**MEKATRONIKA**
e-ISSN: 2637-0883
VOL. 5, ISSUE 2, 6 – 13
DOI: https://doi.org/10.15282/mekatronika.v5i2.9581

**ORIGINAL ARTICLE**

# Application of Mahalanobis-Taguchi System in Rainfall Trends at UMP Gambang Campus

M.A.M. Jamil[1], M.Y. Abu[1,*], S.N.A.M. Zaini[1], N.H. Aris[1], N.S. Pinueh[1] , W.Z.A.W. Muhamad[2], F. Ramlie[3], N. Harudin[4], E. Sari[5], N.A.A.A. Ghani[6], N.N. Jaafar[1]

[1]Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600, Pekan, Pahang, Malaysia
[2]Institute of Engineering Mathematics, Universiti Malaysia Perlis, Kampus Pauh Putra, Perlis 02600 Arau, Malaysia
[3]Razak Faculty of Technology and Informatics, Department of Mechanical Engineering, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100, Kuala Lumpur, Malaysia
[4]Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia
[5]Universittas Trisakti, Faculty of Industrial Technology, Department of Industrial Engineering, 11440, Kyai Tapa No 1, West Jakarta, Indonesia
[6]Faculty of Civil Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Pahang, Malaysia

**ABSTRACT** – Rainfall is a variable meteorological phenomenon that exhibits spatial variability across different locations. Weather stations collect a wide range of parameters to monitor and analyze rainfall patterns. However, not all parameters are equally significant or efficient in performing classification and optimization tasks. In this study, we propose the use of the Mahalanobis-Taguchi system (MTS) method to classify rainfall occurrences by RT-Method and optimize the parameter selection process by T-Method. The data were collected by weather station Vantage Pro2 in UMP Gambang. By applying RT- Method, we can classify the data sample in term of MD for November, May and April while reducing the number of parameters to only those that significantly contribute to the classification, which from 16 parameters to 8 parameters using T-Method. This approach provides a streamlined and efficient methodology for analyzing rainfall patterns and optimizing weather station data collection processes.

**ARTICLE HISTORY**
Received: 28th July 2023
Revised: 18th Aug 2023
Accepted: 6th Sept 2023
Published: 27th Sept 2023

**KEYWORDS**
*Mahalanobis-Taguchi*
*Mahalanobis distance*
*Rainfall*
*Classification*
*Optimization*

## INTRODUCTION

Rainfall distribution refers to the spatial and temporal patterns of rainfall in each area. It plays a crucial role in various natural processes and human activities, such as agriculture, hydrology, and climate studies. Understanding rainfall distribution is essential for effective water resource management, flood prediction, and drought mitigation strategies.

Rainfall distribution can vary both spatially and temporally. Spatial variations in rainfall distribution are influenced by factors such as topography, vegetation, and climate [1]. For example, in mountainous regions, the distribution of rainfall is often characterized by orographic effects, where rainfall increases with elevation due to the lifting of moist air masses [1]. Temporal variations in rainfall distribution can be observed at different time scales, ranging from daily to seasonal and annual variations. Variability in rainfall can have significant implications for water availability, agriculture, and ecosystem dynamics [2].

In Malaysia, rainfall distribution is influenced by the country's geographical location and monsoon climate. Malaysia experiences two distinct monsoon seasons, the northeast monsoon (NEM) and the southwest monsoon (SWM) [3]. The NEM, occurring from November to March, brings heavy rainfall to the east coast of Peninsular Malaysia and the states of Sabah and Sarawak. The SWM, from May to September, brings rainfall to the west coast of Peninsular Malaysia. The Titiwangsa mountain range in Peninsular Malaysia plays a crucial role in rainfall distribution, causing lower rainfall along the west coast during the winter monsoon [4].

Understanding the significant parameters that contribute to the classification of rainfall is crucial for accurate analysis and prediction. A paper identified the association of precipitation extremes in Malaysia with El Niño and La Niña events. They found that these events significantly influence the characteristics of extreme rainfall in the country [4]. The presence of long-term persistence and regional sea surface temperature anomalies also play a role in the spatial distribution and variability of rainfall in Malaysia [5].

This research uses MTS to classify and optimize the parameters collected to find the significant parameters. MTS is a powerful methodology that combines the Mahalanobis distance measure with Taguchi's orthogonal array design for analyzing multivariate data and making quantitative decisions [6]. It has been widely applied in various fields, including engineering, manufacturing, medicine, and more [7]. The MTS offers several advantages over traditional approaches, such as linear discriminant analysis and logistic regression, as it can handle non-linear patterns and consider correlations between independent variables or factors [8].

