Development of Short-term Flood Forecast Using ARIMA

Wei Ming Wong
Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer,
Universiti Teknikal Malaysia Melaka (UTeM) Melaka, Malaysia

Mohamad Yusry Lee
Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer,
Universiti Teknikal Malaysia Melaka (UTeM) Melaka, Malaysia

Amierul Syazrul Azman
Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer,
Universiti Teknikal Malaysia Melaka (UTeM) Melaka, Malaysia

Lew Ai Fen Rose
Fakulti Pengurusan Teknologi dan Teknousahawan
Universiti Teknikal Malaysia Melaka (UTeM) Melaka, Malaysia

Abstract— **The aim of this study is to use the Box-Jenkins method to build a flood forecast model by analysing real-time flood parameters for Pengkalan Rama, Melaka river, hereafter known as Sungai Melaka. The time series was tested for stationarity using the Augmented Dickey-Fuller (ADF) and differencing method to render a non-stationary time series stationary from 1 July 2020 at 12:00am to 30th July 2020. A utocorrelation (ACF) and partial autocorrelation (PACF) functions was measured and observed using visual observation to identify the suitable model for water level time series. The parameter Akaike Information Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used to find the best ARIMA model (BIC). ARIMA (2, 1, 3) was the best ARIMA model for the Pengkalan Rama, with an AIC of 5653.7004 and a BIC of 5695.209. The ARIMA (2, 1, 3) model was used to produce a lead forecast of up to 7 hours for the time series. The model's accuracy was tested by comparing the original and forecast sequences by using Pearson r and R squared. The ARIMA model appears to be adequate for Sungai Melaka, according to the findings of this study. Finally, the ARIMA model provides an appropriate short-term water level forecast with a lead forecast of up to 7 hours. As a result, the ARIMA model is undeniably ideal for river flooding.**

Keywords—flood, forecast, ARIMA, Melaka, prediction

I. INTRODUCTION

Floods occur all over the world, and they are a serious concern. It is a natural disaster that has become an annual occurrence in many countries, including Malaysia, Indonesia, Brazil, the United States, and others. Floods can no longer be seen as a one-off occurrence because they are related to issues such as disease outbreaks, food shortages, and climate change. Floods continue to occur in Malaysia, despite the fact that the government has implemented a number of large-scale plans, such as the expansion of drainage systems[1]. Floods have been known to cause a lot of damage. They destroy everything in their path: houses, crops, cars, buildings, and anything else. Animals and people become entangled in the flowing water's current and are unable to escape until rescue attempts are made [2].

Hundreds of people have resulted in the death cause by flood in recent decades, either directly or indirectly. Furthermore, relative to all other natural disasters, flood is the most common natural disaster to life today[3].

Floods are a common and temporary condition for partial or full dryland inland. Tidal water floods flow from any water source due to unusual and rapid water accumulation. There are three forms of flood disasters that occur, include monsoon, mud, and flash flood. Eventually, monsoon floods can be represented as flooding due to wind, which produces a lot of rain. mud flood occurs when the mud flows as rainwater, hence causing the mud to be filled by water. Flash floods happen as the drainage network deteriorates in urban environments. When there is heavy rain, the drainage fails to discharge the water quickly and causes the water to overflow.

One method of reducing flood risks and destruction is to use flood forecast. Existing people in flood-prone areas will be warned to evacuate themselves and their belongings before the flood comes. It would greatly reduce flood damage and human deaths.

Flood forecasting has become the subject of researchers around the world. The worst flood impacts have resulted in the implement of flood forecasting techniques. Various flood forecasts have been studied by researchers.

In this study [4], the author use 2 methods in flood forecast which are short-term radar-based extrapolation or longer-term numerical weather prediction for the flood forecast at the Gannam and Bakwoon which located within the Seoul area. The study predicted the observed rainfall for 1 hour ahead with 10 min intervals..

This paper explores ARIMA as a flood forecasting method, taking into account the integration of real-time series into the mathematical model structure. ARIMA has been used in several experiments to forecast floods. The ARIMA model was used in case [5] to create a flood forecasting method for the Segamat River. This scenario is similar to [6] , in which the ARIMA model was used to forecast floods in the Karkheh

river basin.

The aim of this study is to develop a flood prediction model by analyzing the real-time flood parameters for Pengkalan Rama, Melaka river hereafter known as Sungai Melaka using the Box-Jenkins method. Hourly water levels are predicted to alleviate flood related problems caused by the overflow of Sungai Melaka. This study was carried out using Matlab as the software tool for ARIMA modelling. The remaining section presents an approach to the development of the water level series. The ARIMA theories are discussed briefly in section 2. The method for designing the ARIMA model is discussed in Section 3, followed by section 4 that discusses the forecasted results. Finally, section 5 provides a summary of the conclusion.

II. LITERATURE REVIEW

A. Study Area

Melaka's Sungai Melaka River, also known as the Melaka River, is a historical landmark in western Malaysia. Melaka has a total area of 1720 km^2 and a population of 0.93 million people in 2019[7]. Sungai Melaka's flow diagram is shown in Fig. 1 Melaka's topography varies from flat to undulating, with broad coastal plains. To the south of Gunung Ledang, the hills of the Titiwangsa, the largest mountain range in Southeast Asia that stretches from South Thailand to Northen Johor, surround the tapered edge.

In terms of the environment, the country's relatively urbanised nature, with population concentrations in Melaka City, has resulted in the loss of a significant number of natural forests in the immediate hinterland. It is actually one of the few countries with a natural forest density slightly above 3%. Owing to the state's low forest density, one of the key issues is water scarcity. Sungai Melaka is now safe enough for boating and other water sports, and it is a popular tourist destination[3].

Fig. 1. Map of Sungai Melaka and IFOS Location

B. ARIMA Model

Box and Jenkins developed the ARIMA model, which is also known as the Box- Jenkins method. The ARIMA model is a type of time-domain model that is widely used for fitting and forecasting time series with a temporal correlation. The ARIMA model is optimised for stationary time series and integrates d difference into the ARMA model [8].

Monthly rainfall [9][10], streamflow[11], and water level [12] are only a few of the fields where ARIMA has been used

for forecasting [13]. In the study [14], an ARIMA model was used to forecast real-time road traffic data.

ARIMA modelling employs three components that make up the general term for ARIMA (p, d, and q). These three elements, which are used in the order of p, d, and q, are autoregressive (AR), integrated (Differencing), and moving average (MA). Differencing is a function that converts non-stationary series to stationary series, with the AR and MA terms defined by time series temporal correlation [15][16].

Model identification, parameter estimation, diagnostic testing, and prediction are the four phases of ARIMA modelling [17]. ARIMA stands out from the crowd. It is commonly used for short-term forecasts because of its ability to recognise complex patterns in temporary datasets [12]. The efficiency of ARIMA in predicting either a linear or non-linear sequence of intervals is satisfactory, according to a study in [18]. It's also useful for forecasting inter-valued time series.

C. Augmented Dickey-Fuller Test

The ADF (Augmented Dicky-Fuller Test) is used to test the null hypothesis that the time series contains a unit root. This test's aim is to look for seasonal variation, variance, and trend [19]. Time series are trend-stationarity, according to the alternate hypothesis (H0) in ADF. As a result, in ADF, the null hypothesis (H1) denotes that the series is non-stationary. The P-value is calculated at a significance level of 5%, or 0.05 [20]. As a result, if the series is non-stationary, differencing is needed to transform it into a stationary series. As a result, if the series is stationary and can be modelled using the ARMA model, no differencing is needed.

In the study, the ADF statistics are a negative number. The null hypothesis that there is a unit root at any level of trust would be rejected to a greater extent if the findings are negative [21].

D. L-Jung Box Test

The Ljung-Box test is used to determine if the residual is in a random number sequence. The Ljung -Box test tests the null hypothesis, H0: The model has no lack of fit, against the alternative hypothesis, H1: The model has a lack of fit.

The model with a significant value of less than 5% or 0.05 for P was used to describe the existence of any structure in the observed sequence. As a result, if the model has a large P value, it indicates that the model is unfit. A model with a P value greater than 0.05 has a residual that is representative of white noise and is considered adequate[8].

E. AIC and BIC

The Akaike Information Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are criteria for determining the accuracy and goodness of statistical model fitting, as well as efficient methods for evaluating the p and q orders[22].

The accuracy of the model can be determined by using the term Pearson R and R squared for the best fit.

$$AIC = 2k - 2\ln(L) \qquad (1)$$

$$BIC = -2\ln(L) + k\ln(n) \qquad (2)$$

$$r = \frac{\sum(\chi_i - \bar{\chi})(y_i - \bar{y})}{\sqrt{\sum(\chi_i - \bar{\chi})^2 \sum(y_i - \bar{y})^2}} \qquad (3)$$

$$Best\ Fit = r^2\ x\ 100 \qquad (4)$$

The AIC and BIC equations are shown in (1) and (2). L is the log-likelihood in the maximum value of the formula, n is the sample size of the sequence, and k is the number of parameters measured in the model in the equations AIC and BIC. The equation for the Pearson correlation coefficient, r, was shown in (3). This equation was used to examine the scatter data points in order to find the best fit axis.

To analyse both the data's correlation and results, the regression value of the actual water level and forecast water level data was evaluated according (4), these evaluations were determined in the best fit.

III. METHODOLOGY

A. Data

The data use for the study is obtain from the device develop by Universiti Teknikal Malaysia Melaka (UTeM) at the location indicate in Fig. 2 The data collection from the Internet of Things (IoT) Flood Observation System (IFOS) is transmit wireless to the cloud and monitor through online[23]. IFOS not only function as the data collection but also as the water level monitoring system to monitor the water level through the wireless sensor node and alert the authority when the water level is high[24], [25]. In Figure 2 shown the water level data collect at the Pengkalan Rama Jetty, Sungai Melaka with the coordinates of 2° 12'30.3 "N 102° 15'02.8 "E. Data is obtained for this analysis from 1 July 2020 at 12:00am to 30th July 2020 at 7.15am with a 15 minutes time interval.

Fig. 2. IFOS Location in Satellite View

In the preparation of the ARIMA modelling, the water data is collect is ongoing the pre-processing to remove the outlier and filter the noise. The pre-processing procedure are crucial to avoid the modelling process become the disturbance by the various of noise.

In the data collection, there are total of 2810 data. The modelling part of the analysis is divided into training and validation. A total of 2782 data samples that were used for the forecast starting from 12.00 am on the 01 July 2020 till on 29 July 2020 with 15 minutes interval. Meanwhile, the data samples were used for the testing, starting from 12.15

am on 30 July 2020 till 7.15 am on 30 July 2020 with total of 28 data.

B. Plotting The Time Series: ACF and PACF

The key strategies for model identification for autocorrelation (ACF) function and partial autocorrelation (PACF) function are visually examining the series. The ACF and PACF were eventually used to analyze the behaviour and the stationarity of time series. A time series is stationary if its ACF is decreasing rapidly. ACF decay over time implies that the sequence must be differenced because it is non-stationary. Additional checks to validate the non-stationary existence should be performed.

The analysis of ACF and PACF plots led to the recognition of autoregressive (p) and moving-average (q) orders based on the physical observation. If there are autoregressive terms throughout the series, the ACF plot eventually gradually down, while the PACF is suddenly cut off after q lags. In that case, p was assumed to be the autoregressive term. The ACF cut off abruptly after q lags and the PACF plot slowed down when the sequence had moving-average terms. Then, in this case, the moving average term q was considered. Autoregressive and moving average terms slowly decline after a few lags indicate that the model are mix with both p and q term[11].

C. Stationarity Test

Time series stationarity are necessary in ARIMA modelling. Without the stationary of time series, ARIMA cannot be model. In case the time series is not stationary, it is necessary to transform the time series into stationary. Transformation time series into the stationary require to perform the differencing method to convert the time series from non-stationary into stationary time series.

Stationarity test is to examine the stationarity of the time series. The Augmented Dickey-Fuller (ADF) Test is a method to test the stationarity of the time series. If the water level series was tested to be non-stationary then a differencing was required to make the data stationary.

After the differencing was completed, the data was shown to be stationary and was proven correct with the ADF test, which proved that the null hypothesis was rejected with a P-value less than 5% or 0.05, indicating that the data is stationary. As such, when analysing time series that exhibits both trends and seasonality, non-seasonal and seasonal differencing both require to be applied as two separate operations, one after the other. Increasing the differencing to go beyond two differencing is not recommended because over-differencing could lead to unnecessary levels of time series dependency

D. ARIMA Model Identification

The aim of this step is to define a possible ARIMA model to represent the time series' behaviour. The ACF (autocorrelation function) and PACF (partial autocorrelation function) were used to identify the series' behaviour . The ACF and PACF were applied for the model's order evaluation. ACF and PACF knowledge was use to help in determining the types of models to be built. It was then decided to go with the final model, which was chosen using

the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). These parameters assist in identifying models because the models with the lowest criterion value are the best.

Once a suitable model was established, determination of ARIMA model paramater was achieved . The predicted values were calculated based of the maximum loglikelihood for the AR and MA in the model. To make sure that the AR and MA parameters were statistically relevant, they were examined.

The three parameters used in summarising an ARIMA model are AR (p), integration (d), and MA (q). The parameters p and q imply that the order or AR and MA are extracted from p and q, while the parameter d indicates that the time series needs to be differencing to ensure it is stationary. The number of p and q in ARIMA(p,1,q) can be determined as p = 1, 2, 3, and q = 1, 2, 3. The lowest AIC and BIC value determines the model that suits best.

E. Diagnostic Checks

Diagnostic checking is the next step in the ARIMA model building process. This technique necessitates as a test of the chosen model's suitability. To assess if the residuals are correlated with white noise, diagnostic statistical methods such as residual plots were used. The residuals are evaluated using the Ljung-Box method to assess the accuracy of the chosen model.

A set of synthetic time series would be used to compare with the original time series in the best model that passes the diagnostic test. The degree to which the synthetic series resembled the original data set is determined. The model is said to be in good fit if the synthetic series pattern is identical to the original series pattern.

F. Series Comparison and Forecasting

The forecasting model is developed as the final step of ARIMA modeling. As a result, the ARIMA model's calculated model parameters will be compiled and used to forecast future time series. In this analysis, the pearson R and determination coefficient (R^2) are used to compare the observed and forecast values to identify its correlation. These performance parameter also used to denote that the performance is moving in the same direction as the actual value.

IV. RESULTS

The outcomes of each analysis will be presented at this section. Fig. 3 depicts the time series of water levels. The plot for the Pengkalan Rama Jetty can be seen in the figure, and it show the inconsistent trend of the water level.
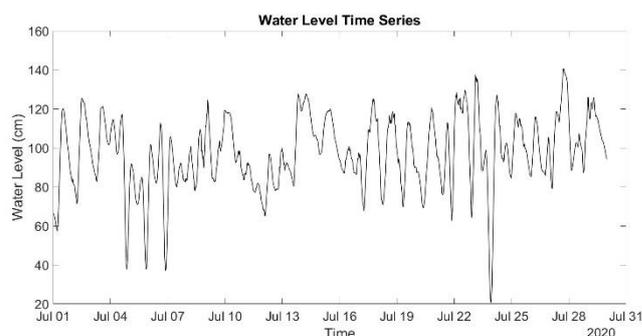
Fig. 3. IFOS Water Level Time Series Data

Only stationary time series can be used in ARIMA modelling. As a consequence, it's critical to demonstrate that the time series is stationary using the ACF and PACF plots..

A. Stationary Test

The ACF and PACF plots of the water level sequence are shown in Fig. 4 and Fig. 5. The ACF plot shows a slow decay, which means that the data may be non-stationary.
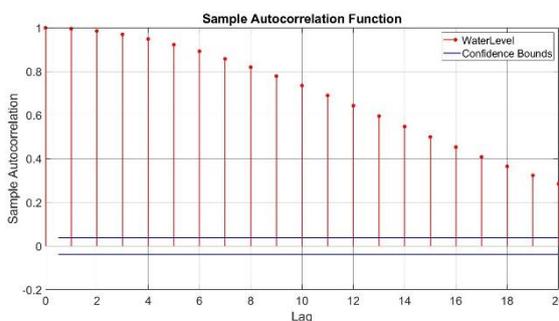
Fig. 4. ACF of Water Level Series
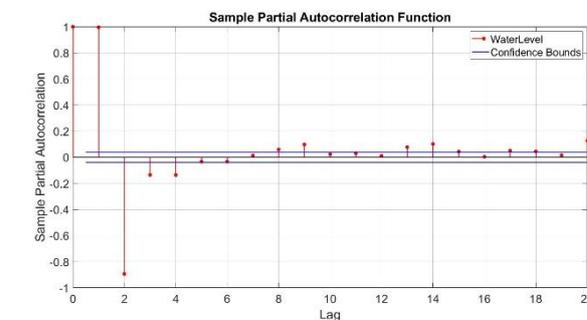
Fig. 5. PACF of Water Level Series

The water level series appears to be stationary in the ACF and PACF plot, but this is unconfirmed. To confirm the initial assumption that the time series was stationary, the Augmented Dickey-Fuller (ADF) test is use to justify it.

TABLE I. PACF OF WATER LEVEL SERIES

P-Value	Test Statistic	Critical Value	Remark
0.094496	-106453	-1.9416	Non-Stationary

The data is not stationary, as shown by the results in Table I. The ADF result indicates that the P-value is greater than 0.05, implying that the null hypothesis is accepted. The test

also confirmed that unit root present in the time series data and the time series is non-stationary..

B. Differencing The Series

The differencing method was applied to the water level series data twice to obtain the optimum value of d. Fig. 6 illustrates the output of the first-order differenced (d =1) and second-order differenced (d = 2) methods.
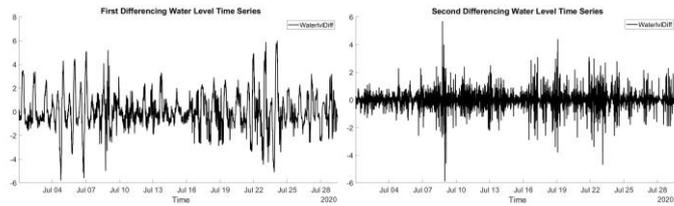
Fig. 6. First and Second Order Differenced
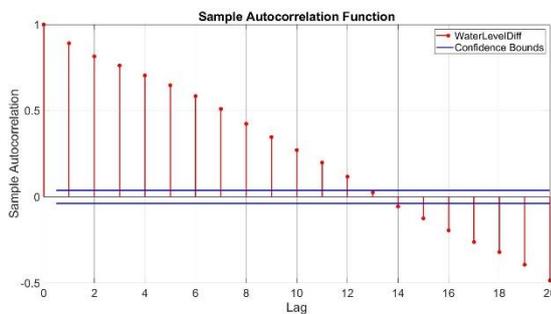
Fig. 7. First Order Differencing ACF
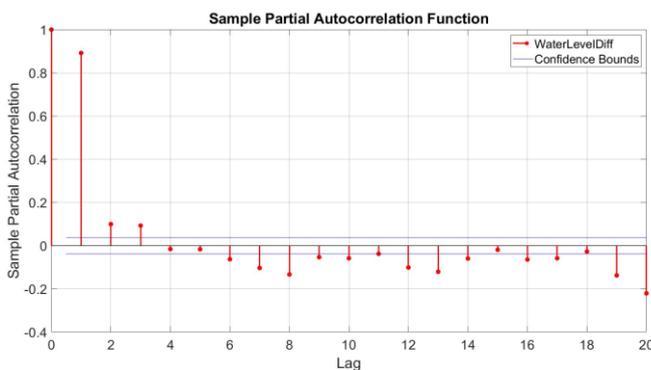
Fig. 8. First Order Differencing PACF
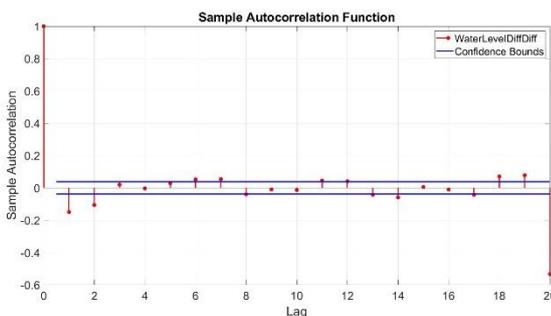
Fig. 9. Second Order Differencing ACF
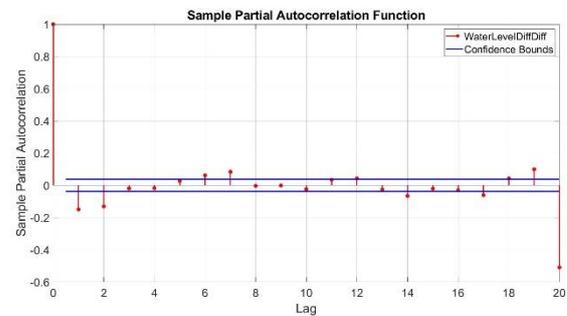
Fig. 10. Second Order Differencing PACF

The first-order differenced (d = 1) and second-order differenced (d = 2) models are plotted in the ACF and PACF plots in Fig. 7, Fig. 8, Fig. 9,and Fig. 10. At lag 1, both were compared. The second-order differenced model had a negative value and was less than -0.4, suggesting that the second-order differenced model was over-differenced [11]. As a result of comparing original water level time series, first-order differenced time series and second-order differenced models, the first order differenced model has positive value of lag 1 and the other lag are within the confidence bound, thus it is consider to be the best, and finalized value for d was one. The stationarity of the water level sequence needed to be checked with another examination. As a result, the stationary ADF test was replicated for validation. The P-value is less than 0.005, as shown by the findings in Table II. After first-order differencing (d = 1), this result shows that the water level series is stationary.

TABLE II.   STANDARD DEVIATIONS OF ORIGINAL AND DIFFERENCED SERIES

P-Value	Test Statistic	Critical Value	Remark
0.001	-11.1023	-1.9416	Stationary

After choosing the best differences, the order of the autoregressive term (p) and moving average term (q) parameters must be determined (d). The differenced series' ACF and PACF assist in deciding the order of p and q. To get the best model, several p and q values were also suggested.

C. ARIMA Modelling and Diagnostic Checking

The ARIMA model is widely used for short-term forecasts because it identifies complex patterns in temporal datasets. The numbers were varied for p = 1, 2 and q = 1,2 for the AIC and BIC calculation to determine the parameter of AR (p) and MA (q) term for ARIMA (p, 1, q) parameter. The model with the lowest AIC and BIC is therefore the best-fitting. Table III shows the AIC and BIC values. The table shows that ARIMA (2,1,3) is the best-fitting model, with an AIC of 5653.7004 and a BIC of 5695.209.

TABLE III.   AIC AND BIC OF ARIMA MODEL

ARIMA (p,d,q)	AIC	BIC
0.1.0	10264.7095	10276.57

0.1.1	8080.7074	8098.499
0.1.2	7220.0273	7243.75
0.1.3	6732.8959	6762.549
1.1.0	5866.3203	5884.111
1.1.1	5833.6714	5857.392
1.1.2	5823.5919	5853.243
1.1.3	5817.1748	5852.756
2.1.0	5840.0441	5863.764
2.1.1	5805.3194	5834.969
2.1.2	5745.9879	5781.567
2.1.3	5653.7004	5695.209
2.1.4	5747.6534	5795.092
3.1.0	5818.123	5847.771
3.1.1	5819.5637	5855.141
3.1.2	5821.4918	5862.998
3.1.3	5652.8662	5700.302

D. Goodness of Fit

The ARIMA (2,1,3) model was choosed to predict the water level time series. Figure 9 shows good fit of the original and predicted data blend together.
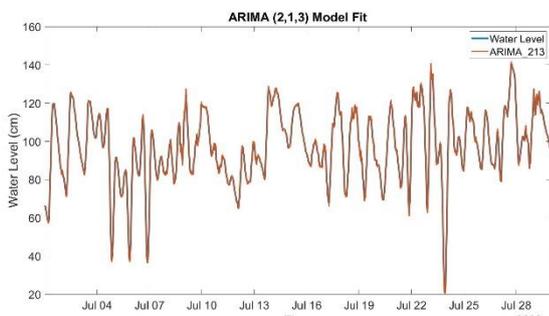
Fig. 11. Model Fit of the observed Water Level and predicted

TABLE IV.   ARIMA (2, 1, 3)

Parameter	Value	Standard Error	t Statistic	P-Value
Constant	0.00012	0.000794	0.15	0.8843
AR{1}	1.937	0.008228	235.42	0
AR{2}	-0.95155	0.007819	121.7	0
MA{1}	-1.2117	0.01531	79.15	0
MA{2}	0.07536	0.018168	4.15	3.358e-05

MA{3}	0.1982	0.01288	15.39	1.9614e-53
Variance	0.44456	0.007185	61.87	0

The variances were constants since the residuals were homoscedastic. For the residuals, homoscedasticity was critical because it determined whether the model was consistent in predicting variable values. Data transformation is required if a model with heteroscedastic residuals, because it cannot produce reliable results.

The residual histograms were visually depicted by plotting their distributions. Fig. 11 displays the residual histogram, which shows that the residual is normally distributed. The residual Q-Q plot depicts the residual's pattern, as seen in Fig. 12 and Fig. 13. The plot is usually distributed, as shown by the physical visual observation. The normality of the residual distribution is essential to obtain a satisfactory confidence interval.

Fig. 12. Residual Histogram

Fig. 13. Residual Q-Q Plot

TABLE V.   L-JUNG BOX TEST

P-Value	Test Statistic	Critical Value
0.6673	0.1847	3.4815

Finally, the L-Jung Box test was used to confirm the good fit of the model between water level time series and forecast water level time series. Table V shows these findings. The P-value for the test is greater than 0.005, implying that the test result is significant, as shown in the table. This demonstrates that the ARIMA model accurately predicts the original and forecasted water level series data.

4.5 Forecast result

To check for accuracy, the forecast time series created by the ARIMA model was compared to the original series. The water level series and forecast time series produced by ARIMA (2,1,3) are shown in Fig. 14. It display the contrast of the forecast time series to the water level time series, which starts on 30 July 2020. The Pearson R and R-squared show the statistical accuracy of the forecast.
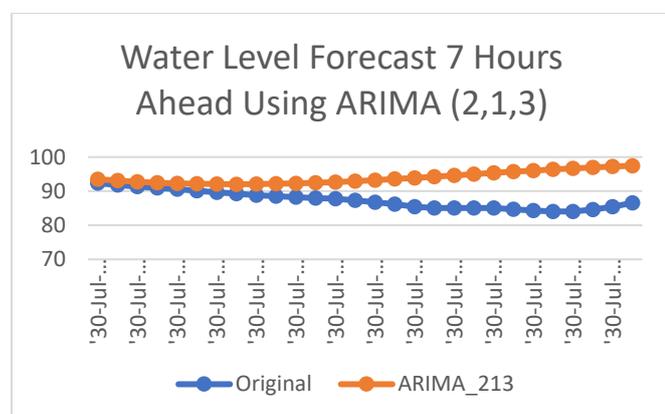
Fig. 14. Original Series and Forecast Series

Tables VI demonstrate the Pearson R and R squared used to determine forecast correlation and accuracy. The result indicate several lead times ranging from 1 to 7 hours. The result demonstrate that the best forecast ahead time was 1 hours with a 99.8523 percent accuracy. As the forecast duration increasing, the forecast accuracy are decreasing.

TABLE VI.   R-SQUARED

Time	R	R^2
1 hour	0.999261	99.8523
3 Hour	0.769121	59.15464
5 Hour	-0.62022	38.46729
7 Hour	-0.73993	54.74961

V. CONCLUSION

On the analysis of Sungai Melaka, the autoregressive integrated moving average (ARIMA) approach was used to carry out an efficient statistical modelling. A forecast series was also generated by the model, which produced sequences of future stage and water level values. The aim of this forecast is to forecast the imminent flood and give the response team enough time to plan. The ARIMA model is perfect for short-term forecasting because it can reliably predict the future. In the case of flash flooding, this short-term forecast may be used.

For the Sungai Melaka river in Pengkalan Rama Jetty, ARIMA modelling using the Box-Jenkins approach was found to be appropriate and suitable. The model's accuracy has been found to decrease as the forecast duration lengthens. The forecast values were 99.8523 percent of best fit in the flood forecast with a lead time of one hour. The forecast value

has the same correlation as the last reported observation.

The current model will serve as a base for future Sungai Melaka hydrological process analysis. The shortcomings of the ARIMA model can be mitigated by integrating other algorithms, such as the Artificial Neural Network (ANN), which take into account the nonlinearity and complexity of most time series prediction problems. This ARIMA simulation can be used in the study of real-time data series with real-time forecasting in the future.

REFERENCES

[1] M. I. Ibrahim, "KPKT sediakan RM25.9 juta bendung banjir kilat," *Berita Harian Online*, 2019.

[2] S. K. Subramaniam, V. R. Gannapathy, S. Subramonian, and A. H. Hamidon, "Flood level indicator and risk warning system for remote location monitoring using flood observatory system," *WSEAS Trans. Syst. Control*, vol. 5, no. 3, pp. 153–163, 2010.

[3] "Review of The National Water Resources Study (2000-2050) and Formulation of National Water Resources Policy-Volume 16-Melaka."

[4] S.-S. Yoon, "Adaptive Blending Method of Radar-Based and Numerical Weather Prediction QPFs for Urban Flood Forecasting," 2019, doi: 10.3390/rs11060642.

[5] N. H. Ab Razak, A. Z. Aris, M. F. Ramli, L. J. Looi, and H. Juahir, "Temporal flood incidence forecasting for Segamat River (Malaysia) using auto-regressive integrated moving average (ARIMA) modelling," *J. Flood Risk Manag.*, no. 11, pp. 56–63, 2016, doi: 10.5874/jfsr.1.56.

[6] K. Hamidi Machekposhti, H. Sedghi, A. Telvari, and H. Babazadeh, "Flood Analysis in Karkheh River Basin using Stochastic Model," *Civ. Eng. J.*, vol. 3, no. 9, pp. 794–808, 2017, doi: 10.21859/cej-030915.

[7] D. of S. Malaysia, "Melaka @ a Glance," Melaka, 2020.

[8] D. Xu, Q. Zhang, Y. Ding, and H. Huang, "Application of a hybrid ARIMA-SVR model based on the SPI for the forecast of drought—A case study in Henan Province, China," *J. Appl. Meteorol. Climatol.*, 2020, doi: 10.1175/jamc-d-19-0270.1.

[9] S. H. Bari, M. T. Rahman, M. M. Hussain, and S. Ray, "Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model," *Civ. Environ. Res.*, vol. 7, no. 1, pp. 69–78, 2015, [Online]. Available: http://www.iiste.org/Journals/index.php/CER/article/view/19069.

[10] Y. F. Huang, M. Mirzaei, and W. K. Yap, "Flood analysis in langat river basin using stochatic model," *Int. J. GEOMATE*, vol. 11, no. 5, pp. 2796–2803, 2016, doi: 10.21660/2016.27.1143.

[11] M. Reza, S. Harun, and M. Askari, "Streamflow Forecasting in Bukit Merah Watershed By Using Arima and Ann," *Portal J. Tek. Sipil*, vol. 9, no. 1, pp. 18–26, 2018, doi: 10.30811/portal.v9i1.612.

[12] Z. Yu, G. Lei, Z. Jiang, and F. Liu, "ARIMA modelling and forecasting of water level in the middle reach of the Yangtze River," *2017 4th Int. Conf. Transp. Inf. Safety, ICTIS 2017 - Proc.*, pp. 172–177, 2017, doi: 10.1109/ICTIS.2017.8047762.

[13] A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models: Literature review," *Water (Switzerland)*, vol. 10, no. 11, pp. 1–40, 2018, doi: 10.3390/w10111536.

[14] D. Xu, Y. Wang, L. Jia, Y. Qin, and H. Dong, "Real-time road traffic state prediction based on ARIMA and Kalman filter," *Front. Inf. Technol. Electron. Eng.*, vol. 18, no. 2, pp. 287–302, 2017.

[15] M. K. Douglas C. Montomery, Cheryl L. Jennings, *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons, Inc., Hoboken, New Jersey., 2008.

[16] S. Chiphang, "Study of Non Timber Forest Products (NTFPS) in Meghalaya State of India: Methods for Empirical Analysis," *Econ. Aff.*, vol. