Short-Term and Long-Term Rainfall Forecasting Using ARIMA Model

M. M. H. Khan*, M. R. U. Mustafa, M. S. Hossain, S. Shams, and A. D. Julius

Abstract—**Rainfall prediction plays a vital role in terms of event preparedness and prevention. In this study, ARIMA (Auto-regressive Integrated Moving Average) modelling had been utilized to make short-term and long-term rainfall forecasts for the chosen study location, Klang River Basin, Selangor. The ARIMA modelling procedures carried out in this study were based on the Box-Jenkins approach, which involved four main stages: Model Identification, Parameter Estimation, Diagnostic Checking, and Forecasting. Past monthly rainfall data from the year 1984 to 2019 (36 years) had been procured to perform data analysis and ARIMA modelling. Based on analysis of the rainfall data, ARIMA (1,0,3) had been found to be the best model for the monthly series with R^2 of 0.78 ,whereas ARIMA (1,0,2) was the best model for the annual series with R^2 of 0.52. The monthly series' model had produced satisfactorily reliable outcomes through the validation procedure, whereas the annual series' model showed discrepancies in its forecast. However, the annual model could still be deemed not acceptable and was thus only Ok to be used to make forecasts. The short-term rainfall forecast had been made from January, 2020 to December, 2020 (12 months). Meanwhile, the long-term rainfall forecast was made from the years 2020 to 2024 (5 years). Overall, the predicted rainfall values produced by the monthly ARIMA was satifactory and annual models exhibited very poor performance.**

Index Terms—**Rainfall forecasting, ARIMA modelling, time series analysis, Klang River**

I. Introduction

In general, the Malaysian climate can be classified as equatorial because it is situated near the equator. Being a tropical country that is predominantly hot and humid throughout the year, Malaysia experiences an average annual rainfall of 2,500 mm. Aside from that, it is also subjected to the Northeast Monsoon season, which lasts from mid-October to March. The Northeast Monsoon originating from China and the North Pacific usually brings along with it high intensities of rainfall compared to any other time of the year [1]. Moreover, the transition between the Northeast Monsoon and Southwest Monsoon seasons also typically brings about sudden, rapid bursts of rainfall that can lead to flash floods, especially in urban areas with poor drainage systems. The chosen study location, Klang River Basin, Selangor, is known to be surrounded by heavily urbanised and densely populated regions [2]. According to [3], in Klang Valley alone, the total population had been estimated to be around 4.4 million (roughly 16% of the country's population), with an approximate growth rate of 5% per annum. The large and continuously rising population in the cities have thus led to ongoing urbanisation and geographical alterations – ultimately causing many people to be exposed to flooding disasters, especially during the wet monsoon season.

Based on the Klang River Basin Environmental Improvement and Flood Mitigation Project in 2007, around 17,000 hectares (13%) of the river basin is prone to floods, with an average of up to three flooding occurrences happening each year. In addition, around half a million people reside within the flood-prone area and are therefore severely affected. Overall, the average estimated flood-related damages surrounding the Klang River Basin area had been reported to reach a tremendous RM 6.3 million annually [4]. Therefore, reliable rainfall predictions are crucial to ensure appropriate flood administration and mitigation, effective water body management, and carry out proper urban planning. This is further supported by [5]. They stated that the prediction of rainfall holds a profound impact on a country's flood prevention and management strategy and waterbody management.

With regards to rainfall prediction, the references [6, 7] explained that one of the most effective approaches available for the analysis of rainfall time series data is the model introduced by Box and Jenkins [8], which is the Autoregressive Integrated Moving Average (ARIMA). Liu [9] compared two-time series models- ARIMA and ARIMA-GARCH and found that both methods are reasonably good. However, they concluded that the ARIMA-GARCH model is more suitable where the variability of the variables is unequal across the data range. Unnikrishnan and Jothiprakash [10] used a hybrid method which incorporates Singular Spectrum Analysis (SSA), Auto-Regressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN). This hybrid approach provides them an impressive accuracy in rainfall prediction. Karimi [11] used two different methods, namely ARIMA and gene expression programming (GEP) to forecast rainfall in Iran and found ARIMA model slightly superior than GEP by obtaining higher R^2 than GEP, whereas RMSE, MAE lower. Rahman [12] and Swain [13] showed that the ARIMA method provides good results for rainfall forecasting, and this method can be used on large-scale issues.

Nyatuame and Agodzo [14], explained that the first part (AR) stands for the autocorrelation between current and past observations, while the second part (I) signifies the level of

Manuscript received July 20, 2022; revised August 22, 2022; accepted January 30, 2023.

M. M. H. Khan and A. D. Julius are with Faculty of Engineering and Quantity Surveying (FEQS), INTI International University, Negeri Sembilan, Malaysia.

M. R. U. Mustafa is with the Department of Civil & Environmental Engineering, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia.

M. S. Hossain is with School of Energy, Geoscience, Infrastructure and Society. Department of Civil Engineering, Heriot-Watt University, Putrajaya 62200, Malaysia.