*CORRESPONDING AUTHOR | M. Y. Abu | ✉ myazid@umpsa.edu.my

Furthermore, the MTS has shown promise as a binary classification algorithm for handling imbalanced data. It has been benchmarked against other algorithms, such as Support Vector Machines (SVM) and Naive Bayes, and has demonstrated superior performance, particularly for imbalanced data with a high ratio [9]. By utilizing the Mahalanobis distance measure, the MTS can effectively classify and predict data in a multidimensional [10]. This makes it a valuable tool for decision-making processes that require accurate and robust predictions.

Overall, the MTS offers a systematic and effective approach for quality improvement and decision-making, making it a valuable tool in various industries and domains. Its ability to analyze complex data, identify patterns, and optimize processes has contributed to its widespread adoption and application in diverse fields [7].

## RESEARCH METHODOLOGY

The data collection of various parameters was collected in UMP Gambang using Vantage Pro2 Weather Station. The data was gathered once a month from the weather station. The raw data recorded in the weather station console was then transferred to the laptop. The data were recorded every 15 minutes for every parameter. Table 1 shows the parameters collected from the weather station.

**Table 1.** Parameter

| Numbers | Parameters | Unit |
|---|---|---|
| 1 | Temperature Outside | ºC |
| 2 | High Temperature | ºC |
| 3 | Low Temperature | ºC |
| 4 | Outside Humidity | ºC |
| 5 | Dew Point | ºC |
| 6 | Heat Index | - |
| 7 | Barometric Pressure | mb |
| 8 | Rain | mm |
| 9 | Rain Rate | mm/hr |
| 10 | Cool Degree-Day | ºC |
| 11 | Inside Temperature | ºC |
| 12 | Inside Humidity | % |
| 13 | Inside Dew | ºC |
| 14 | Inside Heat Index | ºC |
| 15 | Inside EMC | - |
| 16 | Inside Air Density | $kg/m^3$ |

RT Method was utilized for the classification because it could classify the parameters into two variables. The RT Method defines the unit space and signal in accordance with rain. Unit space will be when it is not raining, whereas signal data will be when it is raining. Using Eq. (1), the average value for each parameter within the unit space was calculated. Next, the following Eqs. (2) - (4) can be calculated.

$$\bar{x}_j = \frac{1}{n}\left(x_{1j} + x_{2j} + \cdots + x_{nj}\right) \tag{1}$$

$$Sensitivity, \beta_1 = \frac{L_1}{r} \tag{2}$$

$$Linear\ equation, L_1 = \bar{x}_1 x_{11} + \bar{x}_2 x_{12} + \cdots + \bar{x}_k x_{1k} \tag{3}$$

$$Effective\ divider, r = \bar{x}_1{}^2 + \bar{x}_2{}^2 + \cdots + \bar{x}_k{}^2 \tag{4}$$

Then, the total variations $S_T$, variation of proportional term $S_\beta$, error variation $S_e$, and error variance $V_{el}$, are shown in Eq. (5), Eq. (6), Eq. (7), and Eq. (8) respectively.

$$Total\ Variation, S_{T1} = x_{11}{}^2 + x_{12}{}^2 + \cdots + x_{1k}{}^2 \tag{5}$$

$$Variation\ of\ proportional\ term, S_{\beta 1} = \frac{L_1{}^2}{r} \tag{6}$$

$$Error\ Variation, S_{el} = S_{T1} - S_{\beta 1} \tag{7}$$

$$Error\ variance, V_{el} = \frac{S_{el}}{k-1} \tag{8}$$

The standard Signal to Noise (SN) ratio $\eta_1$ is then calculated as stated in the Eq. (9). The greater the value of $\eta_1$, the stronger the relationship between the input and output.

$$Standard\ SN\ Ratio\ \ \eta_1 = \frac{1}{V_{e1}} \tag{9}$$

The computation of two variable $Y_1$ and $Y_2$ are computed by the sensitivity $\beta$ standard SN Ratio $\eta$ using Eq. (10) and Eq. (11).

$$Y_{i1} = \beta_i \ (i = 1,2, \dots, n) \tag{10}$$

$$Y_{i2} = \frac{1}{\sqrt{\eta_i}} = \sqrt{V_{ei}} \ (i = 1,2, \dots, n) \tag{11}$$

Then, the mean for $Y_1$ and $Y_2$ are computed for all the sample of the Unit Space as stated in the Eq. (12) and Eq. (13).

$$\bar{Y}_1 = \frac{1}{n}(Y_{11} + Y_{21} + \cdots + Y_{n1}) \tag{12}$$

$$\bar{Y}_2 = \frac{1}{n}(Y_{12} + Y_{22} + \cdots + Y_{n2}) \tag{13}$$

Finally, Mahalanobis Distances (MD) of the sample is calculated through Eq. (14)

$$Mahalanobis\ Distance, D^2 = \frac{YA^{-1}Y^T}{k} \tag{14}$$

For the signal data, the sensitivity $\beta_l$ and the linear formula $L^{'}$ were calculated using Eq. (2) and Eq. (3), and the effective divider $r$ were used in the unit space. $Y_1$ and $Y_2$ variables are calculated using signal data β and ŋ. By using Eq. (10), the $Y_1$ value of β can be computed, and Eq. (11) is converted for $Y_2$ so that any scattering from normal conditions can be measured. Finally, Eq. (14) yields the MD value.

