The Development of the SARIMA Model for Flood Disaster Resilience

Heri Sulistiyono[1](✉), Faisal Irshad Khan[2], Humairo Saidah[1], Ery Setiawan[1], I Wayan Yasa[1], I Wayan Suteja[1], Salehudin[1], and I Dewa Gede Jaya Negara[1]

[1] University of Mataram, Mataram 83125, Indonesia
h.sulistiyono@unram.ac.id
[2] Texas A&M University, College Station, TX 77840, USA
fikhan@tamu.edu

Abstract. Some agencies in the world warned that the recent flood disaster has been getting worse all over the world. Engineers have studied flooding and its impact with various simulation models. Yet, there is no model considered the best for all estimations of flood disasters in the future. In this paper, the authors propose a procedure to develop the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The model is for generating future rainfall. In this paper, the authors use the procedure for drainage evaluation in an effort of flood risk resilience. The authors use the rainfall data of the Monjok Station from 2008 to 2018 in the demonstration of the proposed procedure to generate 20 years of rainfall data from 2008 to 2028. This paper evaluates the flood resilience of the Wahidin drainage system. The authors use the Correlation Coefficient and Volume Error to determine the best model. The results showed that the best SARIMA model was (0, 1, 1) with a Correlation Coefficient of 0.71 and a Volume Error of 13%. Based on the model, the 10-year return period of future rainfall is 137.60 mm. This rainfall will cause runoff in the Wahidin 01, 02, and 03 of 4.35 m3/sec, 1.22 m3/sec, and 9.496 m3/sec, respectively. However, because the capacity of the three canals is only 1.32 m3/s, only Wahidin 02 canal is safe until 2028. Wahidin 01 and 03 channels are not.

Keywords: SARIMA Model · Correlation Coefficient · Volume Error · Drainage Evaluation · Flood Risk Resilience

1 Introduction

Please Natural disasters such as floods are becoming more and more common all over the world [1] [2] [3]. According to the National Disaster Management Agency (BNPB), 827 floods have occurred throughout Indonesia [4] [5]. Floods often happen when rainfall is higher than usual. The Lack of water infiltration zones, insufficient drainage systems, meandering rivers, tidal effects in estuaries, and garbage in waterways can increase the severity of flood events [6] [7]. Besides improving watersheds and drainage systems, engineers have to study changes in future rain patterns and characteristics [8] [9] [10] to be able to evaluate the ability of the drainage system in the effort of flood risk resilience.

B. Anshari et al. (Eds.): MIMSE-C-A 2022, AER 215, pp. 211–222, 2023.
https://doi.org/10.2991/978-94-6463-088-6_21

In some cases, the historical rainfall data is insufficient for flood analysis. Consequently, engineers must perform modeling simulations to obtain adequate rainfall data [11] [12]. Among the simulation methods, the Forecasting method based on the same data type becomes preferable when another type is unavailable. This type of modeling uses past events to estimate future events. This paper proposes the Development of the SARIMA Model for Flood Disaster Resilience.

The SARIMA model is a univariate model. The SARIMA Model for forecasting rainfall data only considers the past rainfall events instead of involving other variables [13] [14]. The objectives of this research are:

1. To develop a Seasonal Autoregressive Moving Average (SARIMA) model for the simulation of future rainfall in Mataram City.
2. To estimate the future runoff in the Wahidin main canals based on future rainfall.
3. To evaluate the status of Wahidin drainage main system based on future runoff.

2 Methodology

2.1 The Principles of Proposed Data Forecasting

According to some references, forecasting includes [15] [16] [17] [18]:

1. The random distribution of model residuals.
2. Short-term forecasting is more accurate than long-term forecasting.
3. Group forecasting is more accurate than individual forecasting.

The flowchart in Fig. 1 shows the proposed procedure.

The proposed procedure has three steps. The first step is data preparation. The second step is model development. The model considers seasonal phenomena, previous self-data, and changes in residuals. The third step is model application. In this paper, the application of the model is the evaluation of the drainage system for Flood Disaster Resilience.

