi's T-Method and Multiple Linear Regression Method for Bodyfat dataset (larger sample data)

Method		Training samples						
		10	14	20	30	50	100	176
T-Method	R^2	0.9997	0.99	0.9838	0.9219	0.966	0.9829	0.9906
	RMSE	**0.1682**	**0.3922**	**0.9**	2.714	1.5615	0.8904	0.503
Multiple-linear regression	R^2	NaN	NaN	0.996	0.965	0.981	0.988	0.993
	RMSE	26.0848	32.5183	2.2484	**1.5571**	**0.7715**	**0.4038**	**0.2977**

Note: Analysis is based on 14 input variables, one output variable, and 76 testing data

The findings in Table 2 were consistent with Table 1 as well as the assertion highlighted by [1], [10], which reported that Taguchi's T-Method is performing better than multiple regression analysis in the case of sample size is smaller than double the number of explanatory variables. This trend is shown in Table 2, with Taguchi's T-method having better accuracy up to the training sample 20 and not after 30 samples onwards. Therefore, as the number of samples exceeds the explanatory variables, which is 14, the estimate of regression coefficients is unreliable, which could lead to multicollinearity, and predictive accuracy declines intensely. However, since Taguchi's T-Method regression formulas are configured for each item, a prediction is possible, even using a few sample data [1]. The R^2 value for the Multiple Regression method in Table 1 and training samples 10 and 14 in Table 2 shows that the computational is not feasible due to a smaller number of samples than the explanatory variables (input variables). The RMSE value as well is significantly higher compared to Taguchi's T-Method for training samples 10,14 and 20 in Table 2

One of the elements that need to look into more detail is the weightage SNR. Referring to the analysis trend shown in Table 2, an analysis of the weightage SNR trend for individual input variables is plotted and shown in figure 3 below. The higher the SNR value, the lower the variation of the data is, and the higher the weightage of the particular input variables towards overall prediction. Variable 2 and 4 are excluded from the trend analysis since it involves zero SNR value that is due to negative value led by higher error variance compared to the variation of the proportional term, as shown in equation 2.8 and highlighted previously. The trend shows that input variable 1 with 10 sample data is significantly high SNR compared to others. The remaining data show minimum differences among the SNR weightage.

What can be highlighted in figure 3 is that the overall weightage SNR does not show significant differences except for the variables 1 for 10 samples, which indirectly shows that Taguchi's T-Method indeed can predict with an acceptable range of accuracy throughout the number of sample data considered with simple calculation procedures. There is a various study conducted recently in searching for a possibility to improve Taguchi's T-method prediction model accuracy and various findings been shared which proved to show that improvement is made and still have opportunity to be further enhanced [1], [2], [9], [11]–[16].

Figure 3. SNR weightage trend by the increment of sample data for respective variables

5. Conclusion

In this overview study on Taguchi's t-Method, it can be summarized that:

1. Taguchi's T-Method is a useful prediction tool for small sample size, which lower than the explanatory variables.
2. Prediction using Taguchi's T-method is straightforward and easy to be computed.

There comes in mind, is it a need for a small sample data analysis in the industry? The answer is yes, and the manufacturing industry mainly deals with new product development that has minimal samples to be tested. Besides, the industry that deals with high-cost material is also facing problems in getting extensive samples data for the test development stage or even the production stage. Another example is the end of life products that generally very limited and takes years to get extensive samples data for any further analysis, perhaps. Taguchi's T-Method is just one of the various prediction tools that have capability in predicting for both limited and extensive sample data with its SNR weightage as its robust element. The industrial practitioner with particular expertise should have no problem to apply this method since useful data comes from experience is much more representable, and a tool only helps in visualizing the overall performance accurately.

Acknowledgment

This work is supported by the Potential Academic Staff (PAS) Grant Scheme awarded by the Universiti Teknologi Malaysia (Grant No. Q.K130000.2756.03K33).

