INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 7 (2025) 34-40
https://publisher.uthm.edu.my/ojs/index.php/ijie

Spatial Rainfall Interpolation and Evaluation for Seasonal Precipitation in Peninsular Malaysia

Eugene Zhen Xiang Soo[1]*, Ren Jie Chin[2], Janice Lynn Ayog[1], Nurmin Bolong[1], Foo Wei Lee[2]

[1] *Faculty of Engineering, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, MALAYSIA*
[2] *Department of Civil Engineering, Lee Kong Chian Faculty of Engineering, Universiti Tunku Abdul Rahman, 43000 Kajang, Selangor, MALAYSIA*

*Corresponding Author: esoo@ums.edu.my
DOI: https://doi.org/10.30880/ijie.2025.17.07.004

Article Info

Received: 6 January 2025
Accepted: 30 July 2025
Available online: 23 December 2025

Keywords

Rainfall interpolation, local polynomial interpolation (LPI), inverse distance weighting (IDW), seasonal precipitation

Abstract

Rainfall, a major driving force in hydrology and water resources planning, poses challenges due to potential malfunctions in conventional rain gauges, necessitating effective methods for filling the data gap. This study assesses the efficacy of two spatial rainfall interpolation techniques, i.e., Local Polynomial Interpolation (LPI) and Inverse Distance Weighting (IDW), for seasonal rainfall estimation in Peninsular Malaysia. Interpolated values from both methods are compared to ground observations, and their performance is evaluated through cross-validation using statistical measures such as RMSE, MAE, and R^2. Additionally, Geographically Weighted Regression (GWR) is employed to analyse the relationship between interpolated rainfall and ground elevation. The findings reveal that LPI outperforms IDW, demonstrating higher R^2 values and lower MAE and RMSE, highlighting its superior accuracy in rainfall estimation and underscoring the importance of method selection in handling missing rainfall data.

1. Introduction

Peninsular Malaysia (PM), located in Southeast Asia, experiences a tropical climate marked by significant seasonal precipitation variations due to the monsoon system's influence. The region's climate significantly affects water resource management, agriculture, and disaster preparedness. In this context, accurately depicting spatial rainfall distribution becomes crucial for informed decision-making and effective planning [1]-[3].

Rainfall interpolation enables meteorologists to estimate the precipitation values at sites that lack direct measurement. This process is fundamental for generating comprehensive and high-resolution rainfall maps. Such maps are indispensable for applications in agriculture, where farmers rely on precise rainfall data to optimise irrigation schedules and crop management practices. Similarly, spatial rainfall information is essential to assess the water availability and reservoir operations, ensuring a sustainable supply for agricultural and domestic needs. In disaster preparedness, accurate spatial rainfall is needed to anticipate and mitigate the impact of natural disasters. Flood risk assessments, for instance, heavily depend on understanding the spatial distribution of rainfall, particularly in regions prone to flash floods. Landslide susceptibility mapping is another critical application, as heavy and localised precipitation can trigger slope failures, posing risks to infrastructure and communities [4]-[7]. The geographical heterogeneity of PM, featuring diverse landscapes such as mountains, coastlines, and urban areas, amplifies the challenge of capturing the spatial variability of rainfall. This challenge becomes more pronounced during the monsoon seasons, characterised by convective and localised precipitation events. Accurate rainfall interpolation methods are indispensable for addressing these complexities and providing decision-makers with reliable information tailored to the region's specific topographical and climatic characteristics [8], [9].

Previous studies in PM have utilised various interpolation approaches, such as kriging, inverse distance weighting (IDW), and geostatistical approaches [10]-[12]. For instance, FirdausHum & Talib [11] compared multiple interpolation techniques such as Ordinary Kriging, IDW, and Local Polynomial Interpolation (LPI) in Selangor, concluding that kriging methods generally provide superior accuracy for monthly precipitation estimation based on cross-validation statistics. Fung et al. [12] extended this by evaluating spatial interpolation methods alongside spatiotemporal modelling approaches like Geographically Weighted Regression (GWR) and Multi-scale GWR (MGWR) to capture rainfall distribution changes across Peninsular Malaysia, emphasising the importance of accounting for both spatial and temporal variability due to climate influences. Meanwhile, Fagandini et al. [13] focused on gap-filling techniques for missing daily rainfall data, comparing interpolation methods to improve data completeness, which is critical for reliable hydrological modelling.

