environmental and earth sciences proceedings

MDPI

Proceeding Paper

Assessment of Moving Average (MA) Method for Rainfall Prediction in Yogyakarta, Indonesia †

Nur Ain Jamal [1], Norazian Mohamed Noor [1,2,*], Izzati Amani Mohd Jafri [1] and Gurawan Jati Wibowo [3]

1 Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Arau 02600, Perlis, Malaysia; izzatiamani@studentmail.unimap.edu.my (I.A.M.J.)
2 Centre of Excellence Geopolymer and Green Technology (CEGeoGTech), Universiti Malaysia Perlis, Arau 02600, Perlis, Malaysia
3 Civil Engineering Department, Universitas Muhamamdiyah Surakarta, Jl. Kebangkitan Nasional No.101, Penumping, Kec. Laweyan, Kota Surakarta 57141, Jawa Tengah, Indonesia; gurawan.jati@ums.ac.id
* Correspondence: norazian@unimap.edu.my
† Presented at the 6th International Conference on Green Environmental Engineering and Technology (IConGEET2024), Bali, Indonesia, 29–30 August 2024.

Abstract: This study investigated a time series model using a moving average (MA) for predicting rainfall trend in seven areas (Panggang, Gedangan, Kedung Keris, Ngawen, Wanagama, Tepus, and Playen) located in the Gunung Kidul Province, Yogyakarta, Indonesia. A database with daily rainfall data covering the period of 2010–2019 obtained from the Central Statistical Body of Gunungkidul District (BPS), Indonesia, were analysed. In this study, the MA was developed using Microsoft Excel. Six performance indicators, including Root Mean Square Error (RMSE) and Index of Agreement (IA), were used to evaluate the goodness-of-fit of the time series model. The results specify that the MA is a reliable model, with the RMSE and IA values in the ranges of 10.7–18.3 and 0.80–0.85, respectively. Small error and high agreement rates between the observed and predicted values indicates that prediction using MA method has good potential to be used as one of the prediction tools for rainfall modelling.

Keywords: rainfall prediction; time series model; moving average

Academic Editor: Sara Yasina Yusuf

Published: 6 March 2025

Citation: Jamal, N.A.; Mohamed Noor, N.; Jafri, I.A.M.; Wibowo, G.J. Assessment of Moving Average (MA) Method for Rainfall Prediction in Yogyakarta, Indonesia. *Environ. Earth Sci. Proc.* **2025**, *33*, 5. https://doi.org/10.3390/eesp2025033005

1. Introduction

The climate in Indonesia is almost entirely tropical, with coastal plains averaging 28 °C, inland and mountain areas averaging 26 °C, and higher mountain regions averaging 23 °C. The relative humidity in the area is fairly high, ranging between 70% and 90% [1]. Monsoons usually blow in from the south and east from June to September, as well as from the northwest from December to March. Winds are moderate and generally predictable. In Indonesian waters, typhoons and large-scale storms pose little risk to mariners. Indonesia has a variety of climates, with the tropical rainforest receiving the most precipitation, followed by the tropical monsoon and tropical savanna receiving the least.

Due to large rainfall events that occur throughout the year, Indonesia is extremely prone to flood disasters. Flooding is classified as the most severe natural hazard due to its potential to cause significant societal, economic, and human losses. Throughout the years, the vast amount of rainfall in Indonesia serves as a key atmospheric heat source for the Earth's climate system [2–4]. This natural characteristic makes Indonesia particularly vulnerable to rainfall-related disasters, with floods being the most common hydrometeorological calamity. With increasing rainfall intensity, the frequency of flood disasters is growing year after year.

Geographically, Yogyakarta is often regarded as the primary gateway to Central Java, extending from Mount Merapi to the Indian Ocean. Its strategic location positions it as a hub for tourism and economic activities in the Java region. Excessive precipitation frequently causes floods in several areas of Yogyakarta. Torrential rains, lasting around two hours with intensities comparable to normal monthly rainfall, can cause significant flooding, affecting local residents and damaging public infrastructure [5]. Despite the increasing frequency and intensity of such events, there has been a lack of research focused on the performance of rainfall models and trends as alternative solutions for predicting extreme rainfall, particularly in Yogyakarta.