The description of the steps is as follows:

1. The Development of the SARIMA model. Some models that are popular in the time series analysis are as follows [19] [20]:

   a. Time series Plot of rain data to see the presence of seasonality Calibrate model parameters using the Box-Cox method
   b. Determine the autoregressive coefficient using the Autocorrelation function.

The autocorrelation function is the relationship (correlation) between each data and its previous data. The equation to perform the autocorrelation function is

$$\rho_k = \frac{\sum_{t=1}^{n-k}\left(Z_t - \overline{Z}\right)\left(Z_{t+k} - \overline{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \overline{Z}\right)^2},\ \mathrm{k} = 0, 1, 2, \ldots \quad (1)$$

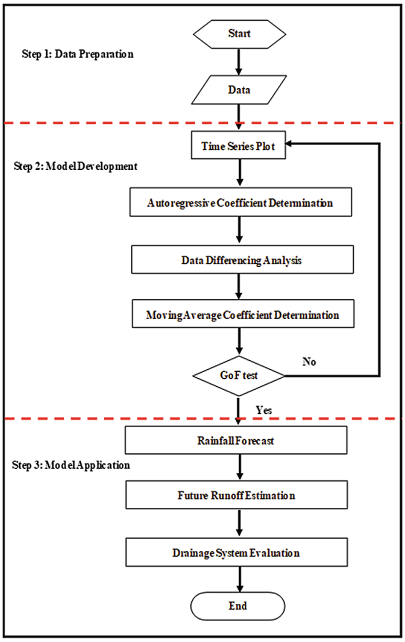

Fig. 1. The Flowchart of Proposed Procedure.

With: ρ_k is a correlation coefficient for a k-lag period, Zt is an observation value in t period, Zt + k is an observation value in t + k period, Z is the average of observation values.

The autoregressive model has the following form

$$Yt = \beta0 + \beta1\ Yt - 1 + \beta2\ Yt - 2 + \ldots + \beta p\ Yt - p + \varepsilon t \tag{2}$$

With: Yt is the stationary series; Yt-1 and Yt-2 are the past value of the series in question; β0, β1, and β2 are the constants and model coefficients; and εt is the model residuals.

c. Perform data differencing by applying data shift (Lagging).

A differencing is the process of calculating the deviation between Zt and Zt-1. Each data has an estimated difference from the previous data. Lambda values in the Box-Cox Plot indicate the need for a transformation to stabilize the mean and standard deviation. Transformation formula is as

$$y(\lambda) = \begin{cases} \frac{X^{\lambda} - 1}{\lambda}, \mathit{if}\,\lambda \neq 0 \\ \log X, \mathit{if}\,\lambda = 0 \end{cases} \tag{3}$$

With: y is the transformed data, X is the data, and λ is the transformation model parameter. The backward shift operator is suitable to describe the difference [18] and is as shown below.

$$BXt = Xt - 1 \tag{4}$$

With: Xt is the value of variable X at time t, X t−1 is the value of variable X at time t1, and B is the backward shift operation.

d. Determine the moving average coefficient using the Partial Autocorrelation function.

A partial autocorrelation function shown below states the relationship between an observation result and the result of the observation.

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1}, j\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1}, j\rho_j} \tag{5}$$

With: *Økk* is the partial autocorrelation coefficient for the -period lag, *Økj* is the partial autocorrelation coefficient for lag period j = 1,2, … *k* – 1

The Moving Average model has the following form:

$$\text{Yt} = \alpha 0 + \varepsilon \text{t} - 1\varepsilon \text{t} - 1 - 2\varepsilon \text{t} - 2 - \ldots - \text{q}\varepsilon \text{t} - \text{q} \tag{6}$$

With: Yt is the stationary series value; εt is the forecast error (error); εt 1 and εt-2 are the past forecasting error; α0, α1, and α2 are the constants and the model coefficients. In case of Engineers must merge Autoregressive-Moving Average (ARIMA) Models to obtain the best solution, the form is becoming

$$\text{Yt} = \beta 0 + \beta 1\text{Yt} - 1 + \ldots + \varepsilon \text{t} - \alpha 1\varepsilon \text{t} - 1 + \ldots \tag{7}$$

With: Yt is the stationary series; Yt-1 and Yt-2 are the past value of the series in question; β0, β1, and β2 are the constants and the model coefficients; εt is the model residuals; εt-1 and εt-2 are the past forecasting error; α0, α1, and α2 are the constants and model coefficients.