Despite these advances, the pronounced seasonal variability in precipitation across Peninsular Malaysia—characterized by distinct monsoon and inter-monsoon periods—poses challenges for spatial interpolation, as rainfall intensity and distribution patterns vary significantly by season and region. This necessitates a nuanced evaluation of interpolation methods to determine their effectiveness in capturing these seasonal dynamics. While kriging and IDW have been widely applied, Local Polynomial Interpolation (LPI) offers an alternative that can adapt locally to spatial heterogeneity, potentially improving estimates where rainfall gradients are complex.

This study seeks to assess the efficacy of two spatial interpolation methods, i.e. Local Polynomial Interpolation (LPI) and IDW, in estimating seasonal-scale precipitation in PM. Through the integration of Geographic Information System (GIS) tools and modelling techniques, the research endeavours to overcome challenges associated with data gaps and spatial variations. The primary goal is to enhance comprehension and application of rainfall data in the context of environmental management and decision-making processes. The anticipated outcome is an improved capacity to leverage rainfall information for more informed and practical decision support.

2. Methodology

2.1 Data Acquisition

This study focuses on 550 rain gauge stations across PM, managed by the Department of Irrigation and Drainage (DID) Malaysia, as shown in Fig. 1. 10-year rainfall records, from 2011 to 2020, were collected. Ensuring the reliability and consistency of the rainfall data before analysis necessitated a rigorous quality check process.

Rain gauge stations without missing data were considered reliable and included directly in the analysis due to their ability to provide comprehensive and continuous information on local rainfall patterns. However, stations exhibiting missing data required additional preprocessing to ensure data quality and analytical accuracy. Following the approach of Burhanuddin et al. [13], the rain gauge stations were classified into two groups based on the proportion of missing data: those with less than 10% missing values and those exceeding this threshold. According to Burhanuddin et al., datasets with less than 10% missing data are generally regarded as good quality and suitable for imputation or further analysis without significantly compromising accuracy. Consequently, stations with missing data below this 10% threshold were retained after appropriate gap-filling procedures, while stations with more than 10% missing data were excluded from the study to avoid potential biases and inaccuracies in rainfall estimation. This exclusion criterion helps to maintain the robustness of the spatial interpolation and overall reliability of the precipitation analysis.

2.2 Spatial Interpolation

This study evaluates two interpolation approaches, specifically LPI and IDW. These two approaches were chosen for their relevance to study objectives and capacity to generate reliable rainfall estimates across the study area. In the case of IDW (Eq. (1)), the correlation between neighbouring points is directly tied to the distance that separates them. This relationship is defined by a distance reverse function for each location, a pivotal factor in the interpolation process. Specifying the nearby radius and determining the appropriate power of the distance reverse function in IDW can be challenging. This depends on setting several sample sites (ideally ≥ 14) and achieving a proper dispersion degree at the local scale level. The accuracy of the inverse distance interpolator is significantly influenced by the power parameter, making the effectiveness of IDW contingent on carefully selecting parameters such as the nearby radius, power parameter, and distribution of sample sites in the study area.

$$Z_o = \left(\sum_{i=1}^{S} Z_i \cdot \frac{1}{d_i^{\,K}} \right) \Bigg/ \left(\sum_{i=1}^{S} \frac{1}{d_i^{\,K}} \right) \quad (1)$$

where Z_o is the predicted rainfall at the unsampled site; Z_i is the observed rainfall at data point i; d_i is the distance between the prediction and observed locations; s is the number of observed sampling points within the site, and K is the exponent that defines the rate of reduction of the weights as distance increases.

