Hourly Rainfall Simulation Using Daily Data

Suroso, Fatimatus Sholihah Marush, Purwanto Bekti Santoso, Irfan Sudono, Edvin Aldrian, and Nelly Florida Riama

Abstract Rainfall is one of the main inputs in the analysis of the hydrological system. However, rainfall data are often found to be inadequate in length and completeness. The purpose of this study is to overcome the problem of the limited availability of high resolution of rainfall data. Rainfall data on a daily scale generally have a better quantity and quality of data available to be used as a basis or reference for deriving hourly rainfall data. The research location is limited to Java Island, with 25 rain gauge stations used. Hourly rainfall has a diurnal cycle, where some rainfall is concentrated in the afternoon or evening. The result of the evaluation using Root Mean Square Error, Mean Absolute Error, Pearson Correlation, and Spearman Correlation shows that the hourly rain simulation model using daily rainfall data can be used for design, especially in the field of water structures due to producing similar statistical characteristics.

Suroso (✉) · F. S. Marush · P. B. Santoso
Department of Civil Engineering, Faculty of Engineering, Jenderal Soedirman University, Purbalingga 53371, Indonesia
e-mail: surosotsipil@unsoed.ac.id

P. B. Santoso
e-mail: purwanto1409@unsoed.ac.id

I. Sudono
Research Centre for Water Resources, Agency of Research and Development, Ministry of Public Works and Housing, Jakarta, Indonesia

E. Aldrian
Agency for the Assessment and Application of Technology (BPPT), Jakarta, Indonesia

N. F. Riama
Meteorology, Climatology, and Geophysical Agency (BMKG), Jakarta, Indonesia
e-mail: nelly.florida@bmkg.go.id

S. A. Kristiawan et al. (eds.), *Proceedings of the 5th International Conference on Rehabilitation and Maintenance in Civil Engineering*, Lecture Notes in Civil Engineering 225,
https://doi.org/10.1007/978-981-16-9348-9_86

1 Introduction

A flood is a pool of water that causes a detrimental impact on human life. A flood is a flow of water whose height exceeds the normal water level. Among all the disasters, floods are the most destructive because they can cause serious damage to human life, infrastructure, agriculture, and social-economic systems [1]. In Indonesia, almost all areas have a high potential for flooding such as West Java [2], Central Java [3], and East Java [4]. Flooding can take in many forms, including slow river flooding, rapid flash floods, accumulation of rainwater in poorly drained environments, and coastal flooding caused by tides and extreme waves.

Rainfall is one of the main inputs in the analysis of the hydrological system. However, rainfall data are often inadequate in length and completeness due to the small number of observations or instrument errors. Insufficient rainfall data series can cause severe problems in hydrological analysis. Accessibility of high-resolution rainfall is one of the main problems for hydrologists due to various reasons such as cost and government policies [5]. Therefore, scientists have attempted to produce accurate high-resolution rainfall data with many computer programs. The simulation of hourly rainfall is an important area of hydrological research, for example for designing hydraulic structures in urban areas or urban drainage systems, and for evaluating the frequency of flooding in small natural catchments characterized by a rapid hydrological response [6]. Hourly rainfall data or even smaller by temporal scale are critical for applications in many sectors, such as urban hydrology, infrastructure design, and risk assessment [7].

Daily rainfall is generally a difficult thing in rainfall modelling because daily rainfall has unique characteristics and is difficult to make a model. Unlike the annual and monthly rainfall, which can be modelled by the simple autoregressive moving average (ARMA) process, the incidence and amount of daily rainfall are challenging to model with the ARMA model [8]. There are many problems in hydrology, agricultural science and engineering that require extensive rainfall records [9], especially for flood risk assessment, water resources management or hydraulic structure design [10]. However, hourly or higher resolution rainfall data tend to have lower temporal and spatial coverage due to high instrument costs and technical limitations.

The purpose of this study is to overcome the problem of the limited availability of high-resolution rain data. High-resolution rainfall data in the hourly or higher (minute) is necessary for the hydrological aspect. Rainfall data on a daily scale generally have a better quantity and quality of data availability. Hence, daily rainfall can be used as a basis or reference for deriving hourly rainfall data.