If the discrepancies or the "differencing" arises in the sequential data values, engineers need an "integrated" part. The model becomes an Autoregressive Integrated Moving Average (ARIMA). The equation is

$$(1 - \text{B})(1 - \theta \text{t B})\text{Xt} = \mu + (1 - \theta \text{t B})\varepsilon \text{t} \tag{8}$$

With: Xt is the stationary series; (1-B) is the differencing of non-seasonal parameter, θt B is the AR non-seasonal parameter, μ is the average of data; εt is the model residuals.

In case of the season affects the movement of sequential changes in the data, then the model form is

$$\Phi_P B^S \phi_{P(B)(1-B)^d (1-B^S)^D Z_t} = \theta_q(B)\Theta_q\left(B^S\right)a_t \tag{9}$$

With: $P_{(B)}$ is the AR Non seasonal parameter, $\Phi_P B^S$ is the AR seasonal parameter, $(1 - B)^d$ is the differencing of non-seasonal parameter, $\left(1 - B^S\right)^D$ is the differencing of seasonal parameter, $\theta_q(B)$ is the MA non seasonal parameter, and $\Theta_q\left(B^S\right)$ is the MA of seasonal parameter.

e. Perform the goodness of fit test

In this paper, the authors use the correlation coefficient and the volume Errors to test the model results [21]. The equations are

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})}} \tag{10}$$

With: r_{xy} is the correlation coefficient, *x* is the observed data, y is the predicted data.

$$VE = \frac{\sum_{i=1}^{n}(x_i - y_i)}{\sum_{i=1}^{n}(x_i)} \tag{11}$$

With: *VE* is the volume error, *x* is the observed data, y is the predicted data.

2. Model Application

   a. The Forecasting rainfall using the Seasonal Autoregressive Moving Average (SARIMA) model
   b. The estimation of the future runoff
   c. The evaluation of the drainage system this paper uses the rational method in the peak runoff calculation. The equation is [22] as follows

$$Qp = 0,002778\ C.I.A \tag{12}$$

With Q_p is the peak runoff rate (discharge) in m^3/sec, C is the runoff coefficient (0<C<1), I is the rainfall intensity in mm/hour, and A is the area of the watershed in hectares.

The formula of concentration time in this paper is as follows [23]

$$t_c = \left(\frac{0.87.L^2}{1000S}\right)^{0.385} \tag{13}$$

With: tc is the time of concentration, L is the length of the main canal from the upstream to the drain in Km, S is the slope of the canal in m/m.

The authors use runoff coefficient (C) as shown in Table 1 [23]

The authors use the Mononobe formula as follows [24]

$$I = \frac{R_{24}}{24}\left(\frac{24}{tc}\right)^{\frac{2}{3}} \tag{14}$$

With: I is the rain intensity (mm/hour), R_{24} is the maximum rain height in 24 hours (mm), and tc is the time of concentration of rainfall.

In this paper, the authors use the open channel approach to calculate the drainage capacity. The formula is as follows [24]

$$\mathrm{Q} = \mathrm{VA} = \frac{1}{n}R^{\frac{2}{3}}S^{\frac{1}{2}}A \tag{15}$$

With: Q is the Flow Discharge, A is the Canal Cross-sectional Area, R is the Hydraulic radius, and S is the Canal slope.

Table 1. Runoff Coefficients.

Description	Runoff Coefficient, C
Businesses	
Urban	0.70–0.95
Suburbs	0.50–0.70
Housing areas	
Single unit	0.30–0.50
separate multi-unit	0.40–0.60
Merged multi-unit	0.60–0.75
Village	0.25–0.40
Apartment	0.50–0.70
Industries	
Light industrial area	0.50–0.80
Dense industrial area	0.60–0.90
Pavements	
Asphalt and concrete	0.70–0.95
Brick and paving	0.50–0.70
Roof	0.75–0.95
Yards, Sandy soil	
Flat, 0%-2%	0.13–0.17
Moderate, 2%-7%	0.10–0.15
Steep, $>$ 7%	0.15–0.20
Playground	0.20–0.35
Graveyard	0.10–0.25
Forest	
Flat, 0%-5%	0.10–0.40
Moderate, 5%-10%	0.25–0.50

3 Results and Discussions

The following sections present a case study to demonstrate the application of the proposed framework. Figure 2 shows the case study location.

Figure 2 shows the existing condition of the Wahidin Main Canal. This canal has three sections: Wahidin 1, Wahidin 2, and Wahidin 3. The catchment area of this canal is 240.79 ha. Table 2 shows the technical data of the Wahidin canal.

This paper uses maximum rainfall data from the Monjok Station from 2009 to 2018. Figure 3 shows the maximum rainfall from the Monjok Station from 2009 to 2018.

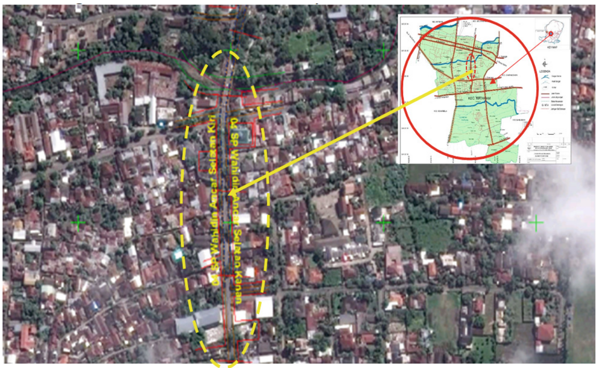

Fig. 2. Case Study Location.

Table 2. Technical Data of the Canals.

Canal	A (ha)	C	S	L (m)	W (m)	D (m)
Wahidin 01	75.52	0.3	0.2%	235.3	1.60	0.60
Wahidin 02	20.11	0.3	0.2%	240.6	1.60	0.60
Wahidin 03	145.1	0.4	0.2%	650.0	1.60	0.60

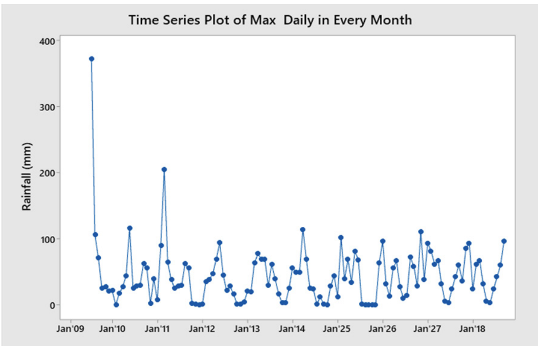

Fig. 3. The Maximum Daily Rainfall from the Monjok Station from 2009 to 2018.

Figure 3 shows the monthly rainfall data from the Monjok Station following the seasons. The monthly rainfall reaches the peak value every January and the base value every August. By the proposed procedure, the next step is to analyze the autocorrelation function plot. Figure 4 shows the autocorrelation function plot for monthly rainfall data.

The data is stationary concerning the mean if the autocorrelation of the first 3-lag is within the confidence line. Figure 4 shows that the autocorrelation of the first lag is out of the confidence line.