For the optimization, T Method 1 were deployed to calculate the degree of contribution in the rainfall data. The MD calculated by the RT Method were used as the output value. Eq. (15) and Eq. (16) determine the average values for each parameter and the output average from the number of samples.

$$\bar{x}_j = \frac{1}{n}\left(x_{1j} + x_{2j} + \cdots + x_{nj}\right) \tag{15}$$

$$\bar{y} = M_0 = \frac{1}{n}\left(x_{1j} + x_{2j} + \cdots + x_{nj}\right) \tag{16}$$

The average parameter and output values determined the unit space. The signal data included unselected sample data. According to Eq. (17) and Eq. (18), the signal data sample was normalized.

$$X_{ij} = x'_{ij} - \bar{x}_j \tag{17}$$

$$M_i = y'_{ij} - M_j \tag{18}$$

For each parameter, compute proportional coefficient, β and SN ratio, η using Eqs. (19) – (25).

$$Proportional\ Coefficient\ \beta_1 = \frac{M_1X_{11} + M_2X_{21} + M_lX_{l1}}{r} \tag{19}$$

$$\text{SN Ratio } \eta_1 = \begin{cases} \dfrac{\frac{1}{r}(S_{B1} - V_{e1})}{V_{el}} & (When\ S_{\beta 1} > V_{el)} \\ 0 & (When\ S_{\beta 1} < V_{el}) \end{cases} \tag{20}$$

$$Effective\ divider, r = M_1^{\,2} + M_2^{\,2} + \cdots + M_l^{\,2} \tag{21}$$

$$Total\ Variation\ S_{T1} = X_{11}^{\,2} + X_{21}^{\,2} + \cdots + X_{lk}^{\,2} \tag{22}$$

$$Variation\ of\ Proportional\ term\ S_{\beta 1} = \frac{(M_1X_{11} + M_2X_{21} \cdots + M_lX_{l1})^2}{r} \tag{23}$$

$$Error\ Variation\ S_{el} = S_{T1} - S_{\beta 1} \tag{24}$$

$$Error\ Variance\ V_{el} = \frac{S_{el}}{l-1} \tag{25}$$

Eq. (26) shows how the β and η for each parameter were used to figure out the summed estimate value of the signal data.

$$\hat{M}_i = \frac{\eta_1 \times \frac{X_{i1}}{\beta_1} \times \eta_2 \times \frac{X_{i2}}{\beta_2} + \cdots + \eta_k \times \frac{X_{ik}}{\beta_{ik}}}{\eta_1 + \eta_2 + \cdots + \eta_k} \tag{26}$$

Next, Eqs. (27) – (33) were used to compute the integrated estimate SN ratio.

$$\text{Integrated SN Ratio } \eta_1 = 10\log\left(\frac{\frac{1}{r}(S_{B1} - V_e)}{V_e}\right) \tag{27}$$

$$Linear\ Equation, L = M_1\hat{M}_1 + M_2\hat{M}_2 + \cdots + M_l\hat{M}_l \tag{28}$$

$$Effective\ divider\ r = M_1^{\,2} + M_2^{\,2} + \cdots + M_l^{\,2} \tag{29}$$

$$Total\ Variation\ S_T = \hat{M}_1^{\,2} + \hat{M}_1^{\,2} + \cdots + \hat{M}_1^{\,2} \tag{30}$$

$$Variation\ of\ Proportional\ term\ S_\beta = \frac{L^2}{r} \tag{31}$$

$$Error\ Variation\ S_e = S_T - S_\beta \tag{32}$$

$$Error\ Variance\ V_e = \frac{S_e}{l-1} \tag{33}$$

The estimated SN ration degrades when a parameter is missing that indicates its importance. Level 1 and level 2 of the OA are utilised for evaluation purposes. The SN ratio may be estimated using OA under different circumstances. Level 1 of the OA is a parameter, while level 2 is not, as indicated by the fact that the OA has two levels. The difference between the SN ratio averages for levels 1 and 2 for each parameter is used to ascertain the relative importance of the parameters in terms of the estimated SN ratio. The degree of contribution was calculated using Eq. (34).

$$\text{Degree of Contribution} = \overline{SNR}_{level-1} - \overline{SNR}_{level-2} \tag{34}$$

## RESULT AND DISCUSSION

The scatter diagrams was constructed from the RT Method result according to the monsoon phenomena. For northeast monsoon, November was selected, meanwhile southwest and transition May and April were selected. In the graph, the variable $Y_1$ and $Y_2$ computed from the RT Method were created to display the classification between the unit space and signal data. The horizontal line represents $Y_1$, and vertical line represents $Y_2$. The unit space group (blue dotted) for November, May and April were 2735, 2901, and 2747 while the signal data (orange dotted) were 141, 75, and 133 respectively.

