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



Performance Evaluation of Missing Data Treatment Methods for the Large-Scale Evaporation

Intan Najiha Akhmar Saharuddin[1], Nurul Nadrah Aqilah Tukimat[1,*], Marlinda Abdul Malek[2], Bambang Winarta[3]

[1] Faculty of Civil Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang Darul Makmur, Malaysia
[2] Faculty of Civil Engineering, Universiti Teknologi Malaysia, Jalan Iman, 81310 Skudai, Johor Darul Takzim, Malaysia
[3] Department of Water Resources Engineering, Universitas Brawijaya, Malang, Indonesia

ARTICLE INFO

Article history:
Received 12 June 2025
Received in revised form 8 September 2025
Accepted 5 December 2025
Available online 29 July 2026

Keywords:
Evaporation, missing values, data treatment and selected method

ABSTRACT

Missing data in evaporation records can cause inaccurate predictions and poor management strategies, leading to significant economic, environmental, and social impacts. This is especially relevant for evaporation datasets in Pahang state because the high amount of missingness might impact the core pattern in the highly variable time series. The best approach for estimating missing evaporation data in various places may differ based on the evaporation pattern and geographical distribution. In this paper, there were 6 evaporation stations in the Pahang state with daily data between 2008 until 2022 (14 years) were used to examine different imputation methods. The selected methods were Arithmetic Mean Method (AMM), Multiple Imputation Technique (MIT) and Normal Ratio Method (NRM). As a result, the purpose of this research was to explore and evaluate three distinct strategies for treating missing data. After investigation, AMM was determined to be the best approach based on root mean square error (RMSE), mean absolute error (MAE), correlation coefficient (r), and percentage of error (% of error).

1. Introduction

The water cycle comprises three primary processes: evaporation, condensation, and precipitation. Evaporation refers to the transformation of liquid water at the surface into water vapor. During this process, water molecules absorb heat energy from the surrounding environment, which subsequently lowers the water temperature and alters the concentration of water molecules (Wirangga et al., 2023). As a fundamental component of the hydrological cycle, evaporation from water surfaces plays a critical role in maintaining the atmospheric water balance, regulating soil moisture content, and influencing both infiltration rates and groundwater recharge (Condon et al., 2020; Goh et al., 2022; Zhao & Gao, 2019). In the context of vegetated surfaces, potential evaporation (PE) represents the volume of water lost to the atmosphere under conditions of unlimited water supply (Kay et al., 2013).

* *Corresponding author.* nadrah@umpsa.edu.my

The significance of evaporation is further underscored by estimates indicating that reservoir evaporation can account for up to 50% of reservoir capacity losses and approximately 20% of total water input in lakes worldwide (Abdallah et al., 2021; Zhao & Gao, 2019).

Evaporation measurement instruments are also susceptible to data loss in hydrological observations. Inadequate maintenance of rain gauges and evaporation measuring instruments can introduce systematic errors, resulting in underestimated measurements or missing observations due to the accumulation of dirt and debris (Eljona & Lorena, 2022). When missing observations constitute a substantial proportion of a dataset, they require careful evaluation to ensure the reliability and integrity of subsequent analyses. The present study utilized a complete dataset obtained from six evaporation stations in Pahang, Malaysia, provided by the Malaysian Meteorological Department (MET Malaysia), covering the period from 2008 to 2022.

Prior to any statistical analysis, missing data must be addressed to ensure that the dataset is complete, homogeneous, and reliable. Gap-filling in daily meteorological datasets is recognized as one of the most critical preparatory steps before conducting climate or hydrological analyses (Fenta et al., 2018; Hassan & Croke, 2013). Accurate estimation of missing climatic data remains a persistent challenge, particularly when working with large time series characterized by varying weather distributions (Simolo et al., 2010). This challenge is further compounded when large volumes of low-quality data records are present.

Recent studies have introduced advanced statistical and machine learning approaches to enhance the accuracy of missing data estimation in hydrological and meteorological datasets. İnan et al. (2025) compared interpolation and machine learning methods for sub-hourly rainfall data imputation in data-scarce environments, demonstrating the effectiveness of hybrid approaches for high temporal resolution datasets. Magyari-Sáska et al. (2025) conducted a comparative analysis of self-imputation methods for river flow datasets and proposed quality assessment frameworks to evaluate imputation performance. Zhang et al. (2025) developed a coupled model with an innovative imputation strategy for missing groundwater monitoring data, demonstrating improved reliability in hydrological applications. Chen et al. (2024) proposed ensemble artificial intelligence models capable of addressing multiple missing precipitation data patterns with enhanced predictive accuracy. Wongoutong (2025) systematically evaluated various imputation techniques under different missing data mechanisms including Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) with particular focus on nonstationary time series datasets. Similarly, Wu et al. (2025) introduced a multidimensional performance evaluation framework for comparing missing value imputation methods, emphasizing the importance of selecting robust techniques for environmental and hydrological applications.