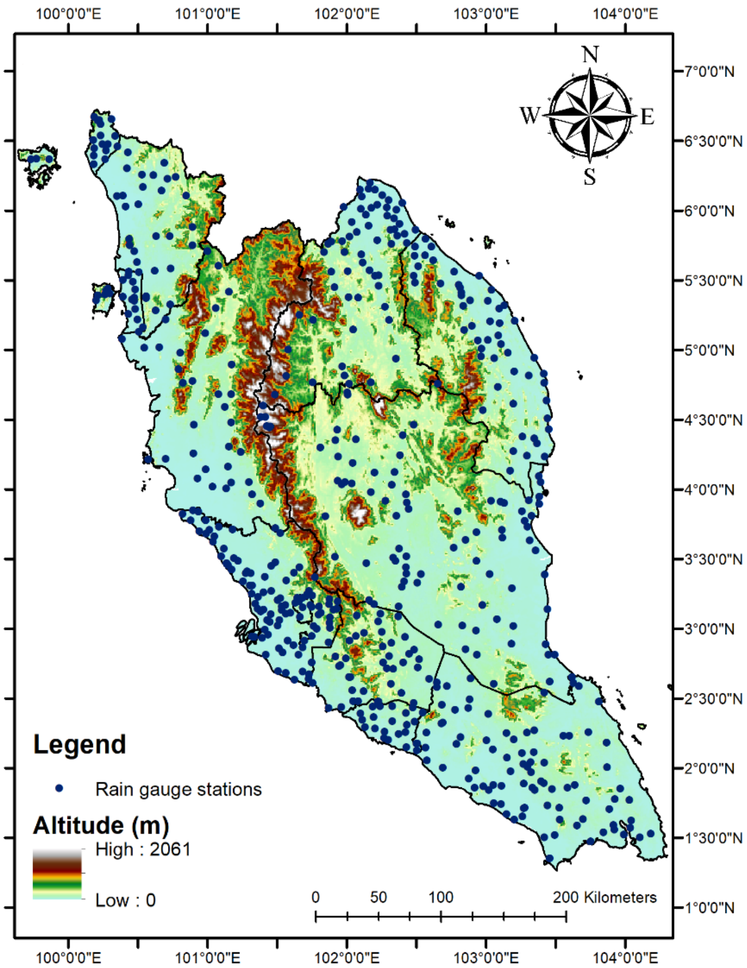

Fig. 1 *Selected rain gauge stations across PM*

IDW is an accurate interpolator, helping circumvent division by zero issues at sample points when $d_i = 0$. This method ensures that the maximum and minimum values of the interpolated surface are confined to the data points. Despite its efficacy in rapid interpolation, IDW exhibits limitations, including susceptibility to outliers and data clustering. Moreover, a notable drawback is the absence of implicit evaluation for estimation accuracy.

On the contrary, Local Polynomial Interpolation (LPI) (see Eq. (2)) proves to be a robust method for fitting polynomials within overlapping neighbourhoods. These neighbourhoods are selected by a search neighbourhood criterion that offers flexibility in choosing form, size, and organisation. LPI effectively captures short-range variations, enhancing its suitability for surfaces susceptible to local changes. Additionally, LPI can adjust the neighbourhood's width and introduce a power parameter through a slider. The power parameters affect the weight assigned to each sample point in the neighbourhood based on its distance from the interpolated point. LPI can furnish surfaces that account for more localised variations through parameter adjustments.

$$Z_o = \sum_{i=1}^{n} w_i . Z_i \tag{2}$$

where Z_o is the predicted value at the unsampled site; Z_i is the observed value at data point i, and w_i is the weight assigned to the data point based on its distance and other parameters determined by the local polynomial fitting.

2.3 Cross Validation (CV)

CV is primarily a model validation approach to evaluate the spatial interpolation model's performance. In spatial interpolation, CV involves dividing the dataset into subsets. The model is trained on a subset of the data and then tested on the remaining subset. The process is iterated multiple times, utilising different subsets for training and testing each iteration. CV assesses the overall predictive performance of the interpolation model but does not explicitly consider spatial relationships or variations. For the present study, three statistical metrics were adopted for the CV, including the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2).

2.4 Geographically Weighted Regression (GWR)

GWR examines spatially diverse relationships between different variables. This technique extends traditional regression by assigning different weights to data points based on spatial proximity. It recognises that relationships between variables can vary across space. GWR estimates a local regression model for each location, considering only nearby data points and assigning higher weights to closer ones. GWR explicitly accounts for spatial variability, allowing for the identification of spatial patterns and trends in the relationships between variables. In the present study, the GWR was applied to investigate the relationship between the interpolated rainfall and the altitude of PM, and the same statistical metrics mentioned in Section 2.3 were adopted.

3. Results

3.1 Seasonal Patterns and Comparative Analysis

Three seasonal periods were recognised in this study, namely Northeast (NEM), Southwest (SWM), and Inter-Monsoons (IM). A seasonal scale assessment was conducted. Fig. 2 shows the IDW and LPI interpolation results presented using a raster figure, while Table 1 shows the overview data for cross-validation (CV) results using IDW and LPI methods.

Based on the outputs, both interpolation methods exhibited varying performance levels across seasons. The NEM emerges as a season where both interpolation methods exhibit heightened accuracy. However, the LPI method stands out with a consistently lower RMSE (1.51 mm/day), a smaller MAE (1.00 mm/day), and a notably higher R^2 (0.78). These metrics collectively indicate the method's proficiency in capturing the intricate spatial variations of rainfall during this season. The superior performance of LPI in the NEM suggests its efficacy in handling complex topographical and climatic features, making it a favourable choice for this season.

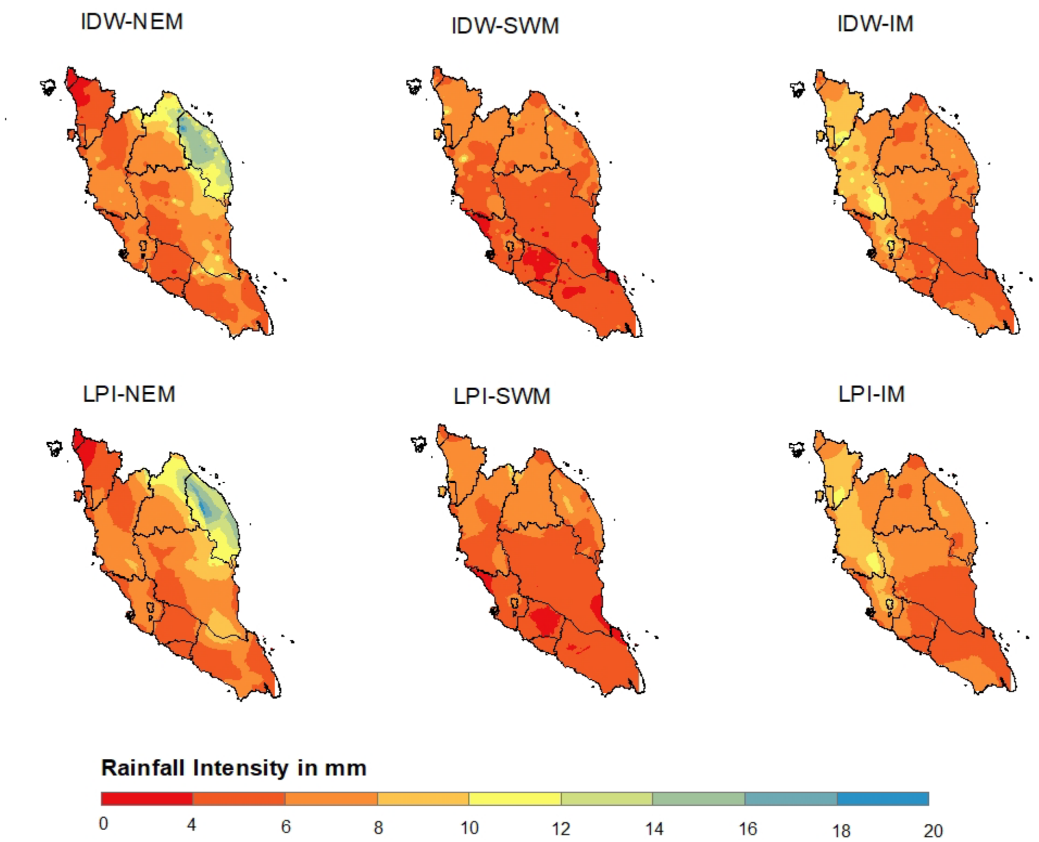

Fig. 2 *Seasonal rainfall maps. IDW represents rainfall map interpolated using Inverse Distance Weighted Method; LPI represents rainfall map interpolated using Local Polynomial Interpolation technique. NEM stands for northeast monsoon, SWM stands for southwest monsoon and IM stands for Inter-monsoon*

Table 1 *CV results*

Season	IDW			LPI		
	RMSE	MAE	R^2	RMSE	MAE	R^2
NEM	1.65	1.09	0.74	1.51	1.00	0.78
SWM	1.01	0.72	0.50	1.02	0.73	0.50
IM	1.18	0.86	0.53	1.15	0.85	0.54

During the SWM, the IDW method displays commendable accuracy with a lower RMSE (1.01 mm/day) and a similar MAE (0.72 mm/day) compared to LPI. However, despite exhibiting a slightly higher RMSE, the LPI method maintains an R^2 of 0.50, comparable to IDW. This suggests that both methods offer relatively comparable performance during the SWM, with IDW showing a slight edge in lower RMSE.

The IM season reveals a nuanced performance comparison between the two methods. LPI again demonstrates its strength with a lower RMSE (1.15 mm/day) and MAE (0.85 mm/day), coupled with a higher R^2 (0.54) compared to IDW. This suggests that the LPI method is better equipped to handle the transitional and less predictable conditions during the IM period.