S. Shams is with Civil Engineering Programme Area, Universiti Teknologi Brunei, Jalan Tungku Link, Gadong BE1410, Brunei.

*Correspondence: shihab.bd@gmail.com (M. M. H. K.)

doi: 10.18178/ijesd.2023.14.5.1447

differencing required to transform a series from non-stationary to stationary. Lastly, the third part (MA) was said to denote the autocorrelation structure of error. According to Naill and Momani [15], there are three main parameters in a general ARIMA model: the autoregressive parameter (p), the number of differencing passes (d), and the moving average parameter (q). These parameters may then be summarised and presented as ARIMA(p, d, q). One of the issues that this study seeks to address is the lack of study on ARIMA modelling with regards to rainfall prediction in Malaysia. Upon thorough research, it had been found that past studies on the subject were significantly limited. This suggests that ARIMA modelling is a method that remains relatively unexplored in Malaysia, especially in terms of hydrological rainfall forecasting. Therefore, this research aspires to shed more light on the potential of ARIMA modelling for rainfall forecasting through the subsequent development of ARIMA models. Moreover, the findings of this study may also serve as a reference or guide for other similar studies in the future, as well as to aid in mitigating and managing hydrological-related disasters (i.e., flooding) surrounding the Klang Basin area in Peninsular Malaysia.

II. Methodology

A. Study Location

This study had been conducted at Klang River Basin, Selangor, which is situated in the West coast of Peninsular Malaysia. Its catchment covers a total area of approximately 1.29 km^2 while stretching across a length of 120 km, and is known as the largest and most significant basin in the state. According to [2], Klang River Basin is also highly urbanized and very densely populated compared to any other river basins in the country.

Raw data used for analysis and modelling had initially been obtained from one of the rainfall stations located within the chosen study area. The selected rainfall station for this study was the Pejabat JPS Klang rainfall station (Station No.: 3014084), situated at coordinates 3°2'20" (latitude) and 101°52'40" (longitude). The selection of this particular station had been duly subjected to data availability for the desired analysis period.

B. Data Collection, Utilization and Soft wares

The raw rainfall data used in this study have been sourced from the Malaysian Department of Irrigation and Drainage (DID), whereby the requested data were that of the selected station’s past monthly readings, from the year 1984 to 2019 (36 years). After obtaining the requested data, any missing data have been interpolated accordingly, which was then followed by sorting, in order to produce two complete sets of monthly and annual data. In this study, precipitation at the chosen location have been forecasted for both short-term and long-term periods. The short-term forecast involved a monthly rainfall prediction which encompassed a period of 12 months (January 2020 to December 2020). Meanwhile, the long-term forecast involved an annual rainfall prediction throughout a 5-year span (year 2020 to 2024). Table I below summarizes the detailed particulars for the short-term and long-term forecasting.

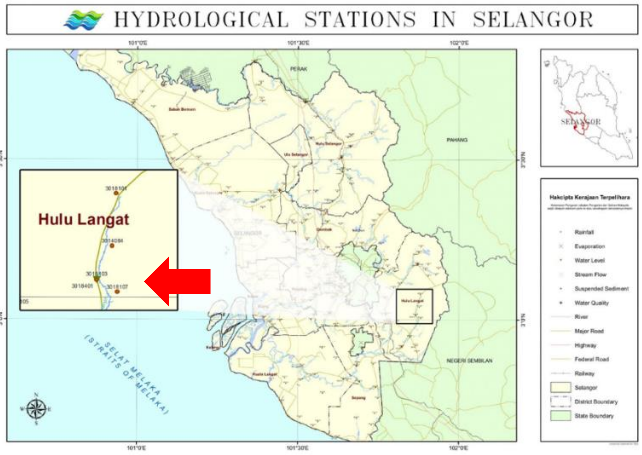


Fig. 1. Map location of Pejabat JPS Klang (DID, n.d.).

TABLE I: Detailed Particulars for Short-Term and Long-Term Forecasting.

No.	Description	Period	Year
1.0	**Short-term forecasting**	**1 year (12 months)**	**2020**
1.1	Past data used for model development	36 years (432 months)	1984–2019
1.2	Past data used for model validation	1 year (12 months)	2019
1.3	Total past data used for short-term forecasting	36 years (432 months)	1984–2019
2.0	**Long-term forecasting**	**5 years**	**2020–2024**
2.1	Past data used for model development	36 years	1984–2019
2.2	Past data used for model validation	5 years	2015–2019
2.3	Total past data used for long-term forecasting	36 years	1984–2019

In order to efficiently sort and analyse the collected rainfall data, the use of appropriate computer programs and software was highly essential. For this study, the MS Excel program had mainly been used to systematically sort and tabulate the rainfall data. Meanwhile, to analyse and model the data, both XLSTAT and Minitab have been used in conjunction with one another.