2 Material and Method

2.1 Data

The data used were obtained from the Meteorology, Climatology, and Geophysical Agency (BMKG). The data required is hourly and daily rainfall data in Java. The location of selected rainfall stations is presented in Fig. 1, and the name of that can be seen in Table 1. The rainfall data used in this study are selected with having relatively complete data from all stations in Java. By knowing the number of NA (Not

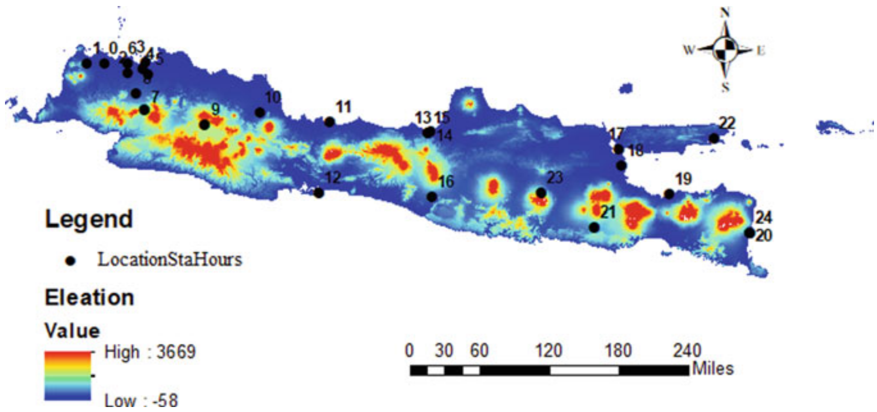


Fig. 1 Map of 25 rain station in Java Island

Table 1 The name of rain station in Java Island

Code	Rain station	ID rain station	Code	Rain station	ID rain station
0	Tangerang	96,735	13	Semarang_Klimatologi	96,835
1	Serang	96,737	14	Semarang_Maritim	96,837
2	Bawean	96,925	15	Ahmad Yani_Semarang	96,839
3	Curug	96,739	16	Surabaya Tg. Perak	96,937
4	Mataram	96,742	17	Surabaya_Perak1	96,933
5	Jakarta BMKG	96,745	18	Juanda	96,935
6	Cengkareng	96,749	19	Karang Ploso	96,943
7	Citeko	96,751	20	Tretes	96,945
8	Dermaga	96,753	21	KarangKates	96,949
9	Bandung	96,783	22	Kalianget_Madura	96,973
10	Jatiwangi_Cirebon	96,791	23	Sawahan	96,975
11	Tegal	96,797	24	Banyuwangi	96,987
12	Cilacap	96,805			

Available) or the countless data, there are three rain recording stations are obtained. The relatively complete rainfall recording stations are Citeko (96,751) and Dermaga (96,753).

2.2 Research Steps

Data Preparation

The following steps in preparing the data are the hourly rainfall recording stations that have a relatively long measurement period, complete and continuous are identified and selected for further analysis. Then at each hourly rainfall recording station that has been selected, the cumulative rainfall calculation is carried out into a daily period. Two data sets of rainfall are collected, so that hourly and daily rainfall data will be obtained at the same time.

Model Development

There are two types of modelling approaches were taken, namely the approach that the rainfall data series for both zero and non-zero rainfall was calculated and another approach that only non-zero rainfall data series were carried out in the modelling stage.

Model Evaluation

After obtaining the hourly rain simulation modelling using daily rainfall data, it is necessary to evaluate it to obtain the performance of the model. There are several measuring tools to find out how well the model developed consists of RMSE, MAE, Pearson Correlation, Spearman Correlation.