The consistent superior performance of LPI during the NEM underscores the method's adaptability to conditions characterised by high variability. The comparable performance of IDW and LPI during the SWM indicates that both methods can adequately capture the spatial patterns of rainfall in this season. LPI emerges as the preferred choice in the IM season, emphasising its versatility in handling transitional periods with fluctuating climatic patterns.

3.2 Relationship Between Seasonal Rainfall and Elevation

The study also delves into the complex relationship dynamics between IDW and LPI-interpolated rainfall and elevation across PM using the Geographically Weighted Regression (GWR). Table 2 summarises the GWR analysis between seasonal interpolated rainfall and elevation in PM. The GWR results offer valuable insights into the predictive power of elevation for rainfall, unveiling spatial nuances that transcend a uniform relationship. IDW demonstrates moderate accuracy in the NEM season with an RMSE of 2.14 mm/day and an MAE of 1.47 mm/day. However, the modest R^2 value of 0.20 indicates that elevation accounts for only a fraction of the variability in rainfall during this season. In the SWM, IDW exhibits improved performance, reflected in a lower RMSE (0.77 mm/day) and MAE (0.69 mm/day). The higher R^2 value of 0.43 suggests that elevation is a more robust predictor of rainfall in this season. The IM season, characterised by transitional climatic conditions, proves challenging for IDW, with higher errors (RMSE and MAE) and an exceptionally low R^2 of 0.01. This underscores the limitations of relying solely on elevation to explain rainfall variability during periods of climatic transition.

Turning to Local Polynomial Interpolation (LPI), the method mirrors IDW in the NEM season, offering a comparable level of accuracy (RMSE = 2.08 mm/day, MAE = 1.44 mm/day, R^2 = 0.21). The SWM, however, sees LPI outperforming IDW with lower errors (RMSE = 0.70 mm/day, MAE = 0.64 mm/day) and a higher R^2 of 0.53. This signals the effectiveness of LPI in capturing the rainfall-elevation relationship during the SWM. Despite advancements over IDW in the IM, particularly in lower errors, LPI still struggles to explain rainfall variability, as evidenced by a low R^2 of 0.02. The transitional nature of the IM season poses inherent challenges in establishing a robust relationship with elevation alone.

The spatial patterns unveiled by GWR emphasise the need for a nuanced understanding of rainfall-elevation dynamics. With its adaptability to capturing complex spatial variations, LPI proves more effective, especially during the SWM. However, the persistent challenge in explaining rainfall variability during transitional periods, as seen in both methods during the IM, calls for a more comprehensive exploration of contributing variables.

This GWR analysis sheds light on the spatial variations in the relationship between rainfall and elevation across seasons in PM. The findings underscore the importance of context-specific methodologies and highlight the need for further research to uncover additional variables influencing rainfall patterns. Such insights are pivotal for the region's informed water resource management and climate resilience.

Table 2 *GWR analysis between seasonal interpolated rainfall and elevation in PM*

Season	IDW			LPI		
	RMSE	MAE	R^2	RMSE	MAE	R^2
NEM	2.14	1.47	0.20	2.08	1.44	0.21
SWM	0.77	0.69	0.43	0.70	0.64	0.53
IM	1.44	1.24	0.01	1.30	1.10	0.02

4. Conclusion

The seasonal study emphasises the vital need to customise interpolation methods to the specific characteristics of each season. While IDW exhibits commendable performance in certain seasons, the LPI method, renowned for capturing intricate spatial patterns, emerges as a robust choice for seasons marked by heightened variability, such as the NEM. This nuanced understanding of seasonal variations provides valuable guidance in selecting the most suitable interpolation method for ensuring accurate and context-specific rainfall estimations in PM.

Implementing the GWR model to investigate the relationship between rainfall and elevation has been a focal point in this study, particularly when assessed on a seasonal basis. The GWR evaluation, integrated with IDW and LPI methods, considered metrics such as RMSE, MAE, and R^2 at the seasonal scale. The outcomes reveal varied performance for IDW and LPI when elevation is incorporated through GWR, emphasising the need for a seasonally nuanced approach in understanding the effectiveness of these interpolation methods in the study area. Future studies can also consider adopting more advanced interpolation methods, such as kriging or machine learning-based approaches, to assess the potential for improving accuracy. Apart from that, relevant variables, such as land use, topography, and vegetation indices, should be included to offer a more comprehensive understanding of spatial variability in rainfall.

