

Analysis of the Correlation Between Rainfall Intensity and Water Level at Machap Dam Catchment Area

Nur Tijani Ilmi Zulkiflie, Arrthy Nair Paramasivam, Puteri Alya Adriana Yusuan, Suhaila Sahat*

Department of Civil Engineering, Centre for Diploma Studies, Universiti Tun Hussein Onn Malaysia, Pagoh Higher Education Hub, 84600 Pagoh, Johor, MALAYSIA

*Corresponding Author: suhailasa@uthm.edu.my

DOI: <https://doi.org/10.30880/mari.2024.05.03.003>

Article Info

Received: 01 March 2024

Accepted: 31 July 2024

Available online: 31 December 2024

Keywords

rainfall, water level, normal ratio method, station, Machap Dam

Abstract

Measuring and recording weather elements such as rainfall, temperature, and moisture is critically important in industries, agriculture, aviation, hydrology, and engineering. This research is conducted to evaluate the use of the Normal Ratio Method to estimate rainfall missing data and whether there will be a significant result between rainfall data from 5 total stations and the water level at the Machap Dam catchment area. The correlation between rainfall intensities and water level at Machap Dam was determined by first obtaining missing data at the 5 total stations using the Normal Ratio Method. The study area involved Kluang, Johor. The analysis of rainfall missing data shows that a minimum of 5 total stations are sufficient to obtain the data. Meanwhile, the correlation between rainfall and dam water level indicates only 1 station has a correlation with R square 0.67, and the other 4 stations have recorded less than 0.5 of the R square values. From the analysis, it can be concluded that the study for short-term periods of rainfall data is irrelevant due to the insufficiency of data collection.

1. Introduction

Natural resource management system's successful design and operation's successful design and operation of natural resource management systems require accurate and complete climatological data [1]. Missing rainfall data has often been one of the most difficult problems that researchers had to deal with in both hydrological and meteorological studies [2][3]. However, the issue of missing data arises for a multitude of reasons. The problems with missing rainfall data are caused by the incorrect method used when measuring rainfall, the relocation of the rain station due to urbanization, equipment malfunction and particularly in flooding areas, and errors in the methods utilized to determine the amount of rainfall [2]. The presence of gaps in environmental data time series is a very common, but extremely serious problem [4][5], as it can produce biased results and, in the worst scenario, prevent important analyses of the variables under consideration from being performed [2]. Analyzing the missing value in the rainfall series is critical to carry out an effective rainfall analysis.

Furthermore, extreme rainfall intensity will cause the river to overflow and cause high tides [6]. When there is heavy and prolonged rainfall in a region, a significant amount of water accumulates on the land's surface. This excess water flows into rivers and streams, increasing their discharge. As the river's water level rises beyond its capacity, it overflows its banks, causing flooding in the surrounding areas. The excessive rainfall on the land cannot be absorbed entirely by the soil or vegetation, especially if the ground is already saturated. This surplus water then becomes surface runoff and flows into the river system. If the rainfall is intense and widespread, the volume

of runoff can be substantial, leading to a rapid rise in river water. High tides are typically caused by the moon's and the sun's gravitational forces on Earth's oceans. However, during extreme rainfall events, the increased river flow and runoff can discharge a massive amount of freshwater into estuaries and coastal areas. Once the flow of water on the surface increases, so does the accumulation of rainwater in the river. As a result, rivers and dams are unable to hold large amounts of water at once.

As a matter of fact, it can be assumed that by studying and analyzing the missing rainfall data, it will be able to take preventive measures to avoid the water level of the rivers and dams overflowing which will end up causing a flood. Besides, this study aims to analyze the missing value in the rainfall series, and the Normal Ration Method (NRM) is used. Machap Dam was chosen as the catchment area for the analysis. The objectives of this research study are, (i) to identify rainfall missing data at the Machap Dam catchment area using the Normal Ratio Method and (ii) to determine the correlation between rainfall data and water level in the Machap Dam catchment area.

2. Material and Methods

2.1 Normal Ratio Method

In this analysis, the Normal Ration Method is used. Moreover, the normal ratio method is used when the normal annual precipitation at any of the index stations differs from the interpolation station by more than 10%, the normal ratio method (NRM) is used [7]. This method will calculate the amount of rain using the data ratio method from other nearby stations. This method has a high ability to estimate missing data [8]. Therefore, the method is the most common method used to record rainfall data that is missing from the station.

$$x = \frac{Nx}{m} \left[\frac{P1}{N1} + \frac{P2}{N2} + \dots + \frac{Pm}{Nm} \right] \quad (1)$$

To use this method, a formula has been reserved which is the formula above where Nx = average annual rainfall at the missing data site, Ni = average annual rainfall at a nearby site. Among the information required to use this formula is total rainfall on the same day with missing data and normal annual rainfall totals from at least three nearby stations [8]. Accurate estimation values are necessary for imputing missing rainfall data to provide precise information on meteorological characteristics.