Rainfall modelling is important because it can describe the relationship between rainfall at a given location and other weather-related variables, such as large-scale climate variables and rainfall observed at other nearby locations. Other than that, rainfall models help to reduce unexplained variation in rainfall amounts, and it provides a principled way to quantify the uncertainty that comes with rainfall processes, which is critical to the efficient design of insurance contracts.

Time series analysis (TSA) is widely used for forecasting in various fields, such as weather prediction including rainfall prediction. The most common time series models used are auto regression (AR), moving average (MA), auto regression moving average (ARMA) or auto regression integrated moving average (ARIMA). Barman et al. [6] compared a few TSA including AR, MA, and ARMA for the time series forecasting of rainfall in India. As compared to the AR and MA models, the outcome showed that the ARMA model worked better. Saputro et al. [7] compared two methods, i.e., MA and Kriging methods, for rainfall predictions to fill in some incomplete data over several months. The results showed that the MA method provided a more accurate reflection of the data's true nature when a shorter time range was used. Conversely, a longer time range produced an average that accounted for a larger set of data points and captured more varied fluctuations. The results indicated that the moving average method outperformed the Kriging method. Thus, this study aimed to investigate the effectiveness of using the moving average (MA) method, a time series model, for the prediction of daily rainfall in the selected province in Yogjakarta, Indonesia.

2. Materials and Methods

This research focused on the selected area of the Gunung Kidul Regency, Yogyakarta, Indonesia, and involved data from seven rainfall monitoring stations located in Panggang, Gedangan, Kedung Keris, Ngawen, Wanagama, Tepus, and Playen as tabulated in Table 1. This research used the daily rainfall data from 2010 to 2019, which was obtained from the Central Statistical Body of Gunungkidul District (BPS), Indonesia.

Table 1. Rainfall stations in Gunung Kidul, Yogyakarta.

No	Stations	Coordinate	
		Latitude	**Longitude**
1	Kedung Keris	110.5946	−7.86922
2	Panggang	110.4241	−8.01411
3	Ngawen	110.7046	−7.81578
4	Playen	110.5188	−7.90919
5	Tepus	110.6258	−8.1055
6	Wanagama	110.5308	−7.89597
7	Gedangan	110.5809	−7.80728

2.1. Time Series Model

The moving average (MA) is one of the popular approaches for determining a time series trend. It has mostly been used to eliminate or reduce randomness in a time series [8]. In this study, Microsoft Excel was used to develop rainfall forecasts for the study area.

In the first step of the analysis, the dataset was divided into two parts which were the model training data and the validation data. The model training data, which consisted of 90% of the data from 2010 to 2018, were used to determine the optimal moving average period for the model. The remaining 10% of the data, representing the year 2019, were used as observed data to validate the performance of the models. This approach ensured that the model was evaluated on an unseen dataset, providing an objective assessment of the accuracy of the model.

In the second step, the MA model was developed where each observation was given equal weight. The model forecasted future values by averaging a specified number of past observations. In mathematical terms, if there is a time series of data points $(X_1, X_2, X_3, \ldots, X_t)$, the model forecasts future values $(\mathrm{X}_t + \mathrm{k})$ as follows:

$$St = Average(X_{t-k+1}, X_{t-k+2}, \ldots, X_t) \quad (1)$$

where St is the moving average (MA) smoothing, k represents the smoothing factor, and t is the time period, ranging from k to N (the total number of observations). In this case, k is the number of past observations used for the averaging process. The parameter value range for the smoothing factor (k) was set between 2 and $t-1$ using the XLMiner Analysis Toolpak Add-on. This MA method smooths out short-term fluctuations in the data, providing a clearer view of the underlying trend. Although simple, this approach effectively captures general trends in the data, making it a practical choice for short-term rainfall prediction [9].