In November, the average unit space value was 1 and the average signal data value was 12.1333. The highest November unit space MD value was 9.3167, while the lowest was 0.0002. For November signal data, the highest MD value was 405.6236 and the lowest was 0.0009. In May, the average unit space value was 1 and the average signal data value was 9.1690. The maximum MD value for the May unit space was 6,141, and the minimum was 0.0001. The highest May signal data MD value was 162.2498, while the lowest was 0.0104. In April, the average unit space value was 1 and the average signal data value was 11.0716. The highest unit space MD value for April was 11,5413, while the lowest was 0.0007. The maximum MD value for April signal data was 174.5749 and the minimum MD value was 0. 0043.

The result demonstrates that the measurement scale constructed by all the variables is substandard due to the overlap between the unit space and the signal data, but it is still acceptable because the average of the signal data is not within the unit space's range value.

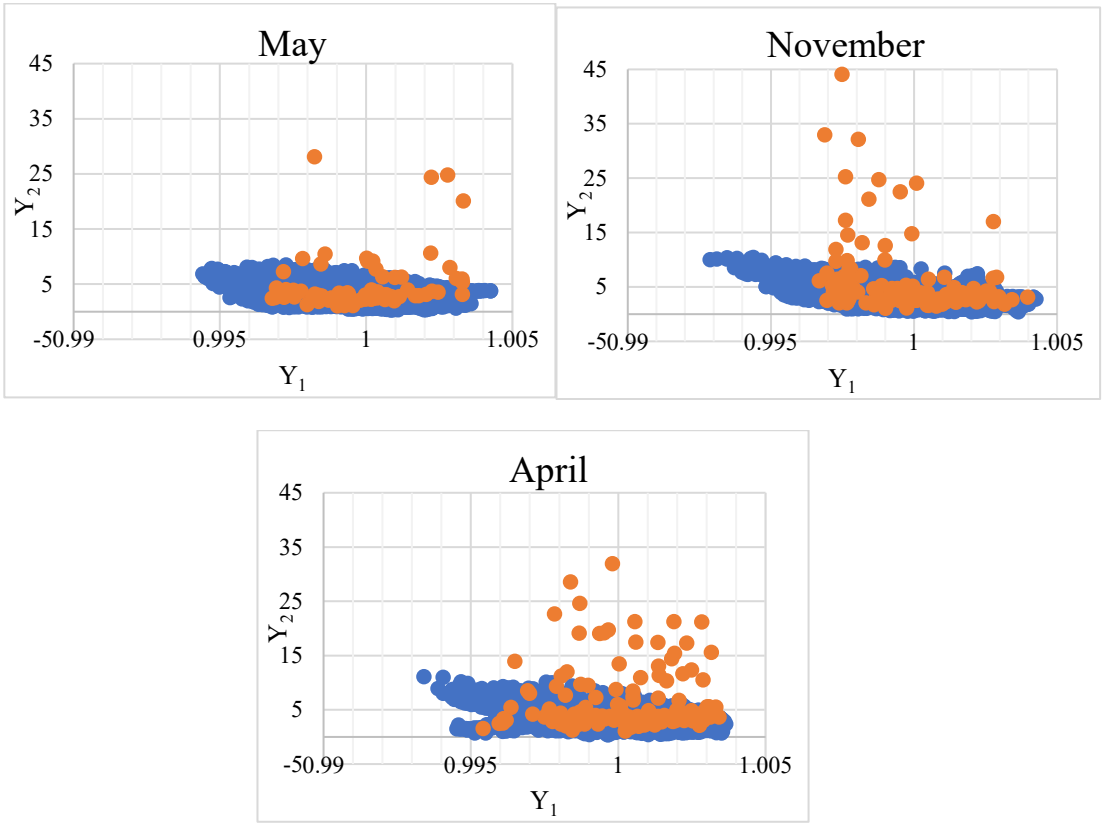



**Figure 1.** Scatter Diagrams of RT Result

In T Method 1, the samples were sorted based on the output value, which in this case the MD value as shown in Figure 2 for November, and the unit space was chosen based on the average value, with 5 samples set to be in the centre of the dark blue dot while the remaining samples were defined as signal data. For the months of May and April, Data (post-sort) were also generated to define the unit space and signal data.