RMSE (Root Mean Square Error)

RMSE is a way to evaluate linear regression models by measuring the accuracy of the estimated results of a model. RMSE is calculated by squaring the error (prediction - observation) divided by the amount of data, then rooted. The RMSE formula is shown in Eq. (1). This prediction model is said to be the best if the value is 0 (zero).

$$\text{RMSE} = \left(\frac{\sum (y_i - \hat{y}_i)}{n} \right)^{1/2} \tag{1}$$

where y = the value of the observation result, ŷ = predicted value, i = sequence of data in the database, N = amount of data.

MAE (Mean Absolute Error)

MAE is one of the methods used to measure the accuracy of the forecasting model. The MAE value shows the absolute average error between the prediction results with the real value. In general, the MAE formula is shown in Eq. (2). This prediction model is said to be the best if the MAE value is 0 (zero).

Table 2 Class of correlation coefficient value

Class	Correlation coefficient value	Information
1	0–0.25	Very weak
2	0.25–0.50	Weak
3	0.50–0.75	Strong
4	0.75–1	Very strong

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (2)$$

where n = amount of data, $\left|y_i - \hat{y}_i\right|$ = absolute error, y_i = predicted value, $\hat{y}_i$ = real value.

Pearson and Spearman Correlation
Pearson correlation is used to determine whether there is a relationship between two variables on an interval or ratio scale. The Pearson correlation denoted by r is shown in Eq. (3).

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} \quad (3)$$

where r = correlation value, x = variable x, y = variable y. The strength of the correlation relationship classified into 4 classes that shown in Table 2.

The Spearman-rho rank correlation coefficient was used for the measurement of correlation in nonparametric statistics.

3 Result and Discussion

3.1 Analysis of the Quality of Hourly and Daily Rainfall Data

Completeness of rain data is an important aspect in knowing the quality of hourly and daily rainfall data. Rainfall data on the hourly and daily time scales used in this study were taken from 1981 to 2019. However, the available rain data, especially on the hourly time scale, is inadequate so that a lot of rain data is not recorded or is called NA (Not Available). From the 25 stations, the percentage of missing values (NA) that is less than 31% is the Dermaga station (96,753) and Citeko (96,751), that are 24.57 and 30.44%. Meanwhile, there are 4 stations that have rainfall data between 30–50%. The remaining 19 stations with more than 70% missing values. For the daily rainfall scale, the percentage of missing values (NA) for the 25 stations is not

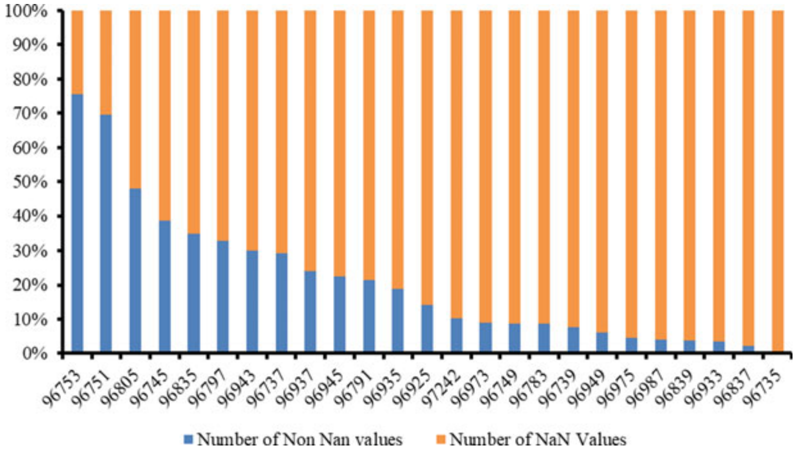


Fig. 2 Percentage of number of non-NA and NA at each station

much different from the hourly rain scale. The following graph shows the percentage of NA numbers at each station as in Fig. 2.

In another study, it was stated that in Malaysia, for the daily rainfall scale, the number of missing values was less than 10% for the period 1975–2004. The number of missing values is estimated using various types of weighting methods. Meanwhile, in Singapore, the daily time scale analyzed from 30 stations for the period 1980 to 2010 shows very few missing values data, namely less than 1%. However, the area used in Singapore is relatively small, around 716 km^2 compared to Java, which is 128.297 km^2.