2.2. Model Evaluation

The time series model of daily rainfall prediction in the Gunung Kidul Regency were evaluated using several performance indicators, namely Mean Absolute Error (MAE), Normalised Absolute Error (NAE), Root Mean Squared Error (RMSE), Index of Argument (IA), Prediction Accuracy (PA), and the Coefficient of Determination (R^2). The performance indicators were calculated using Microsoft Excel. Table 2 shows the formulas used to calculate the performance of the models.

Table 2. Performance Indicators [10].

Performance Indicators	Equation	Better Predictability if
Mean Absolute Error (MAE)	$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \lvert P_i - O_i \rvert}{n}$ (2)	Closer to 0
Normalised Absolute Error (NAE)	$\mathrm{NAE} = \frac{\sum_{i=1}^{n} \mathrm{Abs}(P_i - O_i)}{\sum_{i=1}^{n} O_i}$ (3)	Closer to 0
Root Mean Square Error (RMSE)	$\mathrm{RMSE} = \frac{1}{N}\sum_{i=1}^{N} \lvert P_i - O_i \rvert$ (4)	Closer to 0
Index of Agreement (IA)	$\mathrm{IA} = 1 - \left[\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}\left(\lvert P_i - \overline{O}\rvert + \lvert O_i - \overline{O}\rvert\right)^2}\right]$ (5)	Closer to 1
Prediction Accuracy (PA)	$\mathrm{PA} = \frac{\sum_{i=1}^{n}(P_i - \overline{O})^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}$ (6)	Closer to 1
Coefficient of Determination (R^2)	$R^2 = \left(\frac{\sum_{i=1}^{n}(P_i - \overline{P})(O_i - \overline{O})}{n.S_{pred}.S_{obs}}\right)^2$ (7)	Closer to 1

Where n = Total number of annual measurements of a particular site; P_i = Predicted values of one set annual monitoring record; O_i = Observed values of one set annual monitoring record; $\overline{P}$ = Mean of the predicted values of one set annual monitoring record; $\overline{O}$ = Mean of the observed values of one set annual monitoring record; S_{pred} = Standard Deviation of the predicted values of one set annual monitoring record; S_{obs} = Standard deviation of the observed values of one set annual monitoring record between input and output vectors.

3. Results and Discussion

Table 3 presents the data summary of rainfall intensity for the seven monitoring stations from 2010 to 2019. Four stations (Panggang, Gedangan, Ngawen, Wanagama) had no missing rainfall measurements, except for the Kedung Keris, Tepus, and Playen monitoring stations. The highest mean was recorded in Wanagama (5.673 mm), whereas the lowest mean rainfall intensity was observed in Tepus (4.316 mm). All the stations recorded a 0.00 mm value of median, showing that there was still no rain at the 50th percentile of the ranked rainfall measurements thus indicating that rainfall occurred less than 50% of the time throughout these 10 years.

Table 3. Descriptive statistics.

	Panggang	Gedangan	Kedung Keris	Ngawen	Wanagama	Tepus	Playen
Valid (frequency)	3650	3650	3285	3650	3650	2921	2190
Missing (frequency)	0	0	365	0	0	729	1460
Mean (mm)	5.418	5.136	5.609	4.355	5.673	4.316	5.636
Median (mm)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Standard Deviation (mm)	15.012	12.596	14.376	10.797	14.319	14.094	24.936
Minimum (mm)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum (mm)	285.20	265.50	277.60	190.70	264.20	211.80	560.00