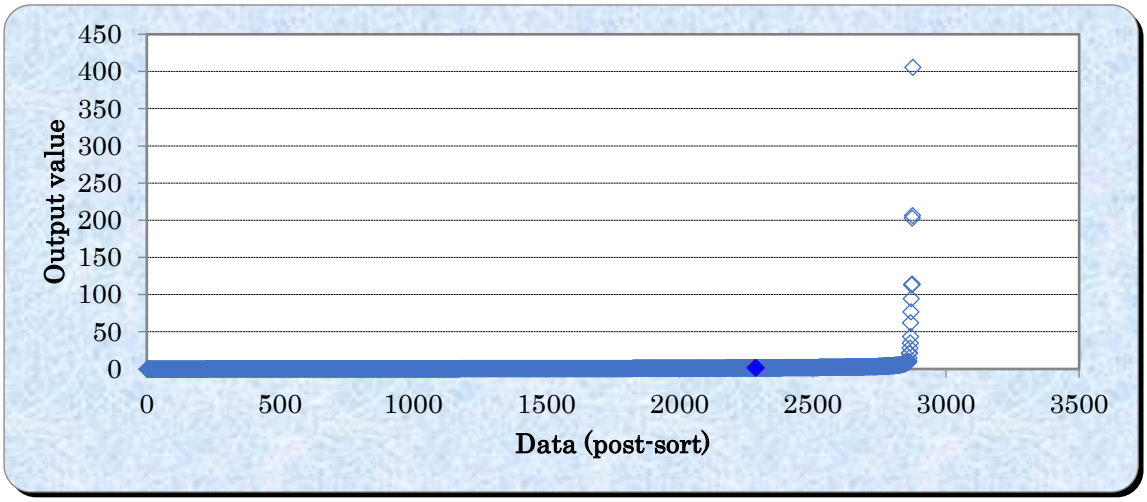
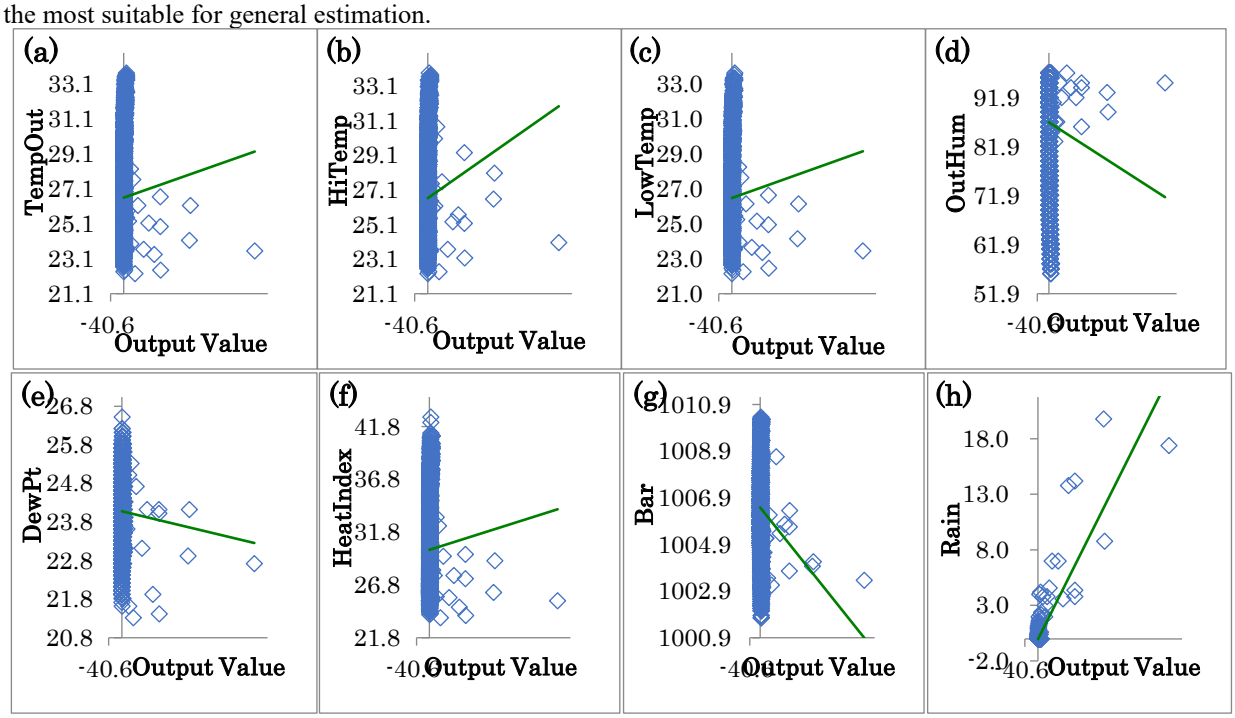



**Figure 2.** Data (post-sort) for November

Figure 3 illustrates the relantionship between a parameter and its output value for November. From the relationship between the normalised output value and the normalised parameter values, the SN ratio and proportional coefficients were determined. The relationship between the output value and the parameter approaches a straight line as the SN ratio increases. The highest SN ratio is 0.0317 in Figure 3(i) which is parameter 9, and the proportional coefficient is positive. In addition, the graph demonstrates that the parameter is suitable for general estimation purposes. Figure 3(n) has the lowest SN ratio -3x10$^{-6}$, and the proportional coefficient is negative which is parameter 14, indicating that the parameter is not particularly helpful for estimation in general.

