SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Interpolating Rainfall Data Trends for Pahang Area using Polynomial Regression

Ismail Md Zani[1], Suzarina Ahmed Sukri[1,*], Yeak Su Hoe[1], Abdulnasir Isah[1]

[1] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

ARTICLE INFO

Article history:
Received 13 March 2026
Received in revised form 30 May 2026
Accepted 2 June 2026
Available online 23 June 2026

Keywords:

Rainfall trend; polynomial regression; rainfall interpolation; Pahang; RMSE

ABSTRACT

Rainfall data are important to support water resource planning. However, rainfall observations often contain missing values and irregular patterns due to limitations in the system. These limitations can affect the accuracy of rainfall trend analysis. This paper aims to analyse rainfall data trends in the Pahang area using polynomial regression. Daily rainfall data for Bentong and Kuantan were obtained from the NASA POWER Data Access Viewer from 2021 to 2024. Polynomial models from degree one to degree five were compared using root mean square error, and the third-degree polynomial was selected because it provided a suitable balance between fitting accuracy and numerical stability. The polynomial trend of 2024 was compared with the average trend from 2021 to 2023. The results show that rainfall patterns in both stations were non-linear, with rainfall generally decreasing during the early part of the year and increasing towards the later period. The comparison produced root mean square error values of 1.3879 mm for Bentong and 1.9855 mm for Kuantan. The lower value for Bentong indicates that it followed a trend very similar to its historical average, while Kuantan had slightly higher variation. Overall, the findings suggest that polynomial regression is a suitable method for producing a smooth representation of rainfall trends in the Pahang area.

1. Introduction

Rainfall information is needed in agriculture, flood monitoring and environmental assessment. However, rainfall observations are not always complete because missing values may occur due to system limitations, equipment problems and irregular records. Ismail *et al.,* [1] stated that incomplete rainfall records may affect the reliability of the analysis. This becomes more important in areas such as Pahang, where rainfall behaviour is influenced by monsoon activity and geographical variation. Mayowa *et al*., [2] reported that the east coast of Peninsular Malaysia often experiences heavy rainfall, especially during the northeast monsoon season.

Interpolation methods have been widely used to estimate rainfall values and improve rainfall representation in locations with limited observations. Deterministic methods such as inverse

* *Corresponding author.*
E-mail address: suzarina@utm.my

https://doi.org/10.37934/araset.60.5.521529

distance weighting (IDW) are often used because the procedure is simple, while geostatistical methods such as ordinary kriging can provide accurate rainfall estimates but require semi variogram modelling and more computational steps [3,4]. Regression-based methods have also been used to estimate rainfall by relating precipitation to physical predictors such as topography and wind speed [5]. In Peninsular Malaysia, Chin *et al.*, [6] reported that local polynomial interpolation performed better than IDW for annual rainfall estimation. FirdausHum and Talib [7] compared several spatial interpolation methods for monthly precipitation in Selangor, while Fung *et al.*, [8] evaluated spatial interpolation and spatiotemporal rainfall modelling in Peninsular Malaysia. These studies show that the performance of rainfall estimation methods may differ according to rainfall dataset, spatial setting and model formulation.

Besides estimating rainfall amount, rainfall trend analysis is also important because long-term rainfall behaviour may provide useful information for climate monitoring and hydrological planning. Mayowa *et al*., [2] showed that rainfall trends on the east coast of Peninsular Malaysia were closely related to seasonal monsoon patterns. Rainfall trends are commonly examined using the Mann–Kendall test together with Sen's slope estimator [2], while modified versions and related procedures have been proposed to address autocorrelation in Mann–Kendall test [9,10]. In addition, fast Fourier transform can be used to identify dominant seasonal cycles in rainfall data [11]. Comparative studies conducted in other regions also showed that interpolation performance may vary according to data distribution, terrain condition and model formulation [12–14]. Liu *et al.,* [15] further explained that autocorrelation may influence the reliability of hydrological trend testing if it is not properly considered. Nevertheless, previous studies have commonly focused on rainfall quantity interpolation or temporal rainfall trend analysis, rather than combining both perspectives in a single representation.