3.2 Statistical Analysis of Hourly and Daily Rainfall Data

From 25 rain stations for the hourly rain scale, the largest average value of rainfall occurred at Semarang Maritim Rain Station (96,837) of 1.17 mm. Meanwhile, for the daily rain scale, there is only one rain station that has an average value below 1 mm, namely Banyuwangi Rain Station (96,987) of 0.06 mm. Then for the hourly scale of rain, the largest maximum amount of rain occurs at the Jatiwangi Cirebon Rain Station (96,791) of 147 mm which occurs precisely at 17.00 on 25 December 2013. For the daily rain scale, the Cilacap rain station (96,805) has the largest maximum amount of rain, which is 302.7 mm. For the minimum amount of rain in the hourly and daily rain scale at each rain station, it is 0 mm, which means that each rain station has dry hours. Furthermore, the standard deviation value for the largest hourly rain scale occurred at Semarang Maritim Rain Station (96,837) of 4.08. Meanwhile, the largest standard deviation value for the daily rain scale occurred at the Cilacap Rain

Station (96,805), which was 19.56. Statistical analysis for 25 rain stations can be seen in Tables 3 and 4.

Diurnal cycles and regional variations in rainfall are essential in the tropics. Because tropical rainfall involves the uptake and release of large amounts of latent heat from evaporation, in the study of energy and water cycles on a large scale such as the Asian monsoon, an understanding of the diurnal rainfall cycle is needed. Most studies have shown that diurnal variation in precipitation in the Tropics peaks in the afternoon overland areas and the morning over adjacent ocean areas (e.g. [11–14]).

Table 3 Statistical analysis of hourly rainfall data

Rain station	Average (mm)	Maximum (mm)	Minimum (mm)	Standard deviation
Tangerang	0.69	55	0	3.16
Serang	0.18	85	0	1.61
Curug	0.31	80	0	2.57
Jakarta BMKG	0.24	82	0	2.15
Cengkareng	0.15	73.9	0	1.69
Citeko	0.37	100	0	2.21
Dermaga	0.41	107	0	2.96
Bandung	0.3	67	0	2.09
Jatiwangi_Cirebon	0.27	147	0	2.46
Tegal	0.18	99.7	0	1.86
Cilacap	0.35	111.2	0	2.56
Semarang_Klimatologi	0.37	94	0	2.71
Semarang_Maritim	1.17	63	0	4.08
Ahmad Yani_Semarang	0.64	85	0	3.37
Surabaya_Perak1	0.23	99	0	2.3
Juanda	0.26	80	0	2.31
Karang Ploso	0.21	77	0	1.89
Tretes	0.8	99	0	4.03
KarangKates	0.62	61.5	0	3.17
Kalianget_Madura	0.18	100	0	1.84
Sawahan	0.53	77.2	0	3.21
Banyuwangi	0.36	74	0	2.5
Mataram	0.18	70	0	1.86
Bawean	0.32	99	0	2.51
Surabaya Tg. Perak	0.21	91.1	0	2.01

Table 4 Statistical analysis of daily rainfall data

Rain station	Average (mm)	Maximum (mm)	Minimum (mm)	Standard deviation
Tangerang	2.13	67.8	0	8.64
Serang	4.33	101.4	0	10.41
Curug	7.55	187.8	0	17.07
Jakarta BMKG	5.68	277.5	0	15.47
Cengkareng	3.58	125.2	0	11.72
Citeko	8.91	245.3	0	15.98
Dermaga	9.7	240	0	18.58
Bandung	7.28	108.4	0	13.39
Jatiwangi_Cirebon	6.6	157.5	0	15.44
Tegal	4.38	140	0	12.15
Cilacap	8.27	302.7	0	19.56
Semarang_Klimatologi	8.78	245.2	0	17.86
Semarang_Maritim	7.14	100	0	16.26
Ahmad Yani_Semarang	1.99	54.5	0	6.78
Surabaya_Perak1	4.31	187	0	14.78
Juanda	5.73	118.2	0	14.15
Karang Ploso	5.11	152	0	12.63
Tretes	7.27	187.4	0	18.35
KarangKates	5.26	91.9	0	11.6
Kalianget_Madura	4.31	157.9	0	11.79
Sawahan	9.96	176	0	18.67
Banyuwangi	0.06	13.3	0	0.71
Mataram	4.4	164	0	12.78
Bawean	7.09	221.2	0	16.53
Surabaya Tg. Perak	4.72	130	0	12.36