The rainfall analysis was extended by developing a prediction model using the MA method for all seven selected stations. Table 4 shows the performance of the model for all the selected stations in comparison with two studies from India [6] and Iran [11], respectively. The Ngawen station indicated the smallest error rate compared to the others, with the lowest values of MAE, NAE, and RMSE (MAE = 7.115 mm: NAE = 0.498: RMSE = 10.721 mm). The highest agreement between the predicted and observed values was detected at the Tepus monitoring station, with the highest values of PA and IA (PA = 0.75: IA = 0.85). The Panggang, Wanagama, and Kedung Keris monitoring stations exhibited higher error rates, leading to less accurate predictions. Although the prediction accuracy was relatively higher, the increased error rate indicated lower overall reliability compared to the other stations. In comparison with Barman et al. [6], all the study areas (in this study) have significantly smaller error rates than the MA predicted values in [6]. Barman et al. [6] predicted rainfall intensity in two areas in India (Assam and Meghalaya) using a long-term monthly dataset (1900–2017). In this study, the MA smoothing factor (k) used was 1, whereas in this study, $k = 2$. In contrast, Mehdizadeh [11] proposed a hybrid model (combination of MA and multivariate adaptive regression splines) for predicting rainfall intensity in Iran, and it resulted in better prediction than using MA alone (similar to the results in this study). Overall, the MA method predicts the rainfall intensity accurately. The IA values for all the study areas were almost more than 0.8, indicating very good prediction accuracy that mimics the observed dataset. Besides that, the small MAE values calculated also indicate good prediction accuracy.

Model validation using the Coefficient of Determination (R^2) was supported by the scatterplot of the correlation between the observed and predicted rainfall data, as shown in Figure 1. From the R^2 values, it can be said that the predicted values were moderately in agreement with the observed values. The range of R^2 for all areas were from 0.47 to 0.56, with Tepus giving the highest R^2 value and Wanagama the lowest.

Table 4. Performance Indicators.

PI	Panggang	Gedangan	Kedung Keris	Ngawen	Wanagama	Tepus	Playen	Barman et al. [6]	Mehdizadeh [11]
MAE (mm)	10.475	8.772	12.211	7.115	9.702	8.350	9.595	121.97	5.01
NAE	0.552	0.587	0.580	0.498	0.603	0.546	0.527	-	-
RMSE (mm)	18.278	13.748	17.420	10.721	14.023	14.291	15.135	150.41	6.58
PA	0.709	0.695	0.692	0.694	0.683	0.750	0.745	-	-
IA	0.806	0.803	0.799	0.801	0.794	0.846	0.840	-	-