Missing data management is a critical concern in effective data governance and scientific analysis (Farzandi et al., 2019; Little & Rubin, 2002). In climatic time series, missing observations commonly arise from a range of factors, including manual data collection errors, instrument malfunction, sensor failure, and data transmission problems. These issues frequently produce both random and systematic missing data patterns in precipitation and hydrological time series (Aguilera et al., 2020).

2. Study area

The study area encompassed the east coast Malaysia of the states in Pahang. Fifteen years (2008-2022) of evaporation station records from six sites utilized in the Statistical Downscaling Model (SDSM) and all stations has completed data. However, in this study, it was assuming 10 missing data from 1 station and has been treated using 3 different method which are Arithmetic Mean Method

(AMM), Multiple Imputation Technique (MIT) and Normal Ratio Method (NRM). Figure 1 shows the maps of location evaporation stations in Pahang.

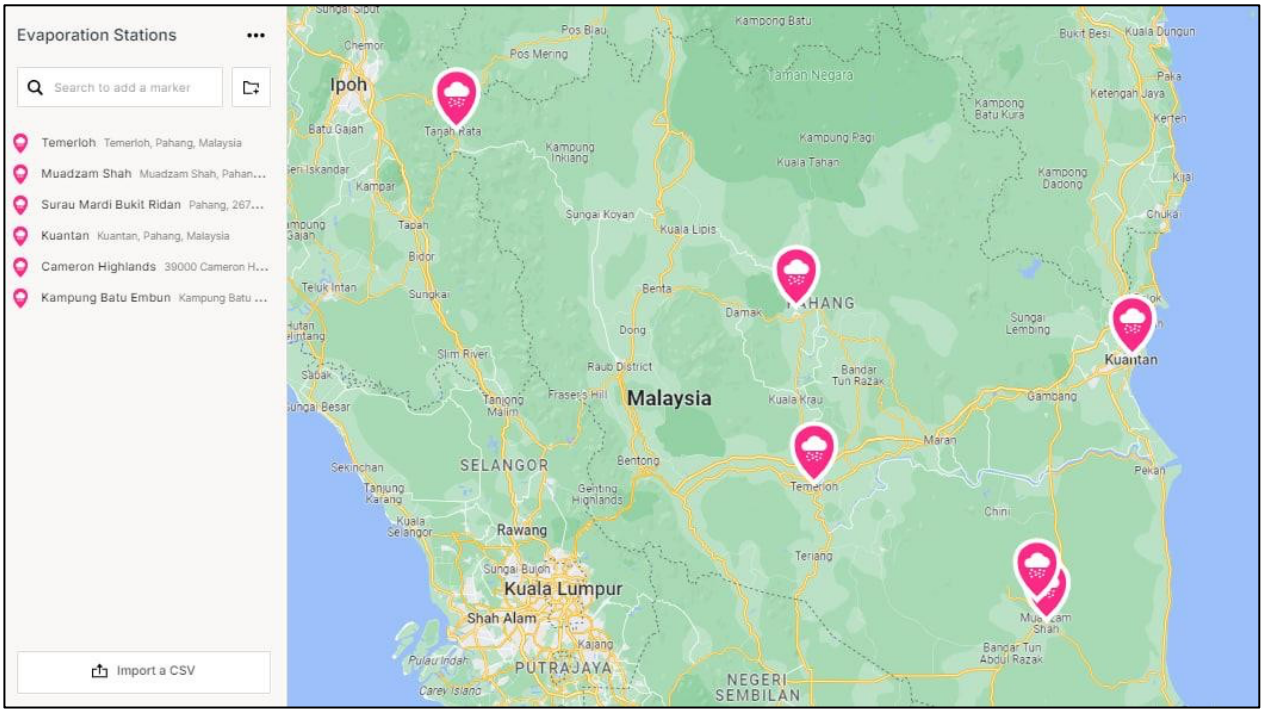


Fig. 1. Location of evaporation station in Pahang

3. Methodology

Figure 2 presents a flowchart illustrating the overall research process. In this study, complete datasets from all evaporation stations were initially obtained. A subset of observations was then intentionally removed to simulate missing data conditions, and the resulting gaps were subsequently estimated using three imputation methods: the Arithmetic Mean Method (AMM), the Multiple Imputation Technique (MIT), and the Normal Ratio Method (NRM). One control station was designated to evaluate the effectiveness and performance of each method, with all imputation procedures applied using data from two neighbouring stations with complete records. Each method was selected based on its distinct characteristics and relative strengths in estimating missing observations.