3.3 Daily Cycle Pattern of Hourly Rainfall

In this study, two seasons were used to determine the daily rain pattern at selected stations, namely the wet and dry seasons. Based on the rain event in Java, it is chosen for the wet season to occur from October to March and the dry season occurs from April to September. Furthermore, a rain station that has the least amount of NA (Non-Available) is selected to determine the daily rainfall pattern based on hourly rainfall data. The selected rain stations are the Dermaga rain station (96,753) and Citeko rain station (96,751). The results of the daily rainfall patterns for the two stations in the two predetermined seasons can be seen in Fig. 3. The boxplot for the daily rain pattern at the Citeko rain station and Dermaga in the wet season and dry season have

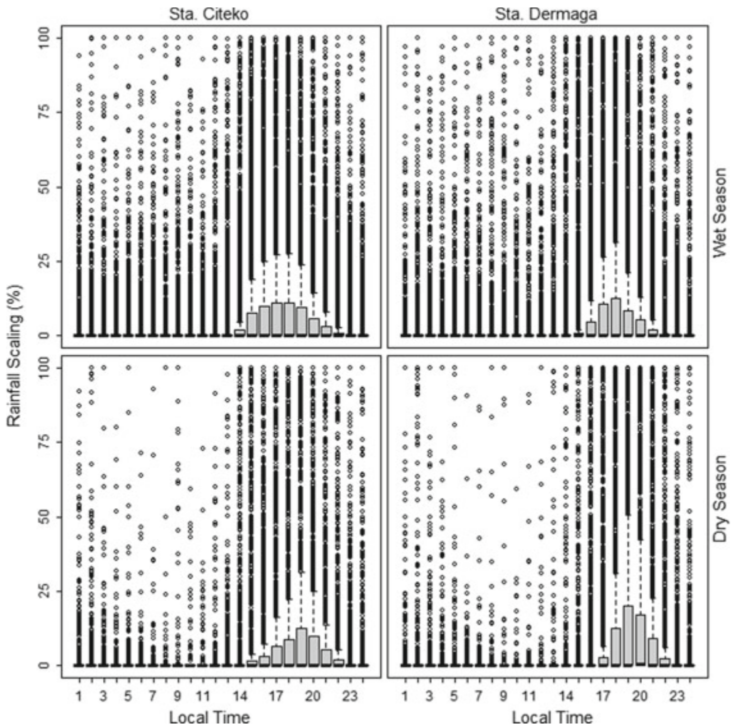


Fig. 3 Boxplot for the daily rain pattern at Citeko and Dermaga rain station

many outlier values, which means that there is a lot of hourly rain data which has very high values differ greatly from other data.

3.4 Simulation Model of Hourly Rainfall

In this research, to develop a simulation model, the assumption is that for rain below 0.1 mm in an hourly period, it is considered 0 mm. The results of the daily rain pattern in Fig. 4 shows that the daily rain has the same pattern for both the Citeko and Dermaga rain stations in two seasons, the wet and dry seasons. Furthermore, a daily rainfall model is made by considering the mean hourly rainfall as shown in Fig. 4. The results of the development of a simulation model from known patterns show that intermittent rainfall has a daily cycle pattern; that is, some concentrated rain occurs in the afternoon or at night. This is because morning to noon has evaporated and formed clouds. So that in the afternoon or evening, the clouds begin to condense so that it rains.