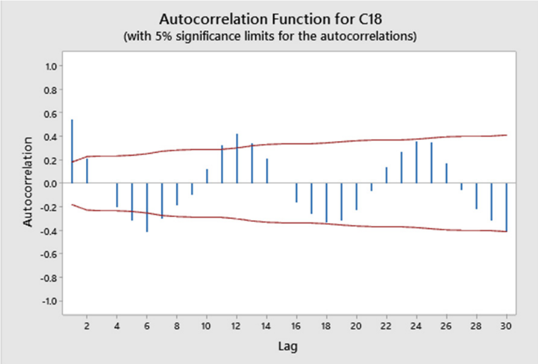

Fig. 4. The Autocorrelation Function Plot for Maximum Daily Rainfall.

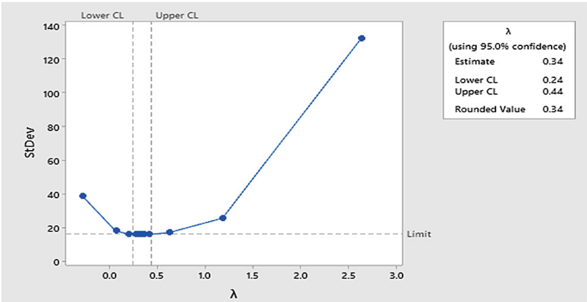

Fig. 5. The box-cox plot method.

Figure 5 shows the Box-Cox Plot analysis to obtain the value of lambda values should be in the Box-Cox transformation to stabilize the mean and the standard deviation. Figure 5 shows that the estimated Lambda value is 0.34. Figure 6 shows the autocorrelation function after the Box-Cox transformation with the Lambda of 0.34.

Figure 6 shows no autocorrelation value sticking out of the confidence limit. The SARIMA modeling using transformed data does not require an autoregressive function anymore. Figure 7 shows the partial autocorrelation function.

Figure 7 shows one partial autocorrelation value sticking out the confidence limit. The transformed data requires a first-order moving average function to model the data. Table 3 shows the ANOVA of the SARIMA.

Therefore, according to Table 3, the SARIMA equation becomes

$$(1 - B)\, yt = (1 + 0.737)\,(1 + 0.873)\, \varepsilon t \tag{16}$$

Figure 8 shows the prediction of rainfall from 2009 to 2028 using the SARIMA model.

Figure 8 shows the predicted rainfall increases in the future. Table 4 shows the goodness of fit tests.

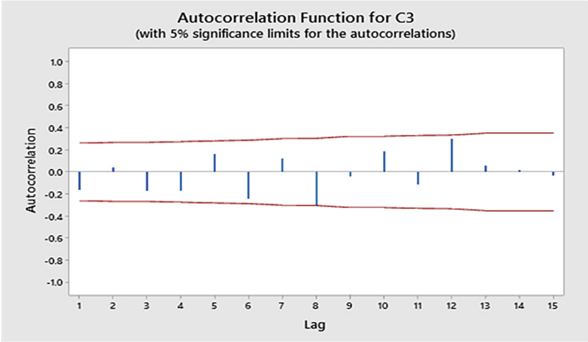

Fig. 6. The Autocorrelation Function after the transformation.

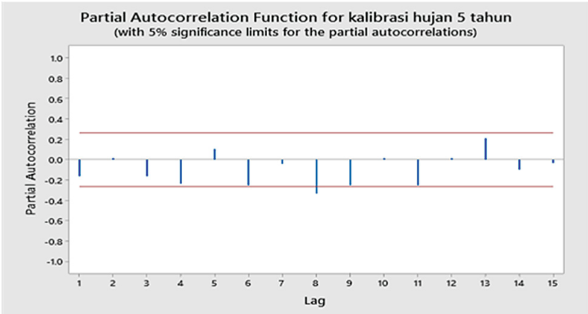

Fig. 7. The Partial Autocorrelation Function after the transformation.

Table 3. The ANOVA.

Parameter	Coef	SE Coef.	T-Value	P-Value
MA 1	0.737	0.096	7.69	0.000
SMA 12	0.873	0.133	6.56	0.000
Constant	0.012	0.011	1.14	0.259

In Table 4, the correlation coefficient is 0.71, and the Volume Error is 13%. This SARIMA model is suitable for generating future rainfall data.