For the month of May, the parameter with the highest SN ratio is 9 with a proportional coefficient of 0.1196, and the parameter with the lowest SN ratio is 13 with a proportional coefficient of -1x10$^{-6}$. The highest SN ratio for April is 0.0845 at parameter 9 with a positive proportional coefficient, while the lowest SN ratio for April is 1x10$^{-5}$ at parameter 6 with a positive proportional coefficient. According to the SN ratio on three month, parameter 9 which is the Rain Rate is the most suitable for general estimation.



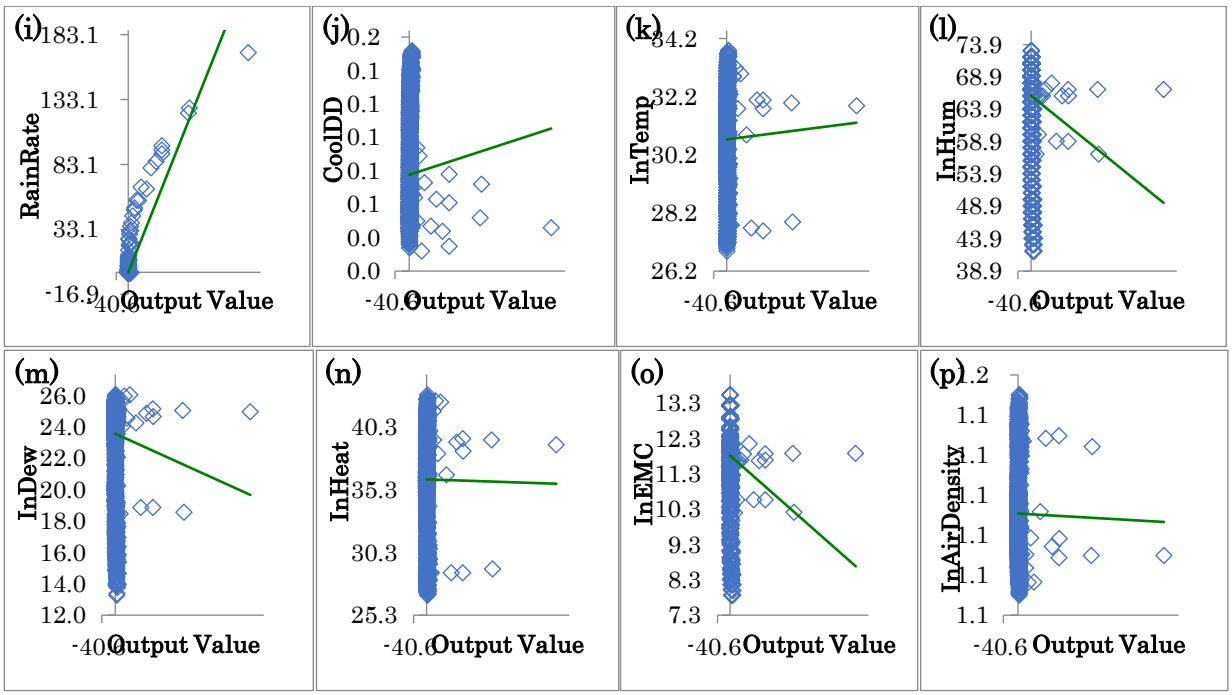



**Figure 3.** Scatter of output value and parameter

Figure 4 illustrates a distribution of actual and estimated signal data values yielding an $R^2$ value of 0.7984 and SN ratio of -14.61db for November, according to a general estimate. The $R^2$ values for the months of May and April are 0.7883 and 0.7813, respectively, while the SN ratio for general estimation are -8.738db and -14.61db. Equations (34), (35) and (36) respectively depict the line equations for November, May, and April.