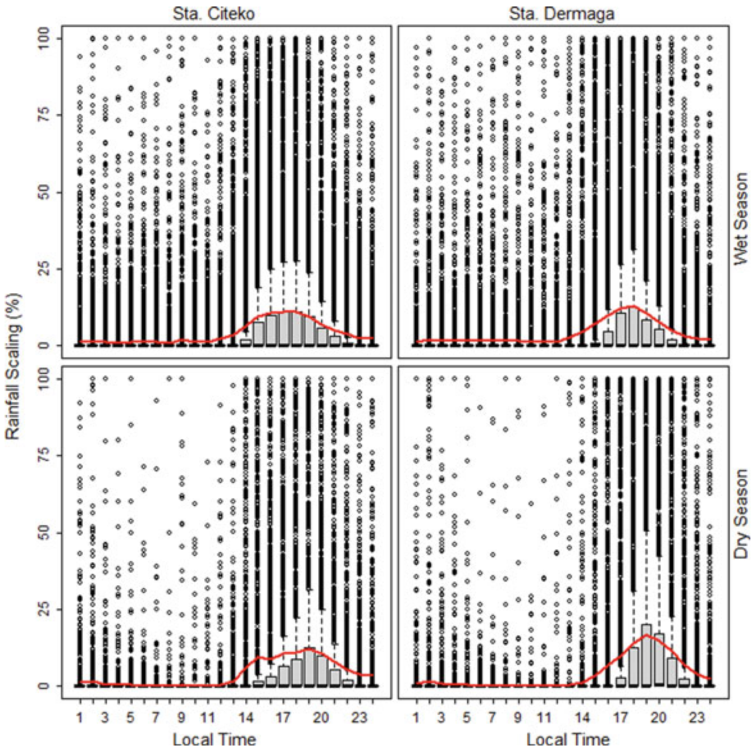


Fig. 4 Daily rain modeling for Citeko and Dermaga rain station

3.5 Model Evaluation

After obtaining the hourly rain simulation modelling using the daily rainfall data, it is necessary to evaluate it to obtain the performance of the model. Evaluation based on the mean value for simulated and observed rain can be seen in Fig. 5. The mean value for the simulated and observed rain has quite a small difference. For the Citeko rain station, the average difference in the mean value for simulated rain and observations in the wet season is 0.193 mm and for the dry season 0.185 mm. Then for Dermaga rain station, in the wet season the average difference in the mean value is 0.214 mm and for the dry season is 0.317 mm.

The next model evaluation is MAE and RMSE. The two prediction models are said to be the best if the value is 0 (zero). For the Citeko rain station in the wet season, the average of MAE value is 1.05 mm, and for the dry season 0.85 mm. Furthermore, the average of RMSE value in the wet season is 2.56 mm and for the dry season is 2.12 mm. For the Dermaga rain station in the wet season, the average of MAE value is 1.17 mm and for the dry is 1.25 mm. Furthermore, the average of RMSE value in the wet season is 2.96 mm and for the dry season is 2.98 mm. From Fig. 6 shows that the pattern of MAE and RMSE values is close to the rain pattern simulations that have been made. The average error value is both above 0 but not too big. The average value for the Citeko rain station tends to be smaller than Dermaga rain station.

The last model evaluation is seen from the value of the Pearson and Spearman correlation coefficient. The average value of the Pearson correlation coefficient for the Citeko rain station in the wet season is 0.29 which indicates that entering into class 2 is sufficiently correlated. Meanwhile, the dry season falls into class 1 with an average of 0.21 which indicates that the correlation is very weak. The Spearman correlation coefficient for the Citeko rain station is smaller than the Pearson correlation. In the wet season, the average coefficient value is 0.24 and 0.15 in the dry season indicates that it is classified as class 1. For the Dermaga rain station, the value of the Pearson and Spearman correlation coefficient in the wet and dry season falls into class 1. For the Pearson correlation in the wet season, the average value is 0.21 and 0.17 in the dry season. Furthermore, the average Spearman correlation value in the wet season is 0.14 and 0.12 in the dry season. Figure 7 shows that the pattern of the Pearson and Spearman correlation coefficient values is close to the rain pattern simulations that have been made. However, the correlation coefficient for both of them is quite small, so it can be concluded that the evaluation of rain simulations and observations has a very weak correlation. For this reason, further modelling is needed if it is to be used for forecasting rain data.