2.2 Multiple Linear Regression Method

Multiple linear regression, as well known as multiple regression, is a statistical technique that predicts the outcome of a response variable using several explanatory variables [9]. Multiple Linear Regression Method is used to analyze the correlation between rainfall intensity and water level. The multiple regressions are graded on a scale of + 1 to -1. + 1 or -1 denotes the correlation between the dependent and independent variables. When one variable rises in tandem with the other, the correlation is significant; when one falls in tandem with the other, the correlation is negative [10]. The value 0 denotes the complete absence of correlation. If the R square value is less than 0.5, there is no correlation between the two variables. R square values more than 0.5 have a correlation and R square values more than 0.8 have a significant correlation [11].

Table 1 Status of R square value

R Square Value	Status
Less than 0.5	No correlation
0.5 to 0.8	Have correlation
More than 0.8	Have significant correlation

2.3 Analysis Data

The study area is from the Department of Irrigation and Drainage (DID) station located at the area where it searches for raw data such as rainfall, and water level that match our objectives. Once the raw data has been collected, it will be sorted to make it easier to calculate. Once calculating the missing data using Normal Ratio Method, the correlation between rainfall intensity and the water level is analyzed using Microsoft Excel.

Among the rainfall stations involved in the study is the station that is nearby stations that surround Machap Dam, which are Machap River Damsite, Lambak Farm, Rubber Plantation See Sun, Renggam Farm, and Simpang Renggam Farm. Machap Dam is located behind Machap Town on Machap River which is Benut River tributary at Kluang, Johor. Fig. 1 shows the location of every station involved.

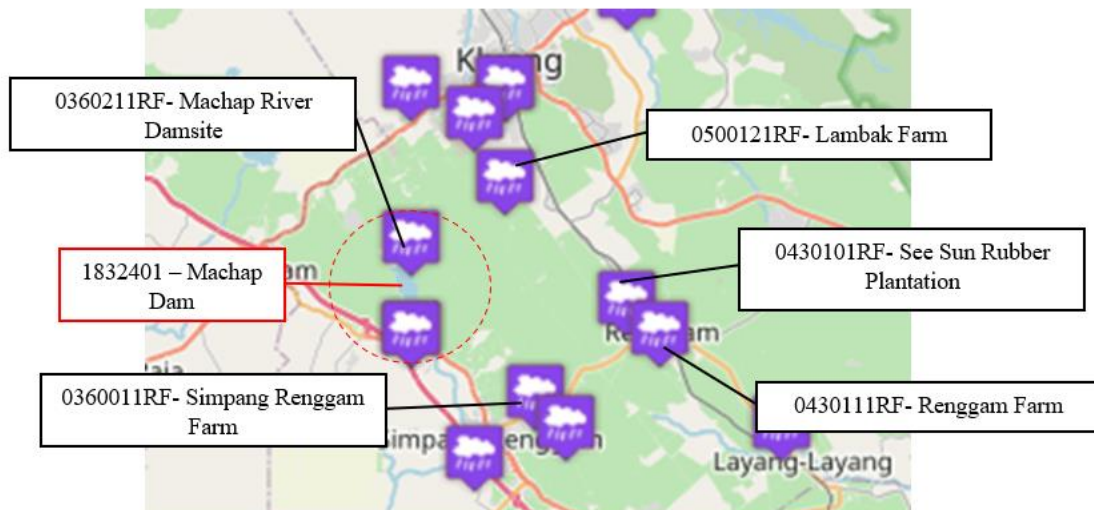


Fig. 1 Rainfall data station

2.4 Data Processing

Annually, rainfall and water level data were extracted and transferred into Microsoft Excel. This procedure was designed to analyze these data. Based on the data, graphs were generated, as well as the correlation and sequences can be identified. Once that data is transferred to Excel, graphs might be easily generated. All the data was converted into graphs to aid in data visualization and analysis. The Regression Method was used to analyze the correlation between rainfall intensity and water level data.