Although previous studies have discussed rainfall interpolation and rainfall trend analysis using several methods, studies on the use of polynomial regression for representing temporal rainfall trends in the Pahang area are still limited. Most existing studies only focused on estimating missing values, mapping spatial rainfall distribution and identifying long-term trend significance using statistical tests. However, less emphasis has been placed on producing a smooth polynomial representation that can describe yearly rainfall behaviour. This study addresses this gap by analysing daily rainfall trends in Bentong and Kuantan using polynomial regression. Therefore, this paper aims to compare polynomial degrees based on fitting accuracy, to select a suitable polynomial degree and to evaluate the suitability of polynomial regression for representing rainfall trends in the Pahang area. The significance of this study lies in providing an interpretable mathematical approach for describing rainfall behaviour over time.

2. Polynomial Trend Modelling of Daily Rainfall

2.1 Mathematical Modelling and Formulation

In this study, rainfall data obtained from the NASA POWER Data Access Viewer were modeled as a function of time using the polynomial approximation in order to capture the rainfall trend. The observed rainfall time series is denoted as

$$\{(t_i, y_i)\}_{i=1}^{N}, \tag{1}$$

where $t_i = \{1,2,3,\dots,365\}$ indicates the indexing of days in one year while $y_i \in \mathbb{R}$ denotes the rainfall amount in millimetres on each day, t_i. To represent the smooth temporal variation of the rainfall, a polynomial function, $p(t_i)$ with degree, d is employed

$$p(t_i) = \sum_{k}^{d} C_k t_i^k \,, \tag{2}$$

where C_k represents the unknown coefficients to be estimated. In this study, a polynomial function with degree three is used due to its ability to capture the nonlinear variation of the rainfall data while maintaining the numerical stability. Hence, the polynomial model can be explicitly written as

$$p(t_i) = C_0 + C_1 t_i + C_2 t_i^2 + C_3 t_i^3. \tag{3}$$

This polynomial function represents the long-term trend component of the rainfall data. Next section provides the detailed explanation on the least squares estimation to compute the coefficients of the polynomial function, C_k for $k = 0, 1, 2$ and 3.

2.2 Polynomial Coefficients Estimation via Least Squares Method

The coefficients of the polynomial function were estimated by minimising the sum of squared errors between the observed rainfall data and the polynomial approximation. The sum of squared errors is defined as

$$S = \min_{C_0, C_1, C_2, C_3} \sum_{i=1}^{N} \left(y_i - p(t_i)\right)^2 . \tag{4}$$

By substituting the third-degree polynomial model into Eq. (4), the objective function becomes

$$S = \min_{C_0, C_1, C_2, C_3} \sum_{i=1}^{N} \left(y_i - (C_0 + C_1 t_i + C_2 t_i^2 + C_3 t_i^3)\right)^2 . \tag{5}$$

The polynomial regression model can be written in the matrix form as

$$\boldsymbol{y} = \boldsymbol{V}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{6}$$

where $\boldsymbol{\epsilon}$ represents the residual errors while $\boldsymbol{y}$, $\boldsymbol{V}$ and $\boldsymbol{\beta}$ can be written as

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \end{bmatrix},$$

$$\boldsymbol{V} = \begin{bmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_N & t_N^2 & t_N^3 \end{bmatrix},$$

where N is the number of days in one year. In order to obtain $\boldsymbol{\beta}$, the Eq. (6) can be written as

$$\boldsymbol{\beta} = (\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V})^{-1}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{y}. \tag{7}$$

Eq. (7) can equivalently be solved by using the following set of equations where this yields to the optimal coefficient vector of $\boldsymbol{\beta}$ that minimises the squared error, S. To ensure a unique solution, the matrix, $\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V}$ is symmetric and invertible under the assumption of linearly independent polynomial basis function. In practice, the computation involves evaluating the Gram matrix, $\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V}$, projecting the data onto the polynomial basis via $\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{y}$ and solving the resulting linear system using numerically stable techniques. The resulting coefficient vector $\boldsymbol{\beta}$ is then used to construct the explicit polynomial function with degree three in the following section as in Table 1 and Table 2.