$$y = 1.0157x \quad (34)$$

$$y = 1.021x \quad (35)$$

$$y = 1.0323x \quad (36)$$

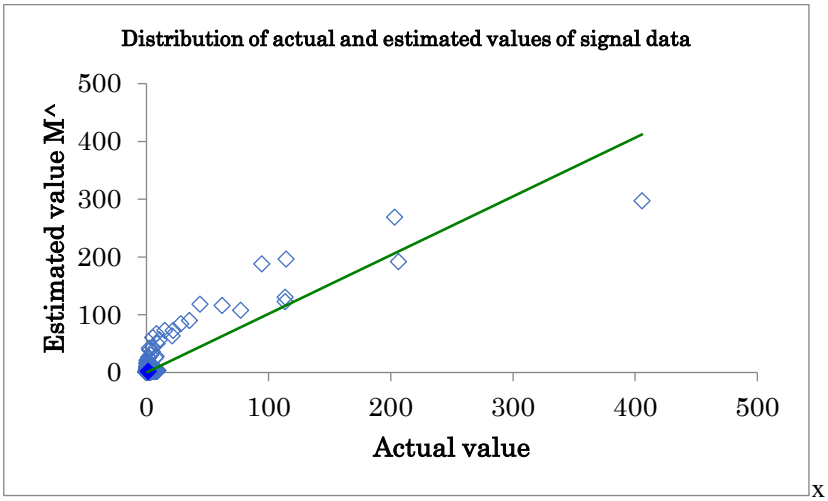



**Figure 4.** Distribution of actual and estimated signal data values for November

Figure 5 show the outcome of analysis for identifying critical parameters and optimizing the recognition and prediction system uses by MTS. The positive bar graph shows the positive degree of contribution, as it means that the use of parameter is affecting the elevation of the MD output. Whereas negative bar graph indicates negative degree of contribution, which means the use of parameter affecting in lowering the output of MD. For all 3 months, parameters 1, 3, 8, 9, 11, 12, 13 and 15 are positive degree of contribution, whereas parameters 2, 4, 5, 6, 7, 10, 14 and 16 are negative degree of contribution. From Figure 5, this research optimized 16 parameters into 8 parameters by reducing 8 parameters using the MTS procedures for all three months. Note that the reduced number of parameters suggested by MTS for future prediction and classification purposes are the significant parameters.