The accurate imputation of missing data using appropriate techniques remains a fundamental requirement in hydrological time series analysis. A thorough understanding of the causes and mechanisms underlying data absence is essential for effective imputation, as it facilitates the establishment of meaningful relationships between measured variables and patterns of data incompleteness (Chiu et al., 2019a). Recent studies have further highlighted the growing application of advanced machine learning and statistical approaches in the reconstruction of hydrological and meteorological datasets. Solanki et al. (2025) demonstrated the effectiveness of machine learning approaches in filling streamflow data gaps within data-scarce catchments. Levy et al. (2025) introduced an automated R package designed for imputing regionally correlated hydrological records, thereby improving computational efficiency in large-scale hydrological datasets. Nair et al. (2025) integrated physical modelling with statistical scaling techniques to address gap-filling challenges in groundwater level time series with irregular temporal resolution. Widyawati et al. (2025) evaluated multivariate missing data imputation methods specifically tailored for climate

datasets, while Kumar et al. (2025) compared multiple imputation approaches for rainfall data in humid tropical river basins, findings that are particularly relevant to Malaysian climatic conditions. Furthermore, Alfarelzi et al. (2025) proposed a multi-method framework for rainfall data imputation that integrates several estimation techniques to enhance prediction reliability. Sharma et al. (2025) similarly demonstrated the applicability of machine learning approaches in imputing missing daily rainfall data under complex environmental conditions.

The performance of each imputation method was assessed using four statistical indicators: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), correlation coefficient (r), and percentage of error (% error). The method yielding the most accurate results based on these statistical criteria was subsequently applied to stations with actual missing observations, as previously identified. Finally, the data distribution following missing data treatment was compared against three stations with complete records to verify that the imputed values remained consistent with observed hydrological behaviour.

Here are the basic of identifying missing data estimation of missing precipitation data:

i. Select stations with few numbers of missing data
ii. Arrange similar data based on day/month
iii. Estimate the missing

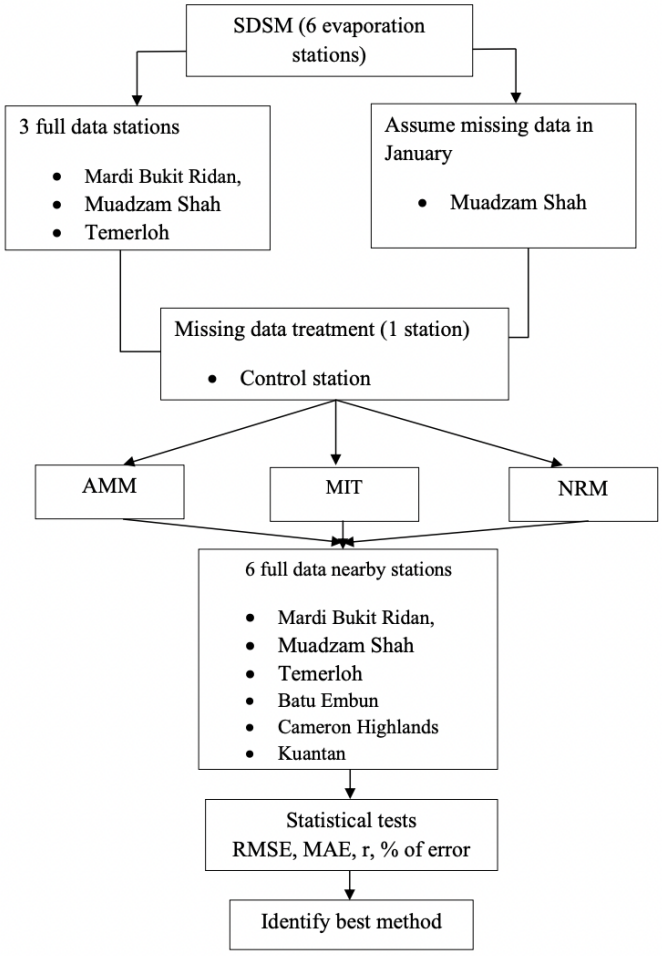


Fig. 2: Flow chart of the methodology

3.1 Arithmetic Mean Method (AMM)

For better estimation purpose data should be arranged with similar month/ day/ year pattern. Missing data is obtained by computing the arithmetic average of the data corresponding to the nearest weather stations.