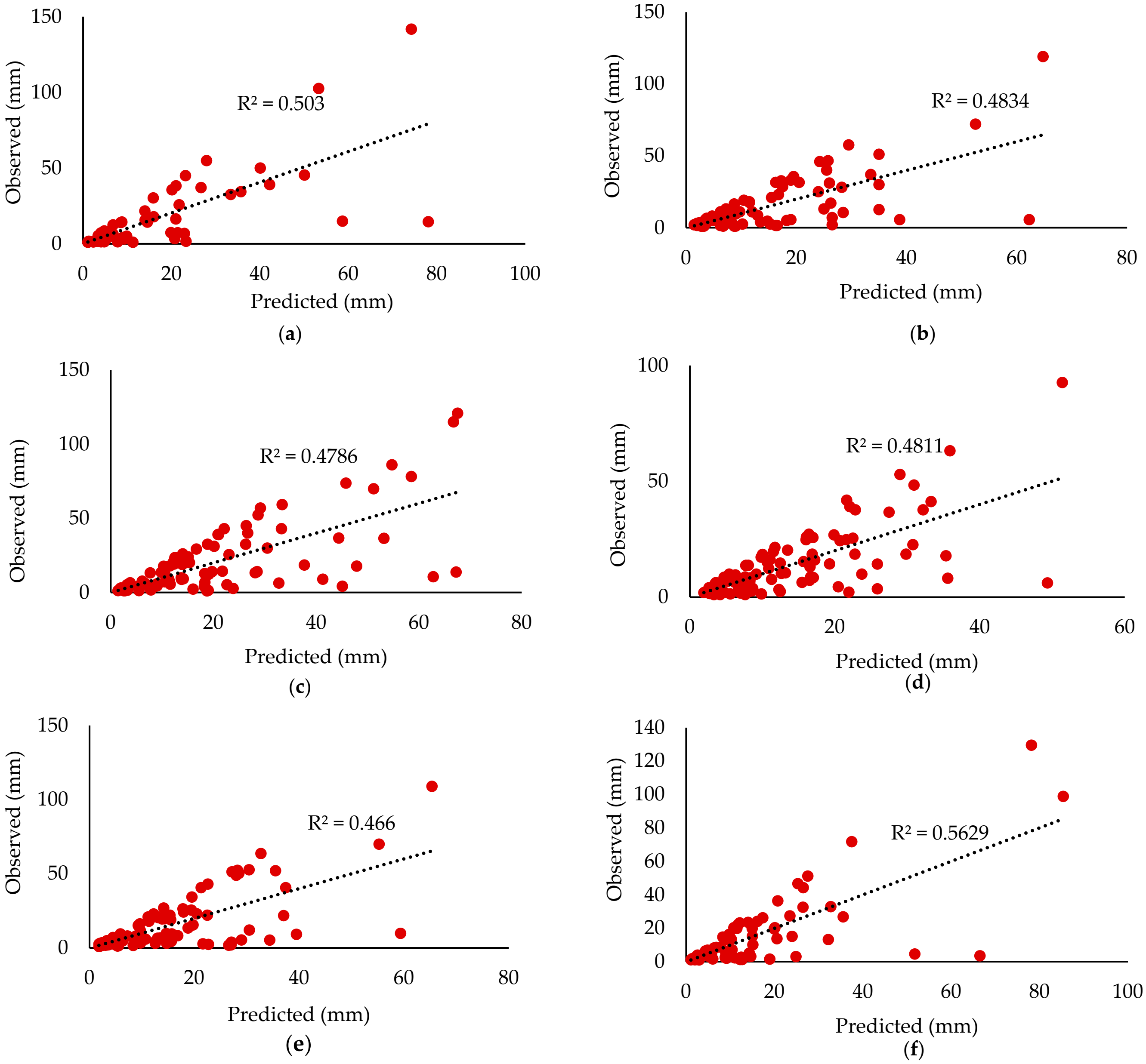

Figure 1. *Cont.*

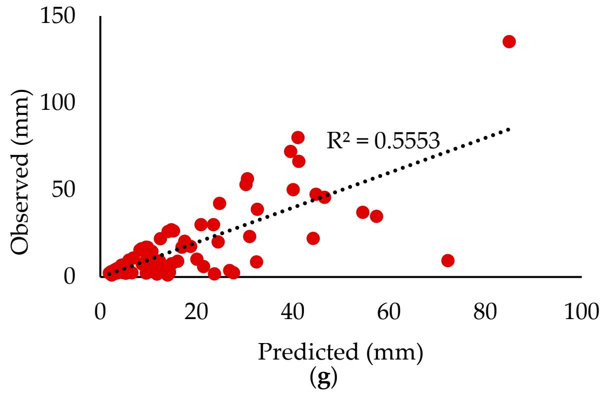

Figure 1. Scatterplot of observed and predicted values of rainfall intensity at; (**a**) Panggang; (**b**) Gedangan; (**c**) Kedung Keris; (**d**) Ngawen; (**e**) Wanagama; (**f**) Tepus; and (**g**) Playen.

Validation of the ability of the proposed moving average model for rainfall prediction was performed to indicate the strength or the weakness of the model. The evaluation of daily rainfall for one year—the data in 2019—was carried out for the measured and predicted data using the MA, as shown in Figure 2. Overall, it can be seen that the trend of the predicted values mimics the trend of the observed values. However, the trend of the predicted values underfitted the observed values for all the places, although the predicted values imitated the trend of the observed values very well.