Figure 9 shows the design rainfall of 2-year, 5-year, and 10-year return periods are 106.87 mm, 126.2117 mm, and 137.67 mm, respectively.

In Table 5, the 10-year return period of discharges of the Wahidin 1, Wahidin 2, and Wahidin 3 are 4.35, 1.22, and 8.35, respectively.

In Table 6, the capacities of the Wahidin 1, Wahidin 2, and Wahidin 3 are 1.32, 1.32, and 1.32, respectively. Therefore, Table 7 shows the status of the Wahidin drainage system until 2028.

In Table 7, one canal is still safe but the other two canals are not.

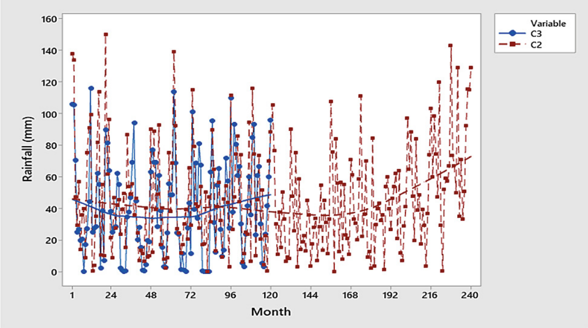

Fig. 8. The time series plot of rainfall and predicted rainfall for the same year.

Table 4. The Goodness of Fit Test.

Parameter	Results
R	0.71
VE	0.13

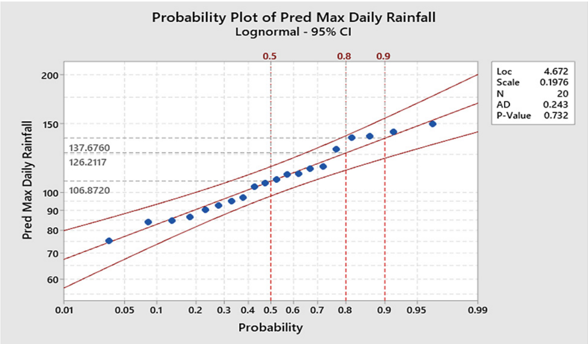

Fig. 9. The Probability Plot of Predicted Maximum Daily Rainfall

Table 5. The Calculation of flow rate.

Canal	A (ha)	C	tc (mnt)	R_{10} (mm)	I (mm/h)	Disc (m^3/s)
Wahidin 01	75.52	0.3	34.34	137.6	69.23	4.35
Wahidin 02	20.11	0.3	31.8	137.6	72.87	1.22
Wahidin 03	145.1	0.4	53.11	137.6	51.77	8.35

Table 6. The Calculation of drainage capacity.

Canal	n	S	A (m^2)	P (m)	R (m)	Q (m^3/s)
Wahidin 01	0.013	0.20%	0.96	3.8	0.25	1.32
Wahidin 02	0.013	0.20%	0.96	3.8	0.25	1.32
Wahidin 03	0.013	0.20%	0.96	3.8	0.25	1.32

Table 7. The Status of the Wahidin Drainage System in 2028.

Canal	Capacity	Discharge	Status
Wahidin 01	1.32	4.35	Not safe
Wahidin 02	1.32	1.22	Safe
Wahidin 03	1.32	8.35	Not safe

4 Conclusion

Based on all studies, the authors can conclude that

1. This study has developed a Seasonal Autoregressive Moving Average (SARIMA) model for the simulation of future rainfall in Mataram City,
2. The prediction of future runoff in the Wahidin main canals no.01, 02, and 03 are 4.35 m^3/s, 1.22 m^3/s, and 8.35 m^3/s, respectively,
3. The Wahidin 02 canal is safe. However, Wahidin 01 and 03 canals are not safe.