Table 1
Polynomial for Bentong and Kuantan (2021-2023)

Station	Year	Polynomial, $p(t)$
Bentong	2021	$2.1 \times 10^{-6}t^3 - 8.0 \times 10^{-4}t^2 + 5.8 \times 10^{-2}t + 9.4$
	2022	$4.3 \times 10^{-7}t^3 + 2.6 \times 10^{-5}t^2 - 4.8 \times 10^{-2}t + 11.8$
	2023	$7.2 \times 10^{-7}t^3 - 7.9 \times 10^{-5}t^2 - 4.3 \times 10^{-2}t + 11.5$
Station	Year	Polynomial, $p(t)$
Kuantan	2021	$1.6 \times 10^{-6}t^3 - 3.7 \times 10^{-4}t^2 - 3.9 \times 10^{-2}t + 13.9$
	2022	$1.0 \times 10^{-6}t^3 - 1.4 \times 10^{-4}t^2 - 4.8 \times 10^{-2}t + 13.6$
	2023	$3.7 \times 10^{-7}t^3 + 3.1 \times 10^{-4}t^2 - 1.4 \times 10^{-1}t + 16.6$

Table 2
Average polynomial for Bentong and Kuantan (2021-2023)

Station	Polynomial, $p_{avg}(t)$
Bentong	$1.1 \times 10^{-6}t^3 - 2.8 \times 10^{-4}t^2 - 1.1 \times 10^{-2}t + 10.9$
Kuantan	$9.8 \times 10^{-7}t^3 - 6.8 \times 10^{-5}t^2 - 7.4 \times 10^{-2}t + 14.7$

3. Results

3.1 Selection of the Optimum Polynomial Degree

To determine the most suitable polynomial degree for rainfall trend representation, polynomial models from degree one to degree five were compared using root mean square (RMSE). The comparison was carried out for Bentong and Kuantan using rainfall data from 2021 to 2024. Table 3 shows the RMSE values obtained for each polynomial degree. Generally, the RMSE decreased as the polynomial degree increased for all datasets.

Table 3
Comparison of RMSE for polynomial degree 1 to degree 5

Station	Year	Degree 1	Degree 2	Degree 3	Degree 4	Degree 5
Bentong	2021	15.776507	15.441505	15.326111	15.186483	15.080166
	2022	11.076705	10.773698	10.766603	10.686178	10.552423
	2023	11.129576	10.684201	10.663898	10.660358	10.607292
	2024	11.069154	10.822267	10.822198	10.744629	10.675408
Kuantan	2021	19.485138	18.874803	18.820214	18.499584	18.344994
	2022	16.342361	15.820523	15.793206	15.720681	15.653364
	2023	18.088335	17.353373	17.349983	17.315576	17.287957
	2024	14.967160	14.405032	14.402206	14.229378	14.154281

Although degree five produced the lowest RMSE for all datasets, the improvement after degree three was relatively small. However, higher-degree polynomials may show more complex curve behaviour and reduce numerical stability, as shown in Figure 1 and Figure 2. Since the third-degree polynomial provided a good balance between fitting accuracy and model simplicity, it was selected as the optimum polynomial degree in this study. Therefore, the third-degree polynomial was then used in the subsequent analysis of rainfall trend for Bentong and Kuantan.

In both figures, the first-degree polynomial gives a straight-line trend. The second-degree polynomial shows only one main curvature, while the third-degree polynomial is able to represent the decreasing and increasing rainfall pattern more clearly. The fourth-degree and fifth-degree polynomials show additional bending towards the end of the year, which indicates more complex curve behaviour. Therefore, the graphical comparison supports the RMSE analysis and further justifies the selection of the third-degree polynomial as the optimum degree.

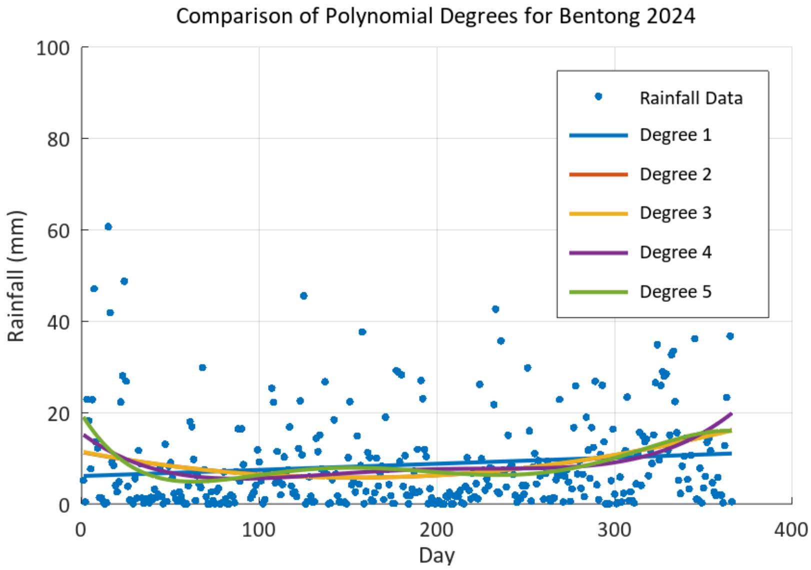

Fig. 1. Comparison of polynomial degrees for Bentong 2024

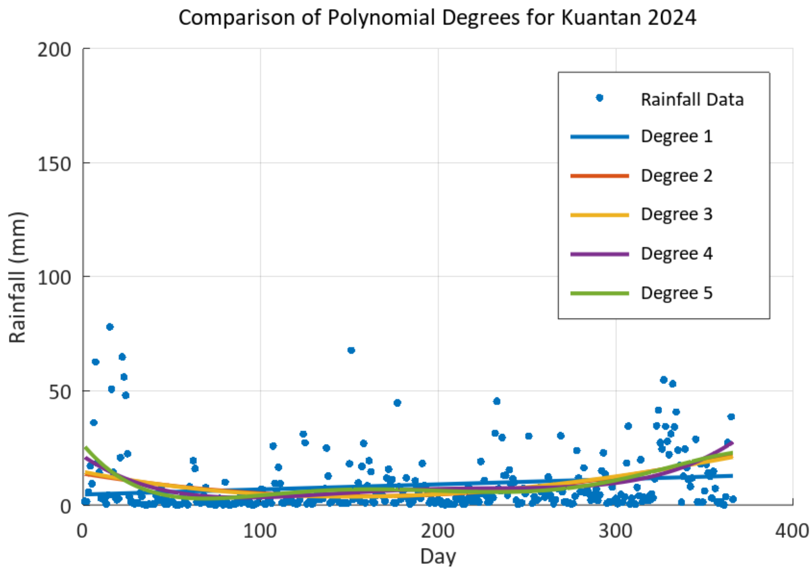

Fig. 2. Comparison of polynomial degrees for Kuantan 2024

3.2 Average Polynomial Trend for Bentong and Kuantan (2021-2023)

The average polynomial trend for Bentong from 2021 to 2023 is shown in Figure 3. The rainfall trend shows a non-linear pattern throughout the year. A decreasing trend can be seen during the early part of the year and an increase trend towards the end of the year. This means that rainfall behaviour in Bentong was not constant over the year.