3. Material and Methods

3.1 Rainfall Missing Data

The average annual rainfall amounts for the stations Machap River Damsite, Rubber Plantation See Sun, Lambak Farm, Renggam Farm, and Simpang Renggam Farm are 1980.5mm, 1169.3mm, 1073.2mm, 1198.7mm, and 1173.2mm. The maximum permitted rainfall ranges from 880.2mm to 1075.8mm. The normal ratio method was used because the average annual rainfall amounts of all stations except Lambak Farm station exceed the maximum rainfall allowed. Because the rainfall data for stations Rubber Plantation See Sun, Lambak Farm, Renggam Farm, and Simpang Renggam Farm is missing in 2018 and 2019, the estimated value for the missing data is 354mm and 1560mm. Meanwhile, another year shows that there is no missing rainfall data. Table 2 shows an example of calculation using the normal ratio method to find the missing data.

Table 2 Missing rainfall data calculation for the year 2018

2018/Station	Normal Annual Rainfall (mm)	Amount of Precipitation (mm)
Machap River Damsite	1980.5	800.5
Rubber Plantation See Sun	1169.3	0
Lambak Farm	1073.2	210
Renggam Farm	1198.7	0
Simpang Renggam Farm	1173.2	0

Calculation:

$$Px = \frac{3541.2}{2} \left| \frac{800.5}{1980.5} + \frac{210}{1073.2} \right| = \frac{1062.1}{3} = 354mm$$

From the calculation that has been done like in the example of calculation in Table 2, Table 3 shows the overall data that has been obtained. Based on the analysis, the missing rainfall just happened in 2018 and 2019.

Table 3 Summary of missing rainfall data

Year	Missing Rainfall Data (Px)
2018	354 mm
2019	1560 mm
2020	0
2021	0
2022	0
2023	0

3.2 Correlation Between Rainfall Intensity and Water Level

Table 4 shows the average rainfall and average water level for five stations. Fig. 2 shows the average rainfall and water level graphs for five stations.

Table 4 Average rainfall and average water level at Machap Dam

Month	Rainfall (mm)	Water level (m)
October	9.35	16.86
November	10	12.63
December	4.76	12.5
January	7.56	13
February	1.34	73.5

R square = 0.668

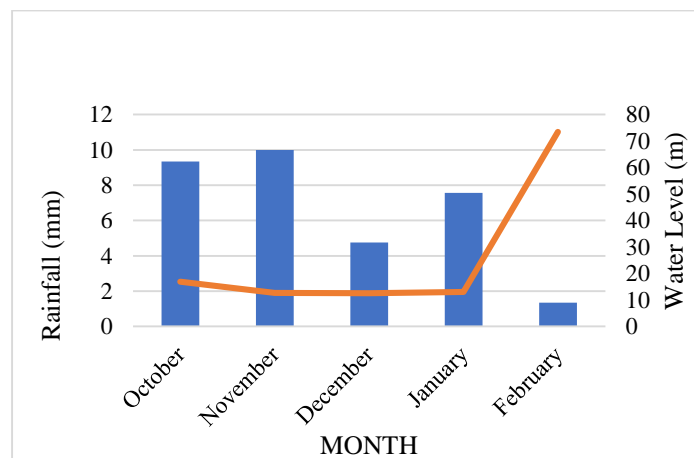


Fig. 2 Graph of average rainfall and average water level at Machap Dam

Table 5 shows the R square value for all stations. Based on Table 1, if the R square value is less than 0.5, there is no correlation between the two variables. R square values more than 0.5 have a correlation and R square values more than 0.8 have a significant correlation. Based on the analysis that has been conducted, it shows that the station of Machap River Damsite and Rubber Plantation See Sun have a correlation between rainfall intensity and water level. Meanwhile, the station of Lambak Farm, Renggam Farm, and Simpang Renggam Farm have no correlation. By recalling Fig. 1, the cause of the correlation does not occur because the station of Lambak Farm and Renggam Farm is too far away from Machap Dam, and the station of Simpang Renggam Farm is located upstream.

Table 5 Summary of R Square Data of Every Station

Station	R Square
Machap Dam	0.67
Rubber Plantation See Sun	0.45
Lambak Farm	0.37
Renggam Farm	0.39
Simpang Renggam Farm	0.20

4. Conclusion

This project needs a lot of research on how to evaluate the missing rainfall data for an area located in Machap Dam using the Normal Ratio Method. Regarding that, monthly precipitation data for every station were approximated using data from nearby stations using the methods to ensure that the real data and the approximated data could be contrasted. Before that, a study on obtaining the best data for this thesis was conducted.