References

1. Hitchcock, K.: 10 Facts About Floods – Statistics & Data In 2022, https://housegrail.com/floods-statistics/, last accessed 2022/08/02.
2. Boivin, G.: Where do floods occur most often, https://guillaumeboivin.com/where-do-floods-occur-most-often.html, last accessed 2022/08/02.
3. Richard.: Qld floods to occur more often due to climate change, https://www.storymirror.com.au/qld-floods-to-occur-more-often-due-to-climate-change/, last accessed 2022/08/05.
4. Hermawan, B.: 2.276 Bencana Alam Terjadi Sepanjang Januari Hingga Oktober, Republika.co.id, last accessed 2022/08/03.
5. Valentino, A.: Banjir semakin sering, dapatkah pendekatan kuno Belanda jadi solusi di tengah ancaman perubahan iklim?, https://www.bbc.com/indonesia/vert-fut-59642242, last accessed 2022/08/02.
6. Tiza, M.T.: The Effects Of Poor Drainage System On Road Pavement: A Review. International Journal For Innovative Research In Multidisciplinary Field 2(8) (2022).
7. Sohn, W., Brody, S.D., Kim, J.H., and Li, M.H.: How effective are drainage systems in mitigating flood losses. Journal Cities 107 (2022).
8. Kuhn, K.: Why Engineers Need To Be Thinking About Climate Change, http://insidescience.org/news/why-engineers-need-be-thinking-about-climate-change, last accessed 2022/08/08.

9. Frueh, S.: Engineering Responses to Climate Change, https://www.nationalacademies.org/news/2021/09/engineering-responses-to-climate-change, last accessed 2022/08/11.
10. Fork, D. and Koningstein, R.: Engineers: You Can Disrupt Climate Change. Decarbonization, Carbon Capture, And Solar-Radiation Management will provide work for decades to come, https://spectrum.ieee.org/engineers-you-can-disrupt-climate-change, last accessed 2022/08/12.
11. Xu, P., Wang, D., Wang, Y. and Singh, V.P.: A Stepwise and Dynamic C-Vine Copula–Based Approach for Nonstationary Monthly Streamflow Forecasts. Journal of Hydrologic Engineering 27(1) (2022).
12. Mishra, A., Mukherjee, S, Merz; B., and Singh, V.P.: An Overview of Flood Concepts, Challenges, and Future Directions. Journal of Hydrologic Engineering 27(6) (2022).
13. Nanlohy, Yonlib Weldri Arnold.: Model Fungsi Transfer Multi Input Untuk Peramalan Curah Hujan Di Kota Surabaya. Institut Teknologi Sepuluh November, Surabaya (2017).
14. Anwar, F. and Jeyanthi, R.: House Price Prediction using SARIMA and VAR Model. International Journal of Engineering Research & Technology 11(4) (2022).
15. Assauri, S.: Teknik dan Metode Peramalan. Jakarta Fakultas Ekonomi, Jakarta (1984).
16. Subagyo, P.: Forecasting Konsep dan aplikasi. BPPE UGM, Yogyakarta (1986).
17. Makridakis.: Metode dan Aplikasi Peramalan. Erlangga, Jakarta (1995).
18. Makridakis.: Metode dan Aplikasi Peramalan, Jilid 1. Erlangga, Jakarta (1999).
19. Montgomery, D.C.: Introduction to Time Series Analysis and Forecasting. John Wiley And Son, Inc, New York (2007).
20. Aulia, F., Yozza, H., and Devianto, D..: Peramalan Curah Hujan Bulanan Kabupaten Tanah Datar Dengan Model Seasonal Autoregressive Integrated Moving Average (SARIMA). Skripsi Jurusan Matematika FMIPA UNAND, Padang (2019).
21. Legates, D.R.; McCabe, G.J.: Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Journal Water Resour. Res. 35 (1), 233–241 (1999).
22. Shyamsunder, B., Vishnuvardhanreddy, K..: Runoff Estimation by Using Rational Method: A Review. International Journal ANS. 10(3), 1-4 (2021).
23. Suripin.: Sistem Drainase Yang Berkelanjutan (Sustainable Drainage System). Andi Offset, Yogyakarta (2004).
24. Triatmodjo, B.: Hidrologi Terapan. Beta Offset, Yogyakarta (2008).