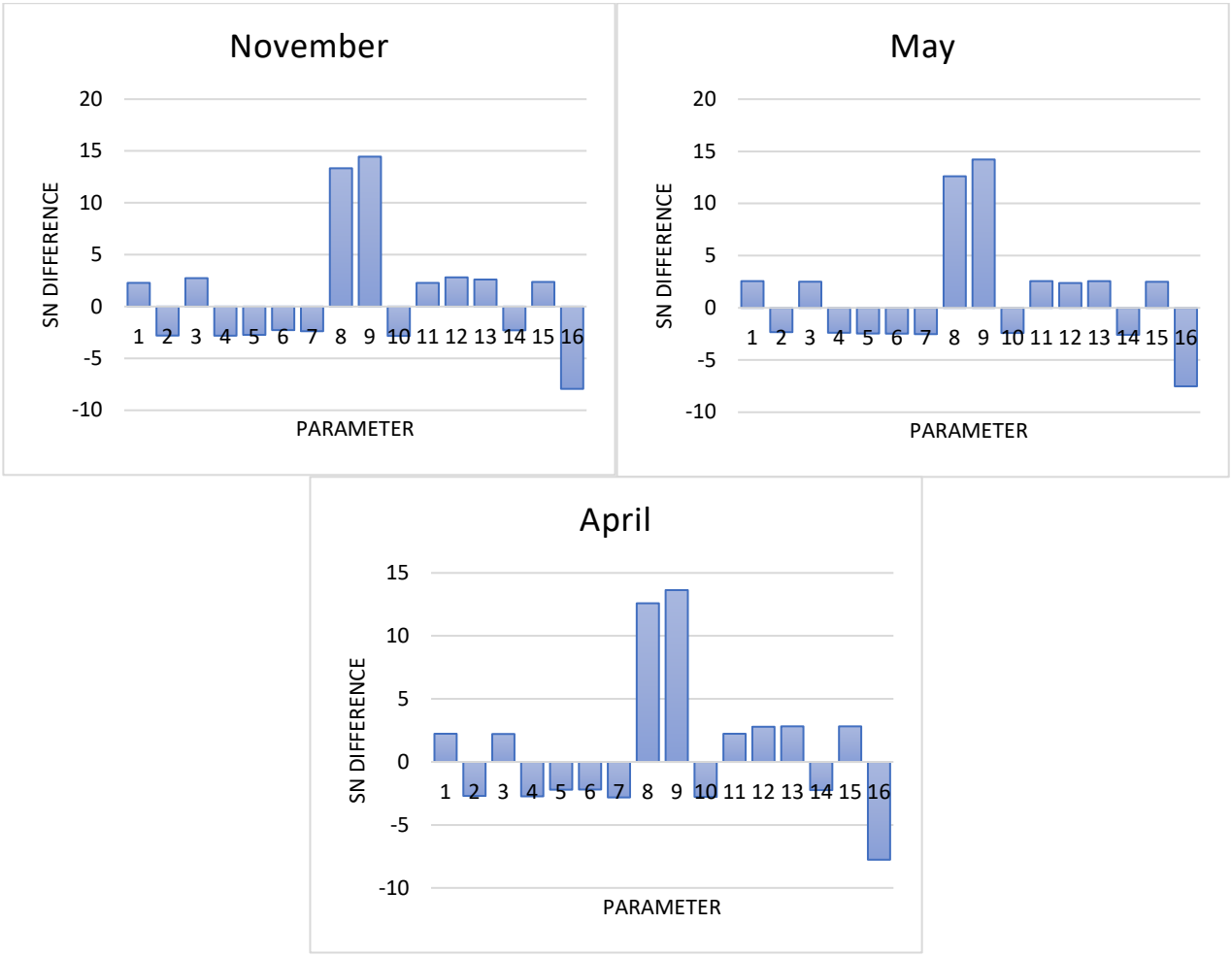



**Figure 5.** Degree of Contribution

## CONCLUSION

Based on the findings of this study, it has been determined that the MTS is capable to effectively distinguish between the unit space and signal data within the context of rainfall data. Based on the computation, the MD value for unit space is determined to be 1 over the course of three months. Additionally, the average MD values for the signal data are found to be 12.1333, 9.1690, and 11.0716, respectively. In addition to its ability to identify the significant parameters for rainfall trends associated with monsoon phenomena, the MTS has been utilized to analyse the data collected in the UMP Gambang. The number of parameters was decreased from 16 to 8 in order to optimise the system. These 8 parameters, as recommended by MTS, are considered significant for future prediction and classification tasks.

## ACKNOWLEDGEMENT.

The authors would like to thank the Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2022/TK10/UMP/03/7 (University reference RDU220104).

## REFERENCES

[1] W. H. Craddock, D. W. Burbank, B. Bookhagen, and E. J. Gabet, "Bedrock channel geometry along an orographic rainfall gradient in the Upper Marsyandi River Valley in central Nepal," *Journal of Geophysical Research*, vol. 112, no. F3, 2007, doi: 10.1029/2006jf000589.

[2] M. Niyongendako, A. E. Lawin, C. Manirakiza, and B. Lamboni, "Trend and variability analysis of rainfall and extreme temperatures in Burundi," *International Journal of Environment and Climate Change*, pp. 36–51, 2020, doi: 10.9734/ijecc/2020/v10i630203.

[3] A. N. Ishak, N. H. T. Ahmad, and M. S. J. Singh, "The diurnal variation of rain intensity in Malaysia for monsoon region using TRMM Satelit Data," *Jurnal Kejuruteraan*, vol. 33, no. 3, pp. 719–731, 2021, doi: 10.17576/jkukm-2021-33(3)-30.

[4] F. Tangang, R. Farzanmanesh, A. Mirzaei, E. S. Salimun, A. F. Jamaluddin, and L. Juneng, "Characteristics of precipitation extremes in Malaysia associated with El Niño and La Niña events," *International Journal of Climatology*, vol. 37, pp. 696–716, 2017, doi: 10.1002/joc.5032.

[5] M. Niyongendako, A. E. Lawin, C. Manirakiza, and B. Lamboni, "Trend and variability analysis of rainfall and extreme

temperatures in Burundi," *International Journal of Environment and Climate Change*, pp. 36–51, 2020, doi: 10.9734/ijecc/2020/v10i630203.

[6] Z. M. Marlan, F. Ramlie, K. R. Jamaludin, and N. Harudin, "Enhanced Taguchi's T-method Using Angle Modulated Bat Algorithm For Prediction," *Bulletin EEI*, vol. 5, no. 11, pp. 2828-2835, 2022, doi: 10.11591/eei.v11i5.4350.

[7] C. G. Mota-Gutiérrez, E. O. Reséndiz-Flores, and Y. I. Reyes-Carlos, "Mahalanobis-Taguchi System: State of the art," *International Journal of Quality & Reliability Management*, vol. 35, no. 3, pp. 596–613, 2018, doi: 10.1108/ijqrm-10-2016-0174.

[8] B. John, and R. S. Kadadevarmath, "A methodology for quantitatively managing the bug fixing process using Mahalanobis Taguchi System," *Management Science Letters*, pp. 1081–1090, 2015, doi: 10.5267/j.msl.2015.10.006.

[9] M. El-Banna, "Modified mahalanobis taguchi system for imbalance data classification," *Computational Intelligence and Neuroscience*, pp. 1–15, 2017, doi: 10.1155/2017/5874896.

[10] N. Harudin, F. Ramlie, W. Z. W. Muhamad, M. N. Muhtazaruddin, K. R. Jamaludin, M. Y. Abu, and Z. M. Marlan, "Binary bitwise artificial bee colony as feature selection optimization approach within Taguchi's T-method," *Mathematical Problems in Engineering*, *2021*, pp. 1–10, 2021, doi: 10.1155/2021/5592132.