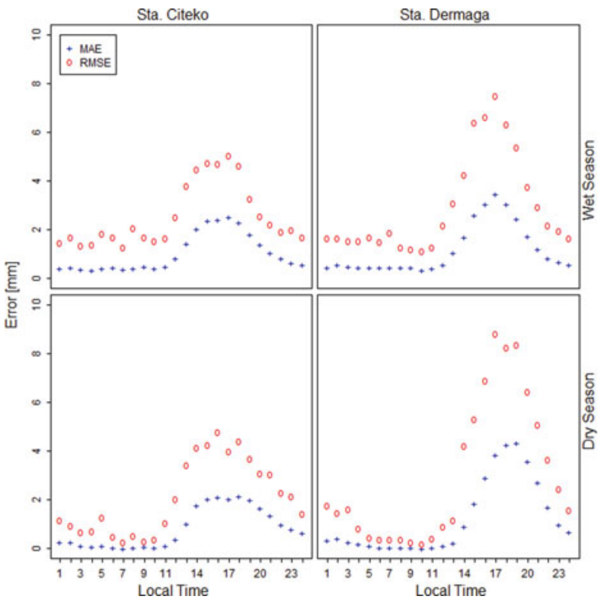
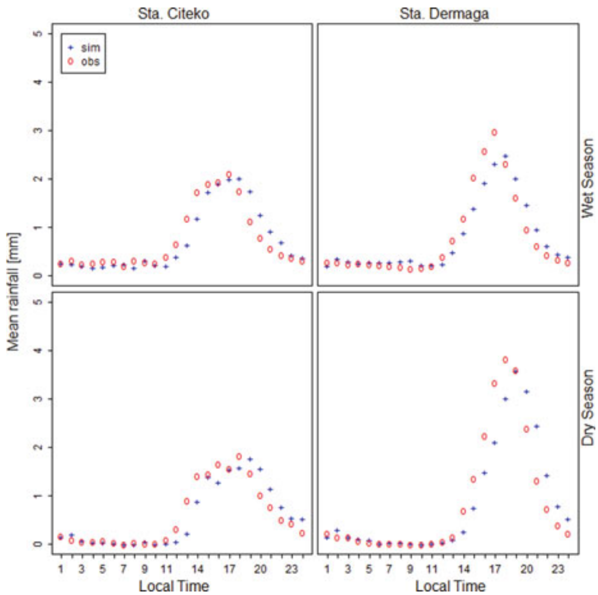


Fig. 5 Comparison of mean value of simulation and observation rain

Fig. 6 Comparison of MAE and RMSE values



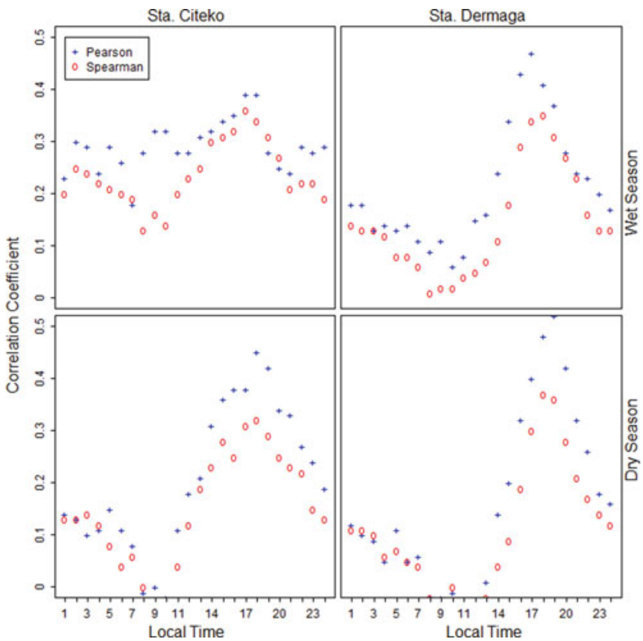


Fig. 7 Comparison of Pearson and spearman correlation coefficient value

4 Conclusion

Based on the quality of hourly and daily rainfall data, it shows that the quality of rain at 25 rain stations in Java Island is not good so that only two rain stations were selected with the least amount of NA (Not Available) data, that are Citeko and Dermaga. The results of the development of a simulation model show that the hourly rainfall has a daily cycle pattern, that is some concentrated rain occurs in the afternoon or at night. The mean values for simulated and observed rain at the Citeko and Dermaga rain stations in both the wet and dry seasons have quite small differences. The average error (MAE and RMSE) value of the two rain stations is above 0 but not too big. The average error value for the Citeko rain station smaller than Dermaga. The value of the Pearson and Spearman correlation coefficient for the two rain stations is quite small. The evaluation of simulated and observed rain has a very weak correlation. Hourly rain simulation modelling using daily rainfall data can be used for design, especially in the field of water construction. However, for forecasting purposes, further modelling is required.

Acknowledgements This research is part of a strategic research collaboration between the Indonesian government and the UK government with the title "Extreme Rainfall and Its Impact on Flood Risk in Indonesia".

References

1. Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. Water 10(11):1–40
2. Jati MIH, Suroso S, Santoso PB (2019) Prediction of flood areas using the logistic regression method (a case study of the provinces Banten, DKI Jakarta, and West Java). J Phys Conf Ser 1367(1):012087 (IOP Publishing)
3. Pratidina G, Suroso S, Santoso PB (2019) Detection of satellite data-based flood-prone areas using logistic regression in the central part of Java Island. J Phys Conf Ser 1367(1):012086 (IOP Publishing)
4. Suroso S, Ghifari IA, Santoso PB (2021) Detection of trend behaviour of extreme rainfall over Java using Mann-Kendall. In: IOP conference series: earth and environmental science, vol 698, issue No. 1. IOP Publishing, p 012010
5. Bakhshi M, Al Janabi F (2019) Disaggregation the Daily rainfall dataset into sub-daily resolution in the temperate oceanic climate region. Int J Mar Environ Sci 13(1):11–16
6. Brigandì G, Aronica GT (2019) Generation of sub-hourly rainfall events through a point stochastic rainfall model. Geosciences 9(5):11–14
7. Diez-Sierra J, del Jesus M (2019) Subdaily rainfall estimation through daily rainfall downscaling using random forests in Spain. Water 11(1).https://doi.org/10.3390/w11010125
8. Suroso S, Bárdossy A (2018) Multisite daily precipitation simulation in Singapore. MATEC Web of Conferences 195:05007 (EDP Sciences)
9. Mehrotra R, Li J, Westra S, Sharma A (2015) A programming tool to generate multi-site daily rainfall using a two-stage semi-parametric model. Environ Model Softw 63:230–239
10. Vandenberghe S, Verhoest NEC, Onof C, De Baets B (2011) A comparative copula-based bivariate frequency analysis of observed and simulated storm events: a case study on Bartlett-Lewis modeled rainfall. Water Resour Res 47(7)

11. Gray WM, Jacobson RW Jr (1977) Diurnal variation of deep cumulus convection. Mon Weather Rev 105(9):1171–1188
12. Murakami M (1983) Analysis of the deep convective activity over the western Pacific and Southeast Asia Part I: Diurnal variation. J Meteorol Soc Jpn Ser II 61(1):60–76
13. Mori S, Jun-Ichi H, Tauhid YI, Yamanaka MD, Okamoto N, Murata F, Sribimawati T (2004) Diurnal land-sea rainfall peak migration over Sumatera Island, Indonesian Maritime Continent, Observed by TRMM satellite and intensive rawinsonde soundings. Mon Weather Rev 132(8):2021–2039
14. Sakurai N, Murata F, Yamanaka MD, Mori S, Hamada JI, Hashiguchi H, Suhardi B (2005) Diurnal cycle of cloud system migration over Sumatera Island. J Meteorol Soc Jpn Ser II 83(5):835–850