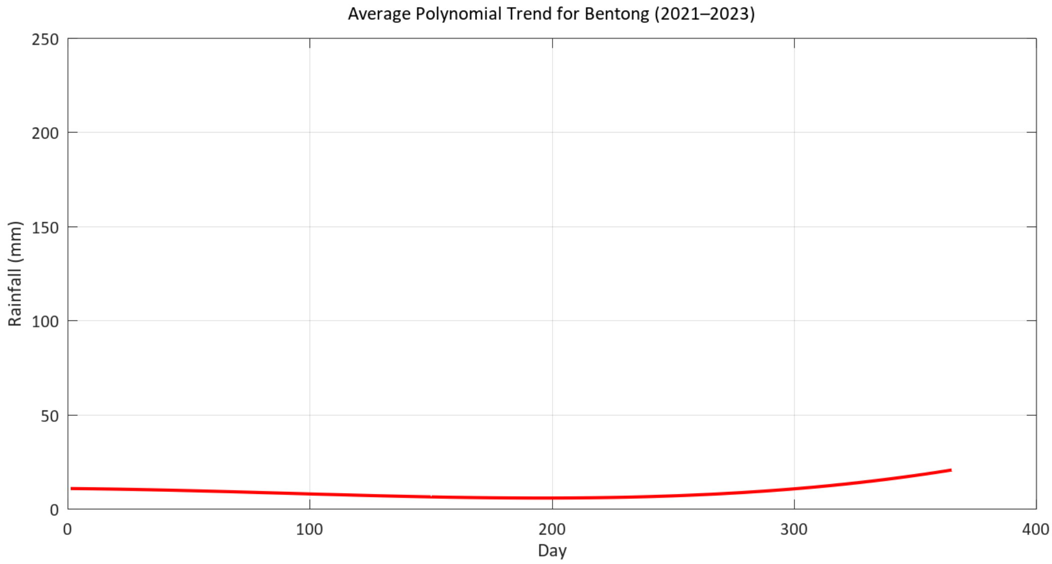

Fig. 3. Average polynomial trend for Bentong (2021-2023)

Similar to Bentong, the rainfall trend for Kuantan also shows a non-linear pattern. The average polynomial trend for Kuantan from 2021 to 2023 is shown in Figure 4, where rainfall decreases during the earlier part of the year and increases again towards the later period. However, the increasing trend at the end of the year was more obvious in Kuantan. This shows that Kuantan had stronger variation in rainfall behaviour due to its coastal location and exposure to monsoon influence. Hence, the polynomial trend curve provides a clear representation of the overall rainfall pattern for the three-year period.

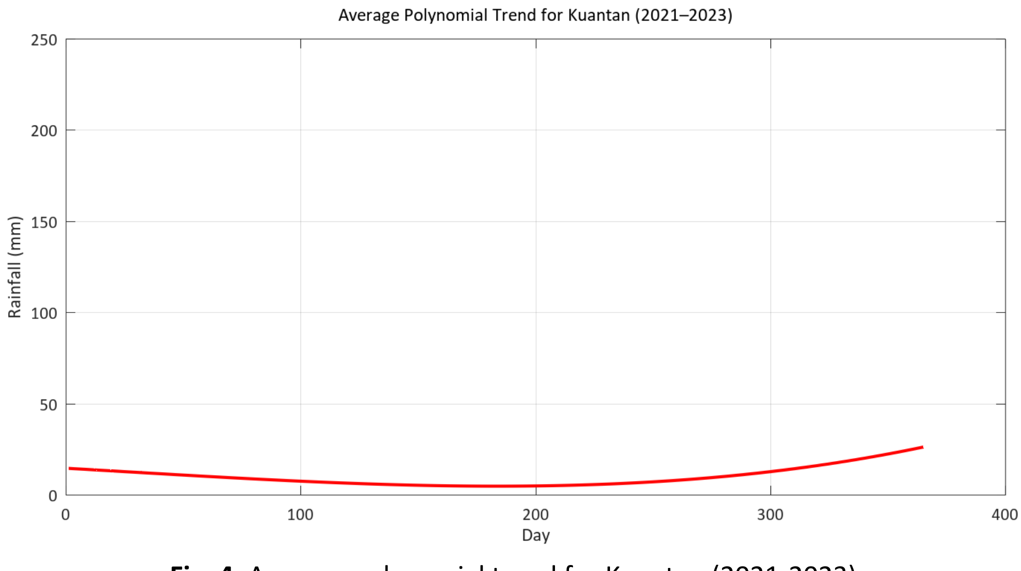

Fig. 4. Average polynomial trend for Kuantan (2021-2023)

3.3 Comparison between 2024 Trend and Average Trend (2021-2023)

The comparison between the polynomial trend for 2024 and the average trend from 2021 to 2023 for Bentong is shown in Figure 5. Both curves follow a similar shape with a decreasing pattern during the early period and increasing trend towards the end of the year. The polynomial trend for 2024 is slightly lower than average trend from 2021 to 2023. Nevertheless, the difference between both curves is small.

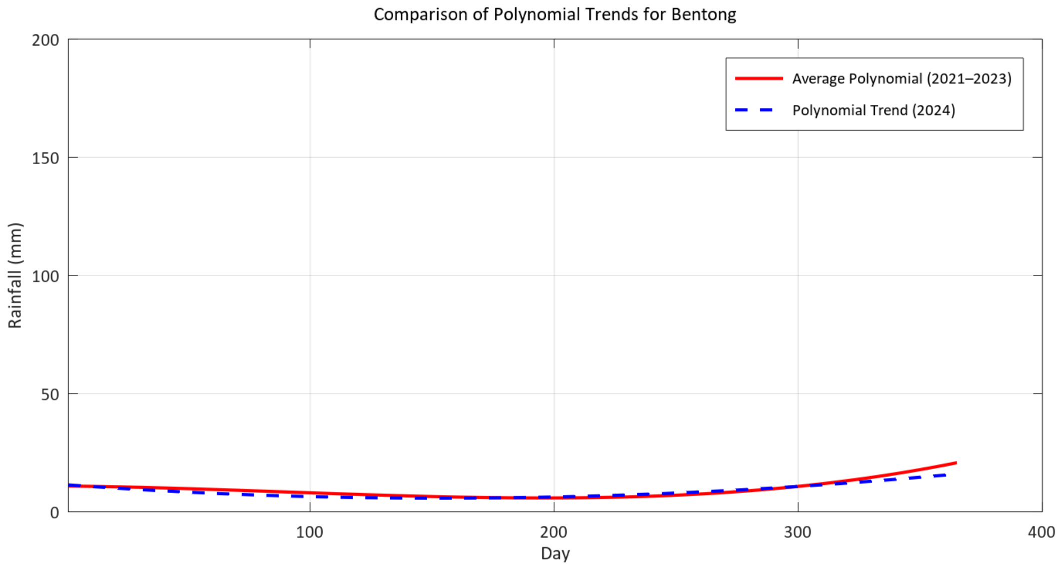

Fig. 5. Comparison of polynomial trend of 2024 and average trend for Bentong

The comparison between the polynomial trend for 2024 and the average trend from 2021 to 2023 for Kuantan is presented in Figure 6. It can be observed that both curves show a non-linear pattern as well as the decreasing trend in the earlier part of the year and increasing at the end of the year. Similarly, the polynomial trend for 2024 is lower than the average trend for some period of the year. In contrast, Kuantan showed a slightly larger deviation from its historical average trend compared to Bentong.

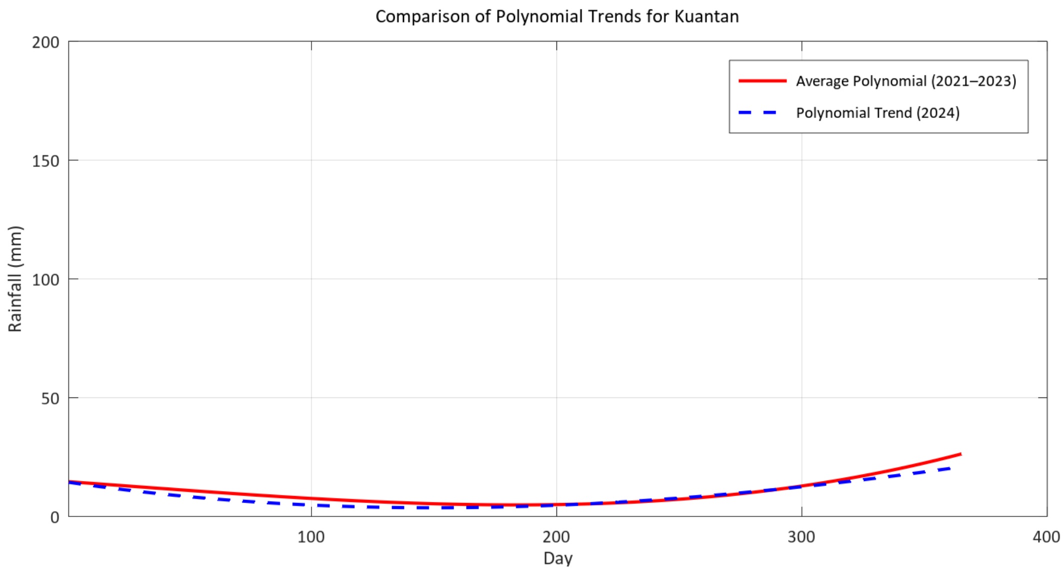

Fig. 6. Comparison of polynomial trend of 2024 and average trend for Kuantan

To further evaluate the difference between the polynomial trend for 2024 and the average trend from 2021 to 2023, RMSE was calculated for both stations as shown in Table 4. Bentong recorded RMSE value of 1.3879 mm, while Kuantan recorded 1.9855 mm. The lower RMSE value for Bentong indicates that the rainfall behaviour in 2024 was closer to its historical average trend. In comparison, the higher RMSE value for Kuantan suggests more variation from the average rainfall pattern.

Table 4
RMSE between 2024 polynomial trend and average trend (2021-2023)

Station	RMSE, mm
Bentong	1.3879
Kuantan	1.9855

4. Conclusions

To determine the most suitable polynomial degree, models from degree one to degree five were compared using RMSE. Even though the fifth-degree polynomial gave the lowest RMSE for all datasets, the difference after degree three was relatively small. Thus, the third-degree polynomial was selected as the optimum degree since it provided a better balance between fitting accuracy and model simplicity. Therefore, this paper analysed rainfall data trends in Bentong and Kuantan, Pahang as the main stations by using third-degree polynomial regression. The results showed that the rainfall trends in both locations presented a non-linear pattern, with rainfall generally decreasing during the earlier part of the year and increasing towards later period. This shows that rainfall behaviour in the study area was not constant and cannot be represented by a simple linear model. The comparison between the 2024 polynomial trend and the average trend from 2021 to 2023 showed that both stations remained consistent with their historical rainfall pattern. However, Kuantan showed a slightly larger deviation compared to Bentong. The RMSE results indicate that Bentong recorded 1.3879 mm, while Kuantan recorded 1.9855 mm. This suggests that Bentong had a more stable rainfall trend in 2024, while Kuantan had slightly higher variation. Overall, polynomial regression was found to be suitable for representing rainfall trends in Pahang. In conclusion, the method was able to produce smooth trend representation and gave a useful interpretation of rainfall behaviour over time.

Acknowledgement
This research was funded by Kurita Water and Environment Foundation (KWEF) 2025 (25Pmy125) under International Grant R.J130000.7354.1U125.

References

[1] Ismail, Wan Norliyana Wan, Wan Zawiah Wan Zin, and Wan Ibrahim. "Estimation of rainfall and stream flow missing data for Terengganu, Malaysia by using interpolation technique methods." *Malaysian Journal of fundamental and applied sciences* 13, no. 3 (2017): 213-217. https://doi.org/10.11113/mjfas.v13n3.578

[2] Mayowa, Olaniya Olusegun, Sahar Hadi Pour, Shamsuddin Shahid, Morteza Mohsenipour, Sobri Bin Harun, Arien Heryansyah, and Tarmizi Ismail. "Trends in rainfall and rainfall-related extremes in the east coast of peninsular Malaysia." *Journal of Earth System Science* 124, no. 8 (2015): 1609-1622.https://doi.org/10.1007/s12040-015-0639-9

[3] Mohamed Salleh, Mohd Khaidir, Noor Fadhilah Ahmad Radib, and Nor Azrita Mohd Amin. "Spatial interpolation for missing rainfall data in Northern Region of Peninsular Malaysia." In *Journal of Physics: Conference Series*, vol. 1863, no. 1, p. 012049. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1863/1/012049

[4] Jamaludin, Suhaila, and Hanisah Suhaimi. "Spatial interpolation on rainfall data over peninsular Malaysia using ordinary kriging." *Jurnal Teknologi (Sciences & Engineering)* 63, no. 2 (2013).https://doi.org/10.11113/jt.v63.1912

[5] Anees, Mohd Talha, Khiruddin Abdullah, M. N. M. Nawawi, Nik Norulaini Nik Ab Rahman, Abd Rahni Mt Piah, M. I. Syakir, Mohammad Muqtada Ali Khan, and Abdul Kadir Mohd Omar. "Spatial estimation of average daily precipitation using multiple linear regression by using topographic and wind speed variables in tropical climate." *Journal of Environmental Engineering and Landscape Management* 26, no. 4 (2018): 299-316. https://doi.org/10.3846/jeelm.2018.6337

[6] Chin, Ren Jie, Sai Hin Lai, Wing Son Loh, Lloyd Ling, and Eugene Zhen Xiang Soo. "Assessment of inverse distance weighting and local polynomial interpolation for annual rainfall: A case study in Peninsular Malaysia." *Engineering Proceedings* 38, no. 1 (2023): 61. https://doi.org/10.3390/engproc2023038061

[7] FirdausHum, Nurul Nadiah Mohd, and S. A. Talib. "Spatial interpolation of monthly precipitation in selangor, Malaysia-Comparison and evaluation of methods." *Journal of Applied and Physical Sciences* 2, no. 1 (2016): 1-9. https://doi.org/10.20474/japs-2.1.1

[8] Fung, Kit Fai, Kim Soon Chew, Yuk Feng Huang, Ali Najah Ahmed, Fang Yenn Teo, Jing Lin Ng, and Ahmed Elshafie. "Evaluation of spatial interpolation methods and spatiotemporal modeling of rainfall distribution in Peninsular Malaysia." *Ain Shams Engineering Journal* 13, no. 2 (2022): 101571. https://doi.org/10.1016/j.asej.2021.09.001

[9] Hamed, Khaled H., and A. Ramachandra Rao. "A modified Mann-Kendall trend test for autocorrelated data." *Journal of hydrology* 204, no. 1-4 (1998): 182-196. https://doi.org/10.1016/S0022-1694(97)00125-X

[10] Yue, Sheng, and Chun Yuan Wang. "Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test." *Water resources research* 38, no. 6 (2002): 4-1. https://doi.org/10.1029/2001WR000861

[11] Agustina, Atika, Hamdi Akhsan, Melly Ariska, Suhadi Suhadi, and Nely Andriani. "ANALYSIS OF RAINFALL PATTERNS IN THE NUSA TENGGARA REGION USING THE FAST FOURIER TRANSFORM (FFT) METHOD AND ITS RELATIONSHIP WITH EL-NIÃ ‘O AND IOD." *Indonesian Physical Review* 8, no. 1 (2025): 268-280.https://doi.org/10.29303/ipr.v8i1.397

[12] Hwang, Yeonsang, Martyn Clark, Balaji Rajagopalan, and George Leavesley. "Spatial interpolation schemes of daily precipitation for hydrologic modeling." *Stochastic environmental research and risk assessment* 26, no. 2 (2012): 295-320. https://doi.org/10.1007/s00477-011-0509-1

[13] Wang, Shuo, G. H. Huang, Q. G. Lin, Z. Li, H. Zhang, and Y. R. Fan. "Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada." *International Journal of Climatology* 34, no. 14 (2014).https://doi.org/10.1002/joc.3941

[14] Di Piazza, Annalisa, F. Lo Conti, Leonardo V. Noto, Francesco Viola, and G. La Loggia. "Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy." *International Journal of Applied Earth Observation and Geoinformation* 13, no. 3 (2011): 396-408. https://doi.org/10.1016/j.jag.2011.01.005

[15] Liu, Saiyan, Yangyang Xie, Hongyuan Fang, Huihua Du, and Pengcheng Xu. "Trend test for hydrological and climatic time series considering the interaction of trend and autocorrelations." *Water* 14, no. 19 (2022): 3006. https://doi.org/10.3390/w14193006