6. References

[1] S. Negishi, Y. Morimoto, S. Takayama, and A. Ishigame, “Daily Peak Load Forecasting by Taguchi’s T Method,” Electr. Eng. Japan (English Transl. Denki Gakkai Ronbunshi), vol. 201, no. 1, pp. 57–65, 2017.
[2] N. Harudin et al., “Increasing T-Method Accuracy through Application of Robust M-estimator,” J. Adv. reseach Dyn. Control Syst., no. 09-special issue, pp. 1287–1294, 2018.
[3] B. Buenviaje, J. E. Bischoff, R. A. Roncace, and C. J. Willy, “Mahalanobis-Taguchi System to Identify Preindicators of Delirium in the ICU,” IEEE J. Biomed. Heal. Informatics, 2016.
[4] F. Ramlie, K. R. Jamaludin, R. Dolah, and W. Z. A. W. Muhamad, “Optimal feature selection of Taguchi character recognition in the Mahalanobis-Taguchi system using bees algorithm,” Glob. J. Pure Appl. Math., vol. 12, no. 3, pp. 2651–2671, 2016.
[5] W. Z. A. W. Muhamad, K. R. Jamaludin, Z. R. Yahya, and F. Ramlie, “A hybrid methodology for the mahalanobis-taguchi system using random binary search-based feature selection,” Far East J. Math. Sci., 2017.
[6] S. Teshima, Y. Hasegawa, and K. Tatebayashi, Quality Recognition and Prediction: Smarter Pattern Technology with the Mahalanobis-Taguchi System. New York: Momentum Press, 2012.
[7] Y. F. William and C. C. M., “The dynamic signal-to-noise method and metrics,” in Engineering Methods for Robust Product Design—Using Taguchi Methods in Technology and Product Development, Addison Wesley Publishing Company, 1995, p. 403.
[8] Billy Fang, “Body fat dataset,” http://statistics.berkeley.edu/, 2017.
[9] S. Suguru and N. Yasushi, “A Study of Multivariate Taguchi ’ s T Method,” in 16th ANQ Congress<<Future Quality Leadership in IT-Age>>, 2018, pp. 250–262.
[10] J. Inoh, Y. Nagata, K. Horita, and M. Arisa, “’ PredictionAccuracies of lmproved Taguchi s T Methods Compared to those of MultipleRegression Analysis,” J. japanese Soc. Qual. Control, vol. 42, no. 2, pp. 265–277, 2012.
[11] N. Harudin, K. R. Jamaludin, M. N. Muhtazaruddin, F. Ramlie, and W. Z. A. W. Muhamad, “Artificial Bee Colony for Features Selection Optimization in Increasing T-Method Accuracy,” Int. J. Eng. Technol., vol. 7, no. 4.35, pp. 885–891, 2018.
[12] N. Harudin, K. R. Jamaludin, M. Nabil Muhtazaruddin, F. Ramlie, and W. Z. A. W. Muhamad, “A feasibility study in adapting Shamos Bickel and Hodges Lehman estimator into T-Method for normalization,” IOP Conf. Ser. Mater. Sci. Eng., vol. 319, no. 1, 2018.
[13] S. Matsui and Y. Nagata, “Proposal for a univariate time series analysis method based on Taguchi ’ s T - method,” in 16th ANQ Congress<<Future Quality Leadership in IT-Age>>, 2018, pp. 227–237.
[14] K. Satoshi and N. Yasushi, “Application of Taguchi ’ s T-method to Circumferential Data Considering,” in 16th ANQ Congress<<Future Quality Leadership in IT-Age>>, 2018, pp. 238–249.
[15] K. Nishino and A. Suzuki, “Taguchi’s T-method using median-median line for small sample with outliers,” Electron. Commun. Japan, pp. 49–56, 2018.
[16] H. Kawada and Y. Nagata, “Studies on the Item Selection in Taguchi’s T-Method,” J. Japanese Soc. Qual. Control, vol. 45, no. 2, pp. 179–193, 2015.