Acknowledgement

This work was supported by the Ministry of Higher Education (MOHE) through FRGS grant (grant number: FRGS-EC/1/2024/TK06/UMS/02/5). We also appreciate the support from the Malaysia Toray Science Foundation (4417/0005). We acknowledge the Department of Irrigation and Drainage Malaysia for making the essential data available.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***Study conception and design:*** *Eugene Zhen Xiang Soo & Ren Jie Chin;* ***data collection:*** *Eugene Zhen Xiang Soo;* ***Analysis and interpretation of results:*** *Eugene Zhen Xiang Soo & Ren Jie Chin;* ***Validation of results:*** *Foo Wei Lee;* ***Draft manuscript preparation:*** *Eugene Zhen Xiang Soo;* ***Proofread & editing:*** *Janice Lynn Ayog & Nurmin Bolong. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., Brakenridge, G. R., Policell, F., Habib, S., & Irwin, D. (2011). Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria Basin: Implications for hydrologic prediction in ungauged basins. IEEE Transactions on Geoscience and Remote Sensing, 49, 85-95.

[2] Seyyedi, H., Anagnostou, E. N., Beighley, E., & McCollum, J. (2014). Satellite-driven downscaling of global reanalysis precipitation products for hydrological applications. HESS, 18, 5077-5091. https://doi.org/10.5194/hess-18-5077-2014

[3] Scofield, R. A., & Kuligowski, R. J. (2003). Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. Weather and Forecasting, 18, 1037-1051. https://doi.org/10.1175/1520-0434(2003)018<1037:saooos>2.0.co;2

[4] Canli, E., Loigge, B., & Glade, T. (2018). Spatially distributed rainfall information and its potential for regional landslide early warning systems. Natural Hazards, 91, 103-127. https://doi.org/10.1007/s11069-017-2953-9

[5] von Ruette, J., Lehmann, P., & Or, D. (2014). Effects of rainfall spatial variability and intermittency on shallow landslide triggering patterns at a catchment scale. Water Resources Research, 50, 7780-7799. https://doi.org/10.1002/2013WR015122

[6] Militino, A. F., Ugarte, M. D., Goicoa, T., & Genton, M. (2015). Interpolation of daily rainfall using spatiotemporal models and clustering. International Journal of Climatology, 35, 1453-1464. https://doi.org/https://doi.org/10.1002/joc.4068

[7] Fábián, Z. (2014). Method of the geographically weighted regression and an example for its application. Regional Statistics, 1, 61-75.

[8] Ceron, W. L., Andreoli, R. V., Kayano, M. T., Canchala, T., Carvajal-Escobar, Y., & Souza, R. A. (2021). Comparison of spatial interpolation methods for annual and seasonal rainfall in two hotspots of biodiversity in South America. Anais da Academia Brasileira de Ciências, 93, e20190674. https://doi.org/10.1590/0001-3765202120190674

[9] Hong, Y., Nix, H. A., Hutchinson, M. F., & Booth, T. H. (2005). Spatial interpolation of monthly mean climate data for China. International Journal of Climatology, 25, 1369-1379. https://doi.org/10.1002/joc.1187

[10] Fagandini, C., Todaro, V., Tanda, M. G., Pereira, J. L., Azevedo, L., & Zanini, A. (2023). Missing rainfall daily data: A comparison among gap-filling approaches. Mathematical Geosciences, 56, 191–217. https://doi.org/ 10.1007/s11004-023-10078-6

[11] FirdausHum, N. N. M., & Talib, S. (2016). Spatial interpolation of monthly precipitation in Selangor, Malaysia-comparison and evaluation of methods. Journal of Applied and Physical Sciences, 2, 1-9.

[12] Fung, K. F., Chew, K. S., Huang, Y. F., Ahmed, A. N., Teo, F. Y., Ng, J. L., & Elshafie, A. (2022). Evaluation of spatial interpolation methods and spatiotemporal modeling of rainfall distribution in Peninsular Malaysia. Ain Shams Engineering Journal, 13, 101571. https://doi.org/https://doi.org/10.1016/j.asej.2021.09.001

[13] Burhanuddin, S. N. Z. A., Deni, S. M., & Ramli, N. M. (2017). Imputation of missing rainfall data using revised normal ratio method. Advanced Science Letters, 23, 10981-10985. https://doi.org/10.1166/asl.2017.10203