$$V_o = \frac{\sum_{i=1}^{n} V_i}{N} \quad (1)$$

where V_o is the estimated value of the missing data, V_i is the value of same parameter at its nearest weather station, and N is the number of the nearest stations.

3.2 Multiple Imputation Technique (MIT)

Multiple imputation is a statistical technique used to handle missing data by imputing multiple sets of plausible values for any missing observations (Presti et al, 2010). This method is particularly useful when the missing data is assumed to be missing at random (MAR), meaning that the probability of missingness may depend on observed data but not on unobserved data (Zhu, 2014). MI is a predictive strategy to managing missing data in multivariate analysis that combines classical and Bayesian statistical approaches. The goal behind MI is to impute a missing value multiple times, resulting in numerous (m) data sets. To apply this method, it must follow the following steps:

$$P_\chi = \frac{\sum_{i=1}^{k} I_i(P_\chi, P_i)}{N} \quad (2)$$

i. Find k similar databases for each missing value, then the observed values will be used to impute MD.

ii. ii. For each MD (P_χ), we use data imputations Pi by applying regression methods to obtain k different estimate results I_i (P_χ, P_i).

iii. The final result P_χ will be obtained by combining all imputations results I_i using average of all the k complete data values, that is:

3.3 Normal Ratio Method (NRM)

If the normal precipitations vary more than 10% then, Px is estimated by weighing the precipitation at various stations by the ratios of normal annual precipitation. The normal ratio method gives, Px as:

$$Px = \frac{Nx}{3}\left(\frac{P_1}{N_1} + \frac{P_2}{N_2} + \frac{P_3}{N_3}\right) \quad (3)$$

4. Data collection

For this study, the MET Malaysia provided evaporation data from 6 evaporation stations in Pahang, covering the years 2008 – 2022. Table 1 shows the list of evaporation station in Pahang.

Table 1
List of Evaporation Station in Pahang

No.	Station ID	Station Name	Latitude (N)	Longitude (E)	Elevation (m)
1.	4023001	Batu Embun	03 58 00	102 21 00	59.5
2.	4513033	Cameron Highlands	04 28 00	101 22 00	1545.0
3.	3629103	Kuantan	03 46 00	103 13 00	15.2
4.	7687308	Mardi Bukit Ridan	03 02 18	103 06 17	26.0
5.	2636182	Muadzam Shah	03 03 00	103 05 00	33.3
6.	3428001	Temerloh	03 28 00	102 23 00	39.1

5. Statistical Analysis

The statistical tests used to compare the efficacy of the selected approaches are Root Mean Squared Error (RMSE), Mean absolute error (MAE), r, and percentage of error. r, RMSE, and MAE are the most widely used and accepted in many statistical analyses. (Amirabadizadeh Et al, 2016; Bennett Et. Al, 2013). The equations are stated as follow:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 (\sum x)^2] + [n\sum y^2 (\sum y)^2]}} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)}{n}} \quad (5)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |x_i y_i|}{n} \quad (6)$$

$$\% \text{ of error} = \frac{|y-x|}{x} \text{x } 100 \quad (7)$$

Where x is the observed value, y represents the computed value and n denotes the number of data observations.

6. Results and Discussion

Table 1 shows the list of evaporation stations in the study where 3 of the stations are full while another 10 have missing data. The minimum missing days is 28 days (7.7%) for Missing 1 station and maximum missing days are 79 days (21.6%) for Missing 3 and Missing 4 stations. Table 2 shows the summary of result for missing data treatment by different methods. It can be seen AMM give the nearest number to the actual data.

Table 2
Result for missing data treatment by different methods

Actual Data	Arithmetic Mean Method (AMM)	Multiple Imputations Technique (MIT)	Normal Ratio Method (NRM)
2.3	1.15	2.57	3.55
1	0.5	2.49	0.25
4.4	2.2	0.97	0.35
1.5	0.75	0.83	1.15
1.1	0.55	1.42	1.29
3.2	1.6	0.90	0.95
0	0	2.46	0.50
0.5	0.25	1.335	1.51
2.6	1.3	2.073	1.75
2.4	1.2	1.119	2.59

Table 3
Statistical values of every method

Method	Actual	Predicted	MSE	RMSE	MAE	r	% of error
AMM	2.3	1.15	0.91	1.32	0.83	1.15	100
	1	0.5	4.97	0.25	1.35	0.5	100
	4.4	2.2	2.98	4.84	1.48	2.2	100
	1.5	0.75	0.58	0.56	0.65	0.75	100
	1.1	0.55	2.71	2.56	1.08	1.6	100
	3.2	1.6	1.28	0.00	0.8	0	100
	0	0	0.06	0.00	0.13	0	0
	0.5	0.25	1.72	0.06	0.78	0.25	100
	2.6	1.3	2.29	1.69	1.25	1.3	100
	2.4	1.2	0.79	1.44	0.47	1.2	100
MIT	2.3	2.57	2.26	0.07	0.88	0.27	10.51
	1	2.49	12.87	2.22	0.97	1.49	59.84
	4.4	0.97	6.33	11.76	2.05	3.43	353.61
	1.5	0.83	0.33	0.45	0.18	0.67	80.72
	1.1	1.42	5.34	0.10	0.99	0.32	22.54
	3.2	0.90	8.70	5.29	0.08	2.3	255.56
	0	2.46	3.72	6.05	1.65	2.46	100.00
	0.5	1.34	0.63	0.70	0.16	0.84	62.55
	2.6	2.07	1.78	0.28	0.90	0.53	25.42
	2.4	1.12	2.38	1.64	0.02	1.28	114.48
NRM	2.3	3.55	1.34	1.56	0.25	1.25	35.21
	1	0.25	16.68	0.56	2.4	0.75	300.00
	4.4	0.35	8.32	16.40	2.2	4.05	1157.14
	1.5	1.15	0.10	0.12	0.08	0.35	30.43
	1.1	1.29	5.08	0.04	1.03	0.19	14.73
	3.2	0.95	2.78	5.06	0.88	2.25	236.84
	0	0.50	1.15	0.25	0.76	0.5	100.00
	0.5	1.51	1.23	1.02	0.08	1.01	66.89
	2.6	1.75	0.40	0.72	0.33	0.85	48.57
	2.4	2.59	0.02	0.04	0.095	0.19	7.34

According to Table 3, the AMM approach has the lowest RMSE, MAE, and percentage of error values. The AMM approach has the least RMSE and MAE values when compared to the MIT and NRM methods. The lower the RMSE and MAE, the more accurate the formula (Nguyen and Ho, 2018). The reduced percentage of inaccuracy between observed data on the ground and anticipated data using

missing treatment methods demonstrated that in this study, AMM is the best appropriate approach for treating missing data.

The graph in Figure 3 clearly illustrates that the trendline in (a) AMM is more exact and consistent throughout a wide range of trends. When compared against the EM and MIT approaches, the AMM method scores the highest. The greater the value of r, the stronger the correlation between the observed data set and anticipated data following missing data treatment. The findings with 10 missing data are the best, demonstrating that AMM is the superior approach among others. The RMSE value is 0.25, and the MAE is 1.35, which are lower than the MIT and NRM approaches. Meanwhile, the r-value is 0.73, which ranks among the greatest. In terms of percentage of mistake, 3.33% is the lowest and highest figure. The MI technique produced the greatest values of RMSE.

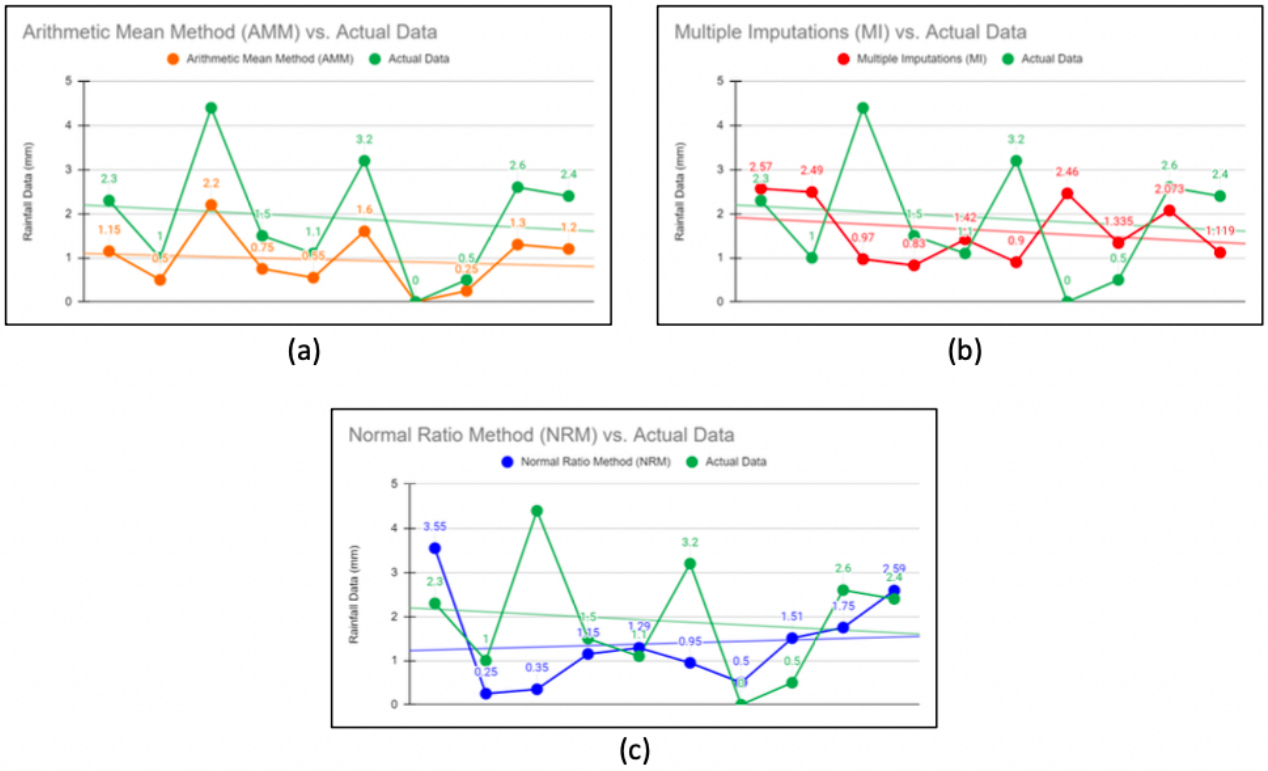


Fig. 3. Comparison of daily rainfall data with measured values, methods abbreviation as following: **a.** Arithmetic Mean Method (AMM); **b.** Multiple Imputations Technique (MIT); **c.** Normal Ratio Method (NRM)

7. Conclusions

This investigation yields several findings. AMM yields the best results with the fewest errors, followed by MIT techniques. Overall, all three techniques generate good forecasts for the suggested missing data. Also, there are no discernible differences between approaches that use a single component, such as location mean value, and those that use numerous combination variables, if the research region is confined to the Pahang zone. Errors in projections rise when precipitation amounts in the area fall, as seen in the dry year comparison data. In addition, disparities were found between the findings of the tested methodologies in other studies, although in more difficult terrain. Furthermore, further data collection may be necessary, as climate change and rainfall trend changes are predicted to provide new information.

Acknowledgement
This research is supported by Universiti Malaysia Pahang Al-Sultan Abdullah under grant number RDU230112 (FRGS/1/2023/TK06/UMP/02/1), Tabung Persidangan Dalam Negeri (TPDN-UMPSA) entitled Evaluating the Performances of Missing Data Treatment Methods for the Large-Scale Evaporation.

References

[1] Abdallah A. M., Parihar C. M. Patra S., Nayak H. S., Saharawat Y. S. and Upendra S. "Critical evaluation of functional aspects of evaporation barriers through environmental and economics lens for evaporation suppression - A review on milestones from improved technologies" *Science of The Total Environment, 788, page 1-19.* https://doi.org/10.1016/j.scitotenv.2021.147800

[2] Aguilera, H., Carolina, G. A., & Carmen, S. H. (2020). Estimating extremely large amounts of missing precipitation data. Journal of Hydroinformatics, 22(3): 578–592.