C. ARIMA Modelling

The ARIMA modelling procedures for this study were based on the Box-Jenkins approach, which involved four principal stages: Model Identification, Parameter Estimation, Diagnostic Checking, and lastly, Forecasting. In the model identification stage, the stationarity of the rainfall data series (annual and monthly) were examined via plotting approach, as well as through stationary tests (Augmented Dickey-Fueller test, Kwiatkowski-Phillips-Schmidt-Shin test and Mann-Kendall test). The data series to be used for ARIMA modelling should ideally be stationary. If a data series was non-stationary, then an appropriate data transformation needs to be carried out (i.e. differencing, log transformation, etc.). Through these steps, the order of the annual and monthly models' d parameter could be determined. Moreover, the presence of any seasonality had also been checked for both the annual and monthly data series.

In parameter estimation, the initial orders of the AR(p) and MA(q) parameters of a time series was estimated by assessing its autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. This essentially produced several potential models – each with varying orders of p and q, within the range of values of one to three. The corrected Akaike Information Criterion (AIC) was thus used for distinguishing the best-fit model out of all the available, tentative models – the best-fit model for the annual and monthly series was selected based on having the least AIC value. After the best-fit model had been identified, its performance was subsequently assessed before being used for actual forecasting. The said assessment involved a sequence of diagnostic checking in terms of independence, homoscedasticity, and distribution of the fitted model's residuals.

After a model had passed the diagnostic checking, it may then be proceeded to be used for forecasting. However, to provide additional assurance of its prediction reliability, the model would be put through a simple validation process. The validation of the annual and monthly models was carried out by comparing their forecasted values against the actual data series. If the outcome shows adequate similarity, then the model is therefore a good model. Lastly, the validated ARIMA models were used for making future short-term and long-term rainfall forecasts. Since forecasting would be carried out for both short-term and long-term periods, two distinct models were therefore developed using the same procedural steps described previously. To develop the models for both short-term and long-term forecasting, 36 years' worth of past rainfall data (from 1984 to 2019) had been utilized, respectively. Thereafter, to validate the short-term forecasting model, past data from the year 2019 had been used, whereas to validate the long-term forecasting model, past data from 2015 to 2019 were used. The summarized details on the models' development and validation periods can be found in Table I.

D. Model Performance Evaluation

The performances of the presented models are evaluated based on their coefficient of determination (R^2). The coefficient of determination (R^2) can be given by following Eq. (1):

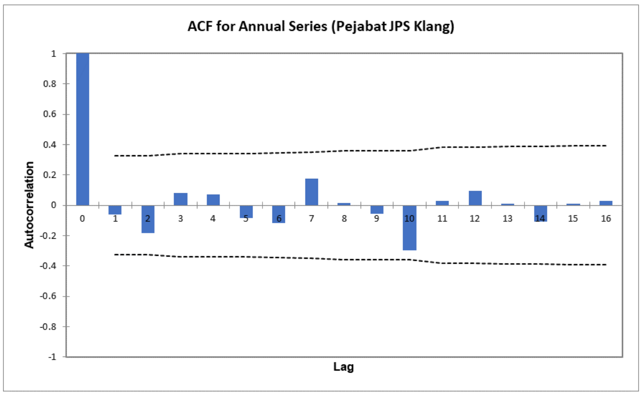
$$R^2 = 1 - \frac{\sum_{i=1}^{N}(R_i - \widehat{R_i})^2}{\sum_{i=1}^{N}(\widehat{R_i} - \bar{R})^2} \qquad (1)$$

where R^2, N, Ri, $\widehat{R_i}$, $\overline{R}$ are determination coefficient, of observations, observed data, predicted values and mean of observed data, respectively. R value close to unity indicates a satisfactory result, while a low value or one that is close to zero implies an inadequate result number.

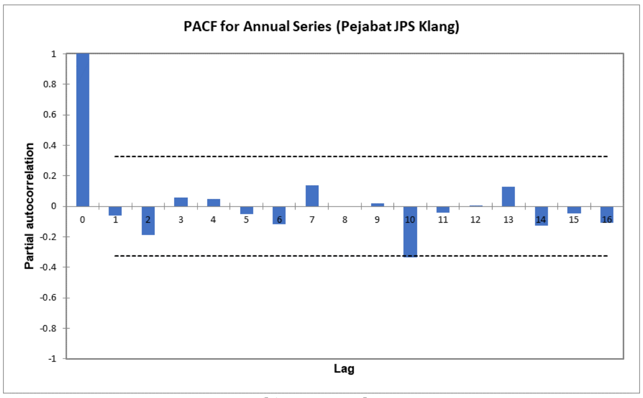
III. Results and Discussion

A. ARIMA Modelling for Annual Series

One of the prominent methods used for initial identification of model orders is through visual inspection of a time series' ACF and PACF plots [14]. Fig. 2(a) shows the ACF plot of the annual series, whereas Fig. 2(b) shows the PACF plot. In order to estimate the initial model parameters, these plots have been jointly utilized. Based on brief observation of the ACF plot, none of the lags appeared to be significantly correlated, therefore pointing to an MA(q) process. This meant that the ARIMA(p,d,q) model for long-term forecasting can reasonably be developed starting with the parameters (0,0,1).



(a). ACF plot



(b). PACF plot

Fig. 2. ACF and PACF plot of the annual series.

To determine the best ARIMA model for the annual series, several potential models with varying p and q orders (between zero to three) have been analysed by their corrected Akaike Information Criterion (AIC) value, as shown in Table II. Since the annual series was stationary and no differencing was required as seen in Table III, the order of d was left to remain at zero. Subsequently, XLSTAT was used to compute the AIC value of each model, whereby the best model would have the lowest AIC value. Based on the results in Table II, the best model was found to be the ARIMA(1,0,2) model, which had the least AIC value of 548.334.

The chosen model was then fitted to examine how well it was able to resemble the original series – the closer the resemblance, the better the fit, hence the more accurate the model. Fig. 3 presents a comparison between the original time series plot and the fitted ARIMA(1,0,2) model. Albeit a gap and the plot of the fitted series could be seen having a rather not so close resemblance to the original series. The R^2 value found to be 0.52. This level of similarity provided a fair indication that the fitted model would not be able to produce satisfactorily accurate forecasts.

TABLE II: ARIMA MODEL FOR ANNUAL SERIES (PEJABAT JPS KLANG RAINFALL STATION)

ARIMA (p,d,q) Model	AIC value
(0,0,1)	625.138
(0,0,2)	609.849
(0,0,3)	599.373
(1,0,1)	570.270
(1,0,2)	**548.334**
(1,0,3)	548.467
(2,0,1)	559.563
(2,0,2)	558.291
(2,0,3)	557.878
(3,0,1)	550.682
(3,0,2)	560.036
(3,0,3)	563.909

TABLE III: RESULTS FOR ADF, KPSS AND MK TESTS FOR PEJABAT JPS KLANG RAINFALL STATION

Data Series (1984 – 2019)	p-value			Conclusion
	ADF Test	KPSS Test	MK Test	
Annual	0.106	0.647	0.673	Stationary
Monthly	<0.0001	0.696	0.816	Stationary

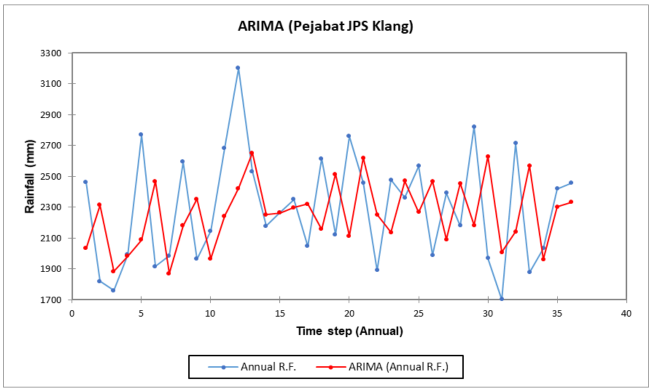
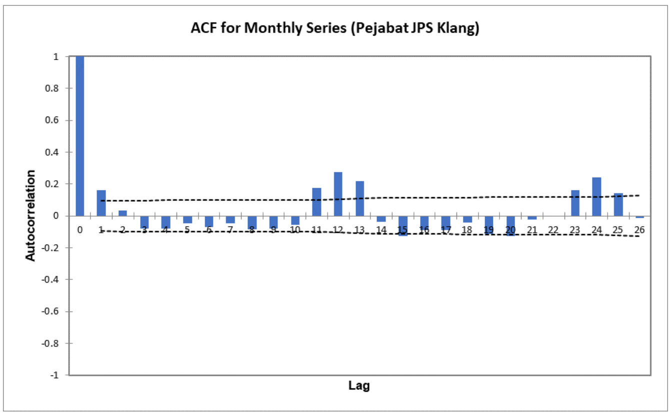


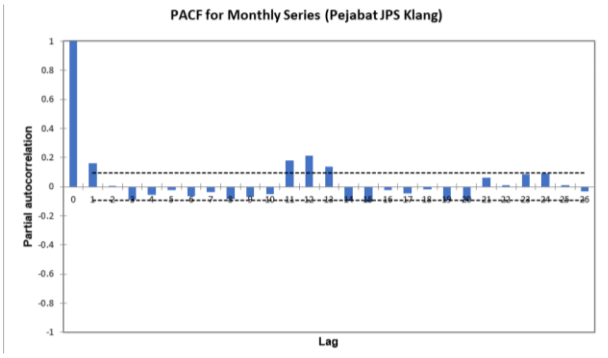
Fig. 3. Comparison of original series vs. fitted series for ARIMA(1,0,2).

B. ARIMA Modelling for Monthly Series

The Fig. 4(a) shows the ACF plot of the monthly series, whereas Fig. 4(b) shows the PACF plot. In order to estimate the initial model parameters, these plots have been jointly utilized. From the PACF plot, a peak was seen at lag 1, which then cuts off immediately. This more or less revealed the presence of a first order AR. On the other hand, the peak observed at lag 1 of the ACF plot – which also cuts off right afterwards – suggested the presence of an MA(q) component. This meant that the ARIMA(p,d,q) model for short-term forecasting can reasonably be developed starting with the parameters (1,0,0).