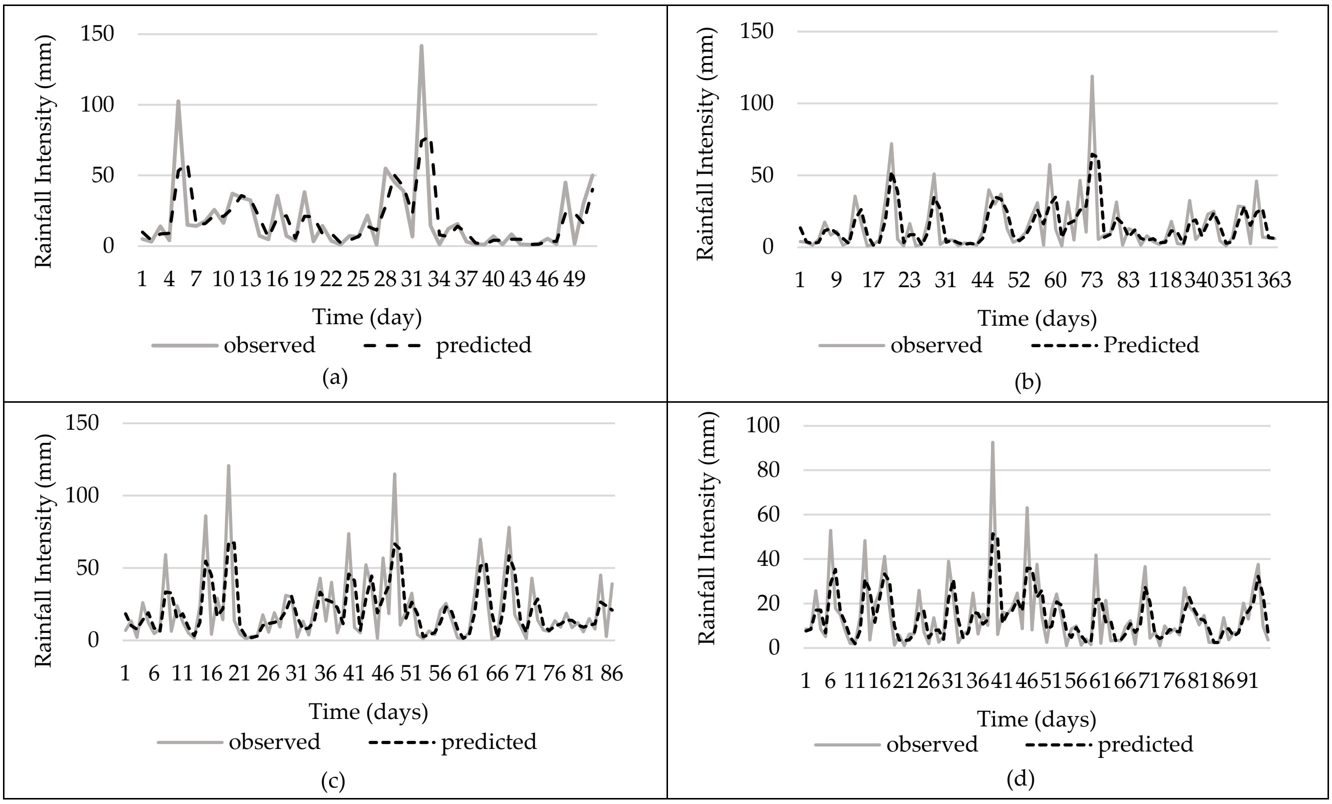

Figure 2. *Cont.*

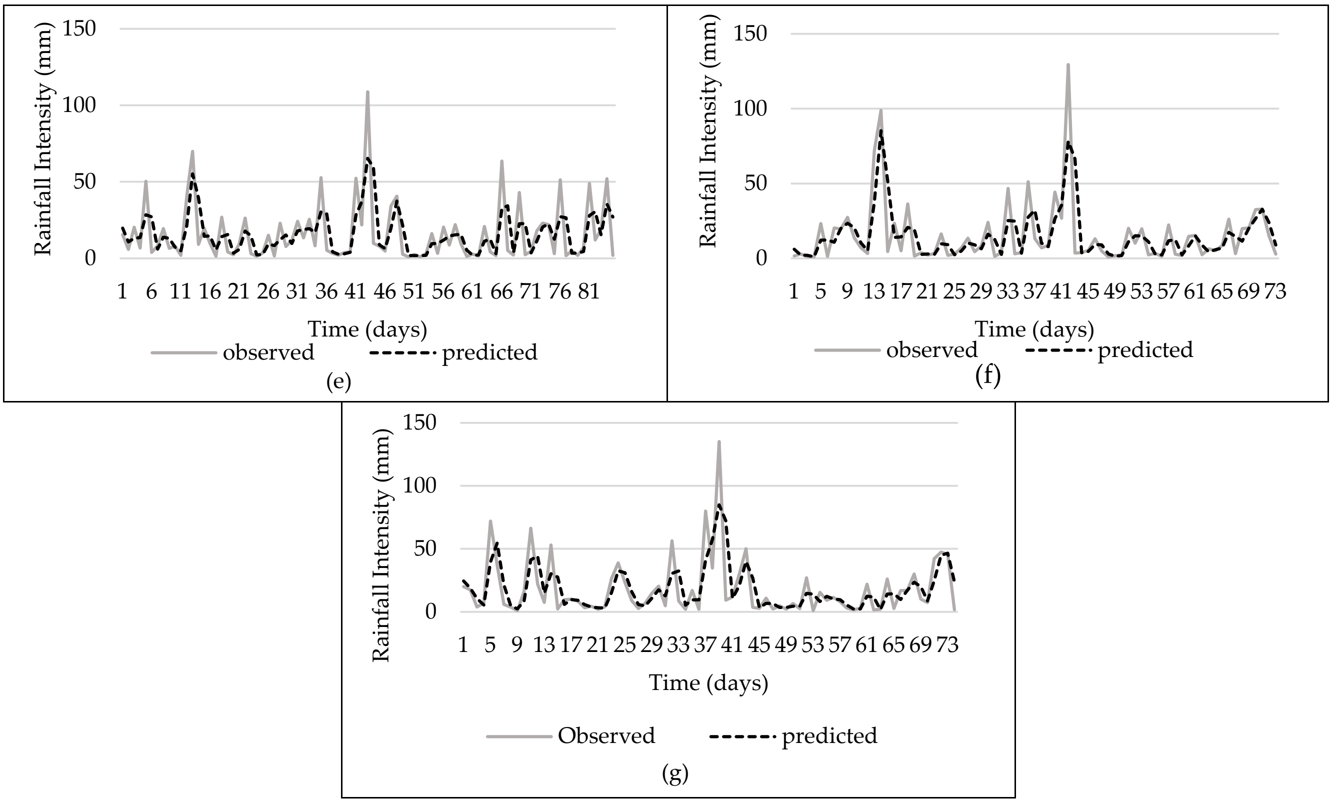

Figure 2. Rainfall Prediction for; (**a**) Panggang; (**b**) Gedangan; (**c**) Kedung Keris; (**d**) Ngawen; (**e**) Wanagama; (**f**) Tepus; (**g**) Playen.

4. Conclusions

In this study, a time series model involving the moving average (MA) was developed to predict daily rain intensity in the Yogjakarta province, Indonesia. A dataset encompassing ten years of daily rainfall data in seven areas in the Yogjakarta province was used for analysis. Six performance indicators (MAE, NAE, RMSE, PA, R^2, and IA) were used to evaluate the predicted values compared to the observed values. The results show that MA is a reliable model that can used to predict the rainfall intensity in the study area, since the performance indicators indicated good agreement between the predicted and observed values. In addition, the predicted rainfall trend had a similar trend with the original data, which shows that the model is suitable to forecast rainfall. To further improve accuracy of the daily rainfall prediction, ARMA and ARIMA can be used as other alternatives of time series models to predict the rainfall intensity. In addition, different models such as multivariate analyses, which includes the multiple linear regression (MLR) and the artificial neural network (ANN) techniques. A modified model can also be used in order to predict rainfall in the future to improve the accuracy of the basic model. A combination model that can be used is a combination of principle component analysis with MLR and ANN. Other parameters such as the water level, stream flow, and water quality can be used with the rainfall dataset in order to enhance the accuracy of the prediction model.

Author Contributions: Conceptualization, N.M.N. and G.J.W.; methodology, N.M.N.; software, N.M.N.; validation, N.M.N. and G.J.W.; formal analysis, N.A.J.; investigation, N.A.J.; resources, G.J.W.; data curation, I.A.M.J.; writing—original draft preparation, N.A.J. and I.A.M.J.; writing—review and editing, N.M.N.; visualization, I.A.M.J.; supervision, N.M.N.; project administration, N.M.N.; funding acquisition, N.M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The author would like to thank the Central Statistical Body of Gunungkidul District (BPS), Indonesia, and Universitas Muhammadiyah Surakarta, Indonesia, for the rainfall dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Koerniawan, M.D. Recreation Park and Thermal Comfort: Site Measurement and Predictive Simulations in Assessment of Open Spaces in Taman Impian Jaya Ancol, Jakarta, Indonesia. In Proceedings of the 5th Yellow Sea Rim International Exchange Meeting on Building Environment and Energy 2010 (YSRIM), Yi Fu Building at Tongji University, Shanghai, China, 27 February–1 March 2010.
2. As-syakur, A.R.; Tanaka, T.; Osawa, T.; Mahendra, M.S. Indonesian rainfall variability observation using TRMM multi-satellite data. *Int. J. Remote Sens.* **2013**, *34*, 7723–7738. [CrossRef]
3. Saragih, I.J.A.; Rumahorbo, I.; Yudistira, R.; Sucahyono, D. Prediksi Curah Hujan Bulanan Di Deli Serdang Menggunakan Persamaan Regresi Dengan Prediktor Data Suhu Dan Kelembapan Udara. *J. Meteorol. Klimatol. Geofis.* **2021**, *7*, 6–14.
4. Matei, M.; Raischi, M.; Ciobotaru, N.; Laslo, L.; Boboc, M.; Zamfir, A.S.; Deák, G. Flood Protective Measures in Divici–Pojejena Wetland, Caras-Severin County, Romania. *J. Environ. Prot. Ecol.* **2017**, *18*, 235–245.
5. Nurfaida, W.; Ramdhani, H.; Shimozono, T.; Triawati, I.; Sulaiman, M. Rainfall trend and variability over Opak River basin, Yogyakarta, Indonesia. *J. Civ. Eng. Forum* **2021**, *7*, 109–120. [CrossRef]
6. Barman, U.; Choudhury, R.D.; Hussain, A.E.; Dahal, M.J.; Barman, P.; Hazarika, M. Comparative Assessment of AR, MA and ARMA for the Time Series Forecasting of Assam and Meghalaya Rainfall Division. In Proceedings of the 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2–4 July 2020; pp. 507–511. [CrossRef]
7. Saputro, D.R.S.; Mattjik, A.A.; Boer, R.; Wigena, A.H.; Djuraidah, A. Pendugaan Data Tidak Lengkap Curah Hujan Di Kabupaten Indramayu Dengan Kringing & Rata-rata Bergerak (Moving Average) Berdasarkan Data Tahun 1980–2000. In Proceedings of the Seminar Nasional Statistika Universitas Diponegoro 2011, Tembalang Semarang, Indonesia, 21 May 2011.
8. Nurlifa, A.; Kusumadewi, S. Sistem Peramalan Jumlah Penjualan Menggunakan Metode Moving Average Pada Rumah Jilbab Zaky. *INOVTEK Polbeng Seri Inform.* **2017**, *2*, 18. [CrossRef]
9. Solver. Moving Average. 2016. Available online: https://www.solver.com/moving-average (accessed on 9 May 2022).
10. Ul-Saufie, A.Z.; Yahya, A.S.; Ramli, N.; Hamid, H. Comparison Between Multiple Linear Regression and Feed forward Back propagation Neural Network Models For Predicting PM10 Concentration Level Based On Gaseous And Meteorological Parameters. *Int. J. Adv. Sci. Technol.* **2011**, *1*, 42–49.
11. Mehdizadeh, S. Using AR, MA, and ARMA Time Series Models to Improve the Performance of MARS and KNN Approaches in Monthly Precipitation Modeling under Limited Climatic Data. *Water Resour. Manag.* **2020**, *34*, 263–282. [CrossRef]