[3] Aieb A., Madani K., Scarpa M., Bonaccorso B. and Lefsih K. "A new approach for processing climate missing

[4] Alfarelzi, H., Susanto, I., & Wibowo, A. (2025). Imputation of missing values for generating rainfall data using a multi-method approach. *Proceedings of ICSGRC 2025*. https://doi.org/10.1109/icsgrc65918.2025.11159759

[5] Amirabadizadeh M., Ghazali A. H., Huang Y. F. and Wayayok A. "Downscaling daily precipitation and temperatures over the Langat River Basin in Malaysia: a comparison of two statistical downscaling approaches." *International Journal of Water Resources and Environmental Engineering, 8, no. 10 (2016),* page 120-136. https://doi.org/10.5897/IJWREE2016.0585

[6] Bennett N. D., Croke B. F. W., Guariso G., Guillaume J. H. A., Hamilton S. H., Jakeman A. J., Marsilli-Libelli S., Newham L. T. H., Norton J. P., Perrin C., Pierce S. A., Robson B., Seppelt R., Voinov A. A., Fath B. D. and Andreassian V. "Characterising performance of environmental models." *Environmental Modelling and Software" 40 (2013),* page 1-20. https://doi.org/10.1016/j.envsoft.2012.09.011

[7] Chen, L., Zhang, Y., Zhou, J., & Wang, X. (2024). Multiple types of missing precipitation data filling based on ensemble artificial intelligence models. *Water*, 16(22), 3192. https://doi.org/10.3390/w16223192

[8] Chiu, P.C., Selamat, A., Krejcar, O. "Infilling missing rainfall and runoff data for Sarawak, Malaysia using Gaussian mixture model based K-nearest neighbor imputation." Lect. Notes Computer Scinces. (2019), page 27-38. https://doi.org/10.1007/978-3-030-22999-3_3

[9] Condon L. E., Atchley A. L. and Maxwell R. M. "Evapotranspiration depletes groundwater under warming over the contiguous United States." Article of Nature Communications*, Department of Hydrology and Atmospheric Sciences, 11, no. 873 (2020)*, page 1-8. https://doi.org/10.1038/s41467-020-14688-0

[10] Eljona M. T. and Lorena M. Z. "Comparison Methods of Estimating Missing Data in Real Data Time Series." *Asian-European Journal of Mathematics, 15, No. 10 (2022).* https://doi.org/10.1142/S1793557122502436

[11] Farzandi M,, Sanaeinejad H., Rezaei-Pazhan H. and Sarmad M. "Improving estimation of missing data in historical monthly precipitation by evolutionary methods in the semi-arid area." Environment, Development and Sustainability, 24, (2022), page 8313-8332. https://doi.org/10.1007/s10668-021-01784-4

[12] Farzandi M., Sanaeinejad H., Ghahraman B., and Sarmad M. (2019). Imputation of missing meteorological data with evolutionary and machine learning methods, case study: long-term monthly precipitation and temperature of Mashhad. Journal of Water and Soil, 33(2), 361–377.

[13] Fenta D. M. and Disse M. "Analyzing the future climate change of Upper Blue Nile River basin using statistical downscaling techniques." *Hydrology and Earth System Sciences, 22, no. 4. (2018).* Page 2391–2408. https://doi.org/10.5194/hess-22-2391-2018

[14] Goh E. G. S., Noranizam M. S., Mohammad F. A., Nurul Ameera M. R., Rudiyanto, Roslaili A. A. "Modelling Water Consumption Efficiency Based on Perlis State Soil Series" *Journal of Advanced Research in Applied Sciences and Engineering Technology, 28, no. 1 (2022).* Page 25-32.

[15] Hassan M. M. and Croke B. F. W. "Filling gaps in daily rainfall data: a statistical approach." *Integrated Catchment Assessment and Management (iCAM) Centre, and the National Centre for Groundwater Research and Training, (2013).* Page 380-386. 10.36334/modsim.2013.a9.hasan

[16] İnan, M. C., Başakın, E. E., & Ekmekcioğlu, Ö. (2025). Interpolation and machine learning methods for sub-hourly missing rainfall data imputation in a data-scarce environment: One- and two-step approaches. *Hydrology*, 12(11), 297. https://doi.org/10.3390/hydrology12110297

[17] Kay A. L. , Bell V. A., Blyth E. M., Crooks S. M., Davies H. N. and Reynard N. S. "A hydrological perspective on evaporation: historical trends and future projections in Britain." *Journal of Water and Climate Change*, 4 no. 3 (2013), page 193-208. https://doi.org/10.2166/wcc.2013.014

[18] Kumar, S., Roshni, T., & Himayoun, D. (2025). Comparison of the multiple imputation approaches for imputing rainfall data: A humid tropical river basin case study. *Water Conservation Science and Engineering*, 403. https://doi.org/10.1007/s41101-025-00403-x

[19] Levy, Z. F., Jurgens, B. C., McCallum, A. M., & Foks, N. L. (2025). ARCHI: A new R package for automated imputation of regionally correlated hydrologic records. *Ground Water*, 13474. https://doi.org/10.1111/gwat.13474

[20] Little R. J. A., & Rubin, D. B. "Statistical analysis with missing data" (p. 408). Wiley Online Library, (2022).

[21] Magyari-Sáska, Z., Dombay, Ş., & Șerban, G. (2025). Experimental comparative study on self-imputation methods and their quality assessment for monthly river flow data with gaps: Case study to Mures River. *Applied Sciences*, 15(3), 1242. https://doi.org/10.3390/app15031242

[22] Nair, A., Collenteur, R., & Bakker, M. (2025). Gap-filling groundwater level time series of irregular temporal resolution using physical modeling (Pastas) and simple statistical scaling. *EGUsphere*. https://doi.org/10.5194/egusphere-egu25-10107

[23] Nguyen L. and Ho T. T. "Early Fetal Weight Estimation with Expectation Maximization Algorithm." *International Technology and Science Press 1, (2018),* page 45-65. https://doi.org/10.31058/j.em.2018.11002

[24] Presti R. L., Barca E. and Passarella G. "A Methodology for Treating Missing Data Applied to Daily Rainfall Data in the Candelaro River Basin (Italy)." Environ. Monit. Assess., 1-4, no. 160 (2010), page 1-22. https://doi.org/10.1007/s10661-008-0653-3

[25] Radi N., Zakaria R. and Azman M. "Estimation of missing rainfall data using spatial interpolation and imputation methods" *AIP Conference Proceedings, 1643, no. 1 (2015),* page https://doi.org/10.1063/1.4907423

[26] Sharma, S., Sood, V., Kaur, M., & Singh, S. (2025). Comparative analysis of machine learning methods for imputing missing daily rainfall data in complex Himalayan terrain. *International Journal of Climatology*, 70122. https://doi.org/10.1002/joc.70122

[27] Simolo C., Brunetti M., Maugeri M. and Nanni T. "Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach." *International Journal of Climatology, 30 (2010)*. Page 1564-1576. https://doi.org/10.1002/joc.1992

[28] Solanki, V., Kumar, A., & Singh, R. (2025). Filling streamflow data gaps in Indian catchments using machine learning. *EGUsphere*. https://doi.org/10.5194/egusphere-egu25-15030

[29] Widyawati, R., Kusuma, W. A., & Nurdiati, S. (2025). Performance of multivariate missing data imputation methods on climate data. *Journal of Applied Informatics and Computing*, 9(6), 11316. https://doi.org/10.30871/jaic.v9i6.11316

[30] Wirangga R., Dan Mugisidi D., Sayuti A. T. and Heriyani O. "The Impact of Wind Speed on the Rate of Water Evaporation in a Desalination Chamber." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, 106, no. 1 (2023).* Page 39-50.

[31] Wongoutong, C. (2025). Performance comparison of various imputation methods for missing data mechanisms (MAR, MCAR, and MNAR) in a nonstationary time series. *International Journal of Mathematics and Mathematical Sciences*, 3031708. https://doi.org/10.1155/ijmm/3031708

[32] Wu, X., Li, Y., & Chen, Z. (2025). A comparative study of missing value imputation methods based on multidimensional performance evaluation. *Proceedings of ICCECT 2025*. https://doi.org/10.1109/iccect64621.2025.11339529

[33] Zhang, Y., Wang, H., & Liu, J. (2025). The superior coupled model with innovative strategy for accurate imputation of missing hydrological monitoring data in water research—A case of groundwater level data. *Water Research X*, 100398. https://doi.org/10.1016/j.wroa.2025.100398

[34] Zhao, G., Gao, H. "Estimating reservoir evaporation losses for the United States: fusing remote sensing and modeling approaches." *Remote Sensing of Environment, 226 (2019).* Page 109-124. https://doi.org/10.1016/j.rse.2019.03.015

[35] Zhu X. "Comparison of Four Methods for Handing Missing Data in Longitudinal Data Analysis through a Simulation Study" *Open Journal of Statistics, 4, no. 11 (2014).* Page 933-944. http://dx.doi.org/10.4236/ojs.2014.411088