The objectives of this research study have been achieved that is to identify rainfall missing data at the Machap Dam catchment area using the Normal Ratio Method and to determine the correlation between rainfall data and water level in the Machap Dam catchment area. Nevertheless, insufficient data for rainfall and water level shows that the result is irrelevant.

According to the findings, only two of the five stations, Machap River Damsite and Rubber Plantation See Sun have a correlation while the other three, Lambak Farm, Renggam Farm, and Simpang Renggam Farm do not. Overall, this data indicates that there is no correlation compared to those that have a correlation and then that the objective of determining the relationship between rainfall data and water level in Machap Dam has been achieved.

We can assume that studying and analyzing missing rainfall data plays a crucial role in developing a comprehensive flood management and water resource strategy. By using data-driven insights, authorities and communities can be better prepared to handle extreme weather events, reduce flood risks, and safeguard lives and property.

Acknowledgement

The authors are thankful to the government agencies (National Hydrological Network Management System (SPRHiN) and DID) for the primary data provided in this research. Communication of this research is made possible through monetary assistance from the University Tun Hussein Onn Malaysia.

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:** Nur Tijani Ilmi binti Zulkiflie, Arrthy Nair A/P Paramasivam, Puteri Alya Adriana binti Yusuan; **data collection:** Nur Tijani Ilmi binti Zulkiflie, Arrthy Nair A/P Paramasivam, Puteri Alya Adriana binti Yusuan; **analysis and interpretation of results:** Nur Tijani Ilmi Zulkiflie, Arrthy Nair A/P Paramasivam, Puteri Alya Adriana binti Yusuan; **draft manuscript preparation:** Nur Tijani Ilmi binti Zulkiflie. All authors reviewed the results and approved the final version of the manuscript.

Appendix A: An Example

Table 6: Graph of average rainfall and average water level



Table 7: Summary Output of All Station

SUMMARY OUTPUT RUBBER PLANTATION									
<i>Regression Statistics</i>									
Multiple R	0.66881142								
R Square	0.44730872								
Adjusted R Square	0.17096307								
Standard Error	4.03212377								
Observations	4								
ANOVA									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	1	26.31615588	26.31615588	1.618657	0.331188579				
Residual	2	32.51604412	16.25802206						
Total	3	58.8322							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	
Intercept	8.78931074	2.93817852	2.991414811	0.095938	-3.852651088	21.4312726	-3.8526511	21.43127257	
16.86	-0.09744014	0.076587966	-1.272264387	0.331189	-0.426971561	0.23209128	-0.4269716	0.232091278	
SUMMARY OUTPUT LAMBAK FARM									
<i>Regression Statistics</i>									
Multiple R	0.60914486								
R Square	0.37105746								
Adjusted R Square	0.0565862								
Standard Error	2.53522237								
Observations	4								
ANOVA									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	1	7.583895074	7.583895074	1.179941	0.390855138				
Residual	2	12.85470493	6.427352463						
Total	3	20.4386							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	
Intercept	4.50980138	1.847397635	2.441164424	0.134713	-3.438909094	12.4585119	-3.4389091	12.45851186	
16.86	-0.05230857	0.04815515	-1.086250776	0.390855	-0.259503455	0.15488632	-0.2595035	0.154886318	
SUMMARY OUTPUT RENGGAM FARM									
<i>Regression Statistics</i>									
Multiple R	0.6260781								
R Square	0.39197379								
Adjusted R Square	0.08796069								
Standard Error	3.19264971								
Observations	4								
ANOVA									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	1	13.14217572	13.14217572	1.289332	0.373921896				
Residual	2	20.38602428	10.19301214						
Total	3	33.5282							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	
Intercept	8.70168079	2.326460032	3.740309597	0.064628	-1.308268822	18.7116304	-1.3082688	18.71163039	
16.86	-0.06885894	0.060642619	-1.135487517	0.373922	-0.329783068	0.19206519	-0.3297831	0.192065194	
SUMMARY OUTPUT SIMPANG RENGGAM FARM									
<i>Regression Statistics</i>									
Multiple R	0.44710734								
R Square	0.19990497								
Adjusted R Square	-0.20014254								
Standard Error	3.26658497								
Observations	4								
ANOVA									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	1	5.33212032	5.33212032	0.499703	0.552892661				
Residual	2	21.34115468	10.67057734						
Total	3	26.673275							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	
Intercept	4.26154556	2.380336105	1.790312533	0.21529	-5.98021408	14.5033052	-5.9802141	14.5033052	
16.86	-0.04386081	0.062046979	-0.706896793	0.552893	-0.310827412	0.22310579	-0.3108274	0.223105792	

Table 8: Summary Output of Machap River Damsite

Regression Statistics									
Multiple R	0.81731114								
R Square	0.6679975								
Adjusted R Square	0.50199624								
Standard Error	2.62968887								
Observations	4								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	1	27.82737286	27.82737286	4.024051	0.182688863				
Residual	2	13.83052714	6.915263571						
Total	3	41.6579							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	8.7112997	1.916234672	4.54605056	0.045136	0.466407361	16.956192	0.46640736	16.95619204	
16.86	-0.10019886	0.049949489	-2.006003723	0.182689	-0.315114165	0.11471644	-0.3151142	0.114716444	

References

- [1] M. E. A. R. C. Moeletsi, Z. P. A. R. C. Shabalala, G. A. R. C. De Nysschen, M. E. Moeletsi, & Walker, S., "Evaluation of an inverse distance weighting method for patching daily and dekadal rainfall over the Free State Province, South Africa," *Water SA*, vol. 42, no.3, July, 2016. [Online serial]. Available: <https://doi.org/10.4314/wsa.v42i3.12>. [Accessed September 15, 2024].
- [2] A.N. Adilah & H. Hannani, "Comparison of methods to estimate missing rainfall data for short term period at UMP gambang," In *IOP Conference Series: Earth and Environmental Science*, vol 682, no. 1, Feb, 2021. [Online serial]. Available: <https://doi.org/10.1088/1755-1315/682/1/012027>. [Accessed September 15, 2024].
- [3] J.P. Hughes & P. Guttorp, "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena," *Water resources research*, vol. 30, no. 5, May, 1994. [Online serial]. Available: <https://doi.org/10.1029/93WR02983>. [Accessed September 15, 2024].
- [4] S.N. Burhanuddin, S.M. Deni & N.M. Ramli, "Imputation of missing rainfall data using revised normal ratio method," *Advanced Science Letters*, vol. 1, no. 23, November, 2017. [Online serial]. Available: <https://doi.org/10.1166/asl.2017.10203>. [Accessed September 15, 2024].
- [5] M. A. Azman, R. Zakaria & S. Z. Satari, "Simulation study of adjusted spatial weighting method to estimate missing rainfall data," *ASM Science Journal*, vol. 13, no. 1-7, 2020. [Online serial]. Available: [https://doi.org/10.32802/asmscj.2020.sm26\(1.20\)](https://doi.org/10.32802/asmscj.2020.sm26(1.20)). [Accessed September 15, 2024].
- [6] P. Luo, D. Mu, H. Xue, T. Ngo-Duc, K. Dang-Dinh, K. Takara, D. Nover & G. Schladow, "Flood inundation assessment for the Hanoi Central Area, Vietnam under historical and extreme rainfall conditions," *Scientific reports*, vol. 22, no. 8(1), August, 2018. [Online serial]. Available: <https://doi.org/10.1038/s41598-018-30024-5>. [Accessed September 15, 2024].
- [7] R.P. De Silva, N.D. Dayawansa & M.D. Ratnasiri, "A comparison of methods used in estimating missing rainfall data," *Journal of agricultural sciences*, vol. 1, no 3(2), May, 2007. [Online serial]. Available: <https://doi.org/10.4038/jas.v3i2.8107>. [Accessed September 15, 2024].
- [8] S. N. Z. Amin Burhanuddin, S. Mohd Deni & N. Mohamed Ramli, "Revised Normal Ratio Methods for Imputation of Missing Rainfall Data," *Scientific Research Journal*, vol. 13, no.1, 2016. [Online serial]. Available:<https://doi.org/10.24191/srj.v13i1.9384>. [Accessed September 15, 2024].
- [9] K.F. Nimon & F.L. Oswald, "Understanding the results of multiple linear regression: Beyond standardized regression coefficients," *Organizational Research Methods*, vol. 16, no. 4, October, 2013. . [Online serial]. Available: <https://doi.org/10.1177/1094428113493929>. [Accessed September 15, 2024].
- [10] R.M. Stolzenberg, "Multiple regression analysis, *Handbook of data analysis*, vol. 21, April, 2004. [Online serial]. Available: <https://doi.org/10.4135/9781848608184.n8>. [Accessed September 15, 2024].
- [11] X.L. Meng, R. Rosenthal & D.B. Rubin, "Comparing correlated correlation coefficients," *Psychological bulletin*, vol. 111, no. 1, January,1992. [Online serial]. Available: <https://doi.org/10.1213/ANE.000000000002864>. [Accessed September 15, 2024].