(a). ACF plot



(b). PACF plot

Fig. 4. ACF and PACF plot for monthly series.

To determine the best ARIMA model for the monthly series, several candidate models with varying p and q orders (between zero to three) have been analysed by their corrected Akaike Information Criterion (AIC) value, as shown in Table IV. Since the monthly series was found stationary as seen in Table III and no differencing was required so it could be called as ARMA model, the order of d was left to remain at zero. Subsequently, XLSTAT was used to compute the AIC value of each model, whereby the best model would have the lowest AIC value. Based on the results in Table IV, the best model was found to be the ARIMA(1,0,3) model taking the “d” as zero interms of ARIMA definition and the least AIC value of 5,229.819.

The selected model was then fitted to examine how well it was able to resemble the original series. Fig. 5 presents a comparison between the original time series plot and the fitted plot (series2) ARIMA(1,0,3) model. Based on assessment, the plot of the fitted series could be seen having a rather close resemblance to the original series, particularly in terms of wave pattern and R^2 value obtained is 0.78. This level of similarity was a fair indication that the fitted model would be able to produce satisfactory forecasts.

TABLE IV: ARIMA MODEL FOR ANNUAL SERIES (PEJABAT JPS KLANG RAINFALL STATION)

ARIMA (p,d,q) Model	AIC value
(1,0,0)	5398.061
(1,0,1)	5333.356
(1,0,3)	5229.819
(2,0,0)	5338.750
(2,0,1)	5332.633
(2,0,2)	5253.448
(2,0,3)	5285.327
(3,0,0)	5323.766
(3,0,1)	5329.226
(1,0,0)	5398.061
(3,0,2)	5341.203
(3,0,3)	5276.468

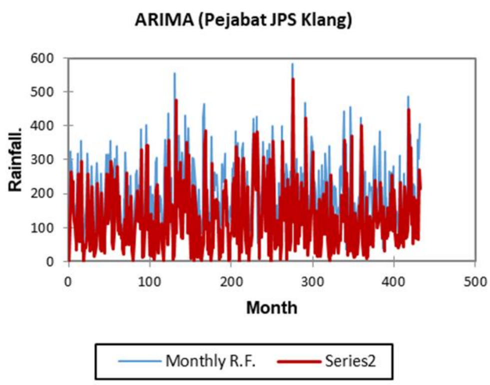


Fig. 5. Comparison of original series vs. fitted series for ARIMA(1,0,3).

C. Validation and forecasting for Annual Series

In order to reaffirm its forecast accuracy, the ARIMA (1,0,2) model had been subjected to a validation procedure, which involved comparison of the original annual series against a forecasted series produced by the said model. The validation procedure was carried out using existing rainfall data from the year 2015 to 2019 (5 years) – which would then reveal how closely the model was able to forecast values that replicate the original ones.

Based on Fig. 6, the plot of the forecasted series appeared to stray off from the actual series, with a noticeable gap that seems to widen across time. This observation suggests possible discrepancies in the model's forecast results, wherein the longer the forecast duration, the less accurate it becomes. Table V shows the rainfall values of both the original and predicted series. Overall, it can be deduced that the ARIMA(1,0,2) model may not be as precise or accurate in making long-term forecasts, but still somewhat acceptable and Ok since the gaps between actual and predicted values were not extremely large.

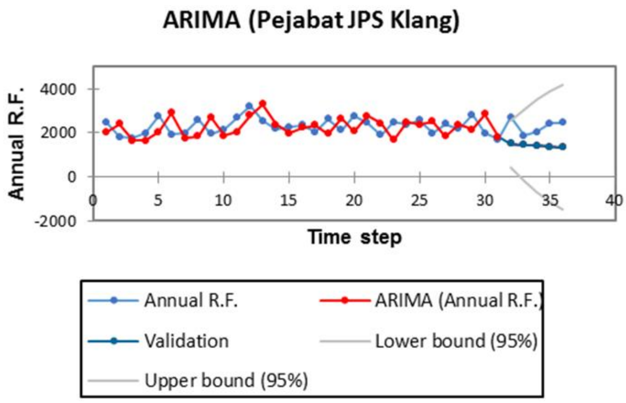


Fig. 6. Annual series validation from 2015–2019.

TABLE V: ANNUAL SERIES VALIDATION WITH 95% C.I.

Lead (Year)	Original R.F. Values (mm)	Predicted R.F. Values (mm)	Lower Limit	Upper Limit
2015	2713.500	1504.587	452.439	2556.735
2016	1878.500	1447.418	-182.198	3077.034
2017	2033.000	1409.239	-714.709	3533.188
2018	2420.000	1372.068	-1131.943	3876.079
2019	2456.500	1335.877	-1481.466	4153.220

Using the developed ARIMA (1,0,2) model, a long-term precipitation forecast for Pejabat JPS Klang had been made, stretching from the year 2020 to 2024 (5 years). Fig. 7 shows the annual forecasted series, with 95% confidence interval. The plot of the model's predicted rainfall values was seen to display very minimal fluctuations, whereby almost resembling a straight line. A similar study previously conducted by Nyatuame and Agodzo [14] also showed a somewhat similar outcome, with regards to the said rainfall pattern. In their study, this particular rainfall pattern (which appeared to become almost constant over time) was described exclusively as, "having a slow declining trend". However, they also acknowledged that a more comprehensive multivariate analysis using digitized data would be needed for further fine-tuning the results of their forecast. This statement subtly highlighted that the results of the annual rainfall forecast in this study would equally require improvement, as well.

Table VI presents a list of the annual forecasted values, including the upper and lower limit values at 95% confidence interval. The confidence interval (otherwise known as 'prediction interval'), assures the relevance of predicted data by keeping them well within the range of the confidence limit. For instance, year 2024 anticipated a total of 2,288.618 mm rainfall – which can be considered relevant, since the value falls between the upper limit (3,178.295 mm) and lower limit (1,398.940 mm). Furthermore, a similar compliance was also observed for the rest of the predicted data set, from the year 2020 to 2023.

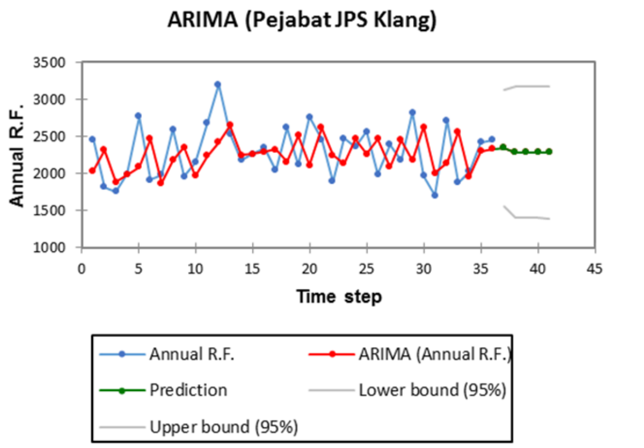


Fig. 7. Annual forecasted series for 2020–2024 (5 years).

TABLE VI: ANNUAL FORECASTED VALUES WITH 95% C.I. FOR PEJABAT JPS KLANG

Lead (Year)	Forecasted R.F. Values (mm)	Lower Limit	Upper Limit
2020	2343.950	1557.017	3130.882
2021	2289.158	1404.757	3173.559
2022	2288.978	1402.814	3175.142
2023	2288.798	1400.876	3176.720
2024	2288.618	1398.940	3178.295

D. Model Validation and Forecasting for Monthly Series

In order to reaffirm its forecast accuracy, the ARIMA (1,0,3) model had been subjected to a validation procedure, which involved comparison of the original monthly series against a forecasted series produced by the said model. The validation procedure was carried out using existing rainfall data from January 2019 to December 2019 (12 months) – which would then reveal how closely the model was able to forecast values that replicate the original ones.

Based on Fig. 8, the plot of the forecasted series appeared to be satisfactorily close to the original series. In addition, the original and forecasted values from Table VII were also within acceptable range, especially for the months of April, May and June. These observations suggest that the model was sufficiently capable of producing satitisfactory forecasts. Hence, it can be deduced that the ARIMA(1,0,3) model was fairly reliable and convincing for making short-term forecasts, and therefore acceptable.

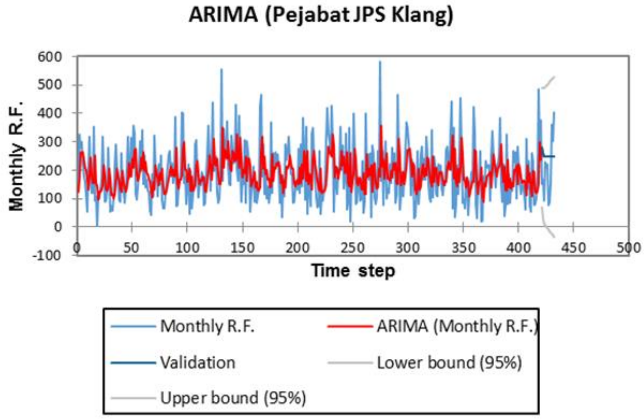


Fig. 8. Monthly series validation from January 2019 to December 2019.

TABLE VII: MONTHLY SERIES VALIDATION WITH 95% C.I.

Lead (Month, 2019)	Original R.F. Values (mm)	Predicted R.F. Values (mm)	Lower Limit	Upper Limit
January	149.000	278.199	69.274	487.124
February	165.000	259.697	30.492	488.902
March	93.500	249.504	11.409	487.599
April	228.000	249.126	5.911	492.341
May	208.000	248.748	0.535	496.962
June	221.000	248.371	-4.728	501.471
July	78.500	247.995	-9.883	505.873
August	88.500	247.619	-14.937	510.175
September	156.500	247.244	-19.894	514.382
October	359.500	246.869	-24.760	518.498
November	304.000	246.495	-29.539	522.529
December	405.000	246.121	-34.236	526.478

Using the developed ARIMA(1,0,3) model, a short-term precipitation forecast for Pejabat JPS Klang had been made, stretching from January 2020 to December 2020 (12 months). Fig. 9 shows the monthly forecasted series, with 95% confidence interval. On the other hand, Table VIII presents a list of the annual forecasted values, including the upper and lower limit values at 95% confidence interval. All of the forecasted monthly data were situated well within their confidence limit. Aside from the rainfall transitions between January (252.851 mm), February (216.871 mm) and March (192.459 mm), the rest of the predicted data set have been seen to display very minor fluctuations.

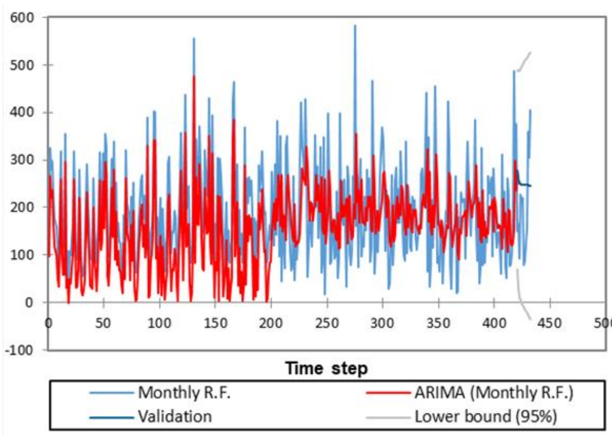


Fig. 9. Monthly forecasted series for 2020 (12 months).

TABLE VIII: MONTHLY FORECASTED VALUES WITH 95% C.I. FOR PEJABAT JPS KLANG

Lead (Month, 2020)	Forecasted R.F. Values (mm)	Lower Limit	Upper Limit
January	252.851	54.381	451.320
February	216.871	8.090	425.653
March	192.459	-18.988	403.906
April	192.440	-19.079	403.958
May	192.420	-19.169	404.009
June	192.401	-19.259	404.061
July	192.382	-19.349	404.112
August	192.362	-19.439	404.164
September	192.343	-19.529	404.215
October	192.324	-19.619	404.267
November	192.304	-19.709	404.318
December	192.285	-19.799	404.369

IV. CONCLUSION

All in all, the purpose of this study had been fulfilled. ARIMA models were developed for making short-term (monthly) and long-term (annual) precipitation forecasts using past rainfall data requested from DID, for the selected rainfall station. The short-term precipitation forecast extended from January, 2020 to December, 2020 (12 months) for the aforementioned study area, using the developed ARIMA(1,0,3) model. To develop the model, monthly rainfall data from 1984 to 2019 (36 years) have been used. The ARIMA(1,0,3) model was chosen as it had the least AIC value among all the other tentative models, as well as having shown to possess a good enough fit when its predicted rainfall values were fitted against the original rainfall values. Furthermore, the outcome of the model's validation also exhibited satisfactory results, whereby the model-predicted rainfall values were observed to be rather close to the original rainfall values. This implied that the monthly model was able to produce fairly accurate and acceptable forecasts. Overall, the results of the short-term precipitation forecasting revealed a somewhat slow, decreasing trend – from the highest recorded value (252.851 mm) in January, 2020 to the lowest (192.285 mm) in December, 2020. Moreover, the predicted rainfall values also showed very minimal fluctuations between the months of March, 2020 up to December, 2020.

On the other hand, the long-term precipitation forecast extended from the year 2020 to 2024 (5 years) for the aforementioned study area, using the developed ARIMA(1,0,2) model. To develop the model, annual rainfall data from 1984 to 2019 (36 years) have been used. The ARIMA(1,0,2) model was chosen as it had the least AIC value among all the other candidate models, as well as having shown to possess a poorly fit when its predicted rainfall values were fitted against the original rainfall values. However, the outcome of the model's validation indicated possible discrepancies, since the plot of the model-predicted rainfall values appeared to increasingly stray off from the actual rainfall values over time. This could imply that the annual model was not able to perform poorly Or, as adequately as desired, in terms of making reliable and accurate forecasts. On the contrary, this was consistent with the claim made by a previous study, whereby ARIMA models were better suited for short-term forecasting. Overall, the results of the long-term precipitation forecasting revealed a somewhat slow, decreasing trend – from the highest recorded value (2,343.950 mm) in 2020, to the lowest (2,288.618 mm) in 2024. Moreover, the predicted rainfall values also showed very minimal fluctuations between the years 2022 up to 2024.

Before being employed for practical uses, both the monthly and annual models should ideally be further primed and improved in order to obtain more accurate and reliable predictions that meet higher expectations. Despite some of their shortcomings, the models, and perhaps also their forecasted results, may somehow be partly useful for supporting future research with regards to flood prediction, flood administration, mitigation strategies and urban planning particularly at the chosen study location. The scope of this study to see the ARIMA outcomes to predict the rainfall for not only short-term but long-term as well because applications of this type of model were done in very limited numbers of researches in past for the study area. Comparative study with SARIMA or other models will be considered while conducting further studies. Aside from that, this study and its findings might help in shedding more light on the potentials of ARIMA modelling for precipitation forecasting in Malaysia, as it may be used as a guide or reference for other similar studies in the future. Additionally, to improve the results of model validation, it is highly encouraged to use a separate set or period of data for model development and model validation, respectively.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

M.M.H.K.: Conceptualization & methodology; M.S.H, M.R.U.M. & A.D.J.: formal analysis; M.M.H.K & A.D.J.: investigation; M.M.H.K.: writing—original draft preparation; M.M.H.K, M.R.U.M, M.S.H., S.S.: writing—review and editing; S.S.: visualization; All authors have read and agreed to the final version of the manuscript.

Funding

This work was supported by the seeding grant (INTI-FEQS-02-02-2021) of INTI International University, Malaysia.

Acknowledgment

The authors would like to thanks INTI University, Malaysia for providing fund INTI-FEQS-02-02-2021, and DID, Malaysia for data delivery.

References

[1] Malaysian Meteorological Department. (2019). Weather Forecast: Current Monsoon Report. [Online]. Available: http://www.met.gov.my

[2] S. M. Saiful, M. Z. Asmawi, M. R. Mispan, R. Ahmad, S. Z. Zulkifli, and N. J. Shafie, "Pollution sources identification of water quality using chemometrics: A case study in Klang River Basin, Malaysia," *International Journal of Engineering & Technology*, vol. 7, no. 4.43, pp. 83–89, 2018. https://doi.org/10.14419/ijet.v7i4.43.25824

[3] P. N. A. Kandari, M. R. Mispan, M. Z. Asmawi, R. Ahmad, S. Z. Zulkifli, and N. J. Shafie, "Flood risk pattern recognition analysis in Klang River Basin," *International Journal of Engineering & Technology*, vol. 7, no. 4.34, pp. 86–90, 2018.

[4] Asian Development Bank, Klang River Basin Environmental Improvement and Flood Mitigation Project Completion Report, 2007.

[5] S. H. Bari, M. T. Rahman, M. M. Hussain, and S. Ray, "Forecasting monthly precipitation in Sylhet city using ARIMA model," *Civil and Environmental Research*, vol. 7, no. 1, pp. 69–77, 2015.

[6] M. Mahsin, Y. Akhter, and M. Begum, "Modeling rainfall in Dhaka division of Bangladesh using time series analysis," *Journal of Mathematical Modelling and Application*, vol. 1, no. 5, pp. 67–73, 2012.

[7] D. Dayal, S. Swain, A. K. Gautam, S. S. Palmate, A. K. Pandey, and S. K. Mishra, "Development of ARIMA model for monthly rainfall forecasting over an Indian River Basin," World Environmental and Water Resources Congress 2019, 2019.

[8] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control. Hoden-Day*, 1976.

[9] Y. Liu, B. Wang, H. Zhan, Y. Fan, Y. Zha, and Y. Hao, "Simulation of nonstationary spring discharge using time series models," *Water Resources Management*, vol. 31, pp. 4875–4890, 2017. https://doi.org/10.1007/s11269-017-1783-6

[10] P. Unnikrishnan and V. Jothiprakash, "Hybrid SSA-ARIMA-ANN model for forecasting daily rainfall," *Water Resources Management*, vol. 34, pp. 3609–3623, 2020. https://doi.org/10.1007/s11269-020-02638-w

[11] B. Karimi, M. J. S. Safari, A. D. Meher, and M. Mohammadi, "Monthly rainfall prediction using ARIMA and gene expression programming: A case study in Urmia, Iran," *Online Journal of Engineering Sciences and Technologies (OJEST)*, vol. 2, no. 3, pp. 8–14, 2019. https://doi.org/10.21859/ojest-02032

[12] M. A. Rahman, L. Yunsheng, and N. Sultana, "Analysis and prediction of rainfall trends over Bangladesh using Mann-Kendall, Spearman's rho tests and ARIMA model," *Meteorology and Atmospheric Physics*, vol. 129, no. 4, pp. 409–424, 2017.

[13] S. Swain, S. Nandi, and P. Patel, "Development of an ARIMA model for monthly rainfall forecasting over Khordha District, Odisha, India," *In Recent Findings in Intelligent Computing Techniques. Advances in Intelligent Systems and Computing*, Springer, Singapore, 2018. https://doi.org/10.1007/978-981-10-8636-6_34

[14] M. Nyatuame and S. K. Agodzo, "Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana," *Journal of Water and Land Development*, vol. 37, pp. 127–140, 2018. https://doi.org/10.2478/jwld-2018-0032

[15] P. E. Naill and M. Momani, "Time series analysis model for rainfall data in Jordan: Case study for using time series analysis," *American Journal of Environmental Sciences*, vol. 5, no. 5, pp. 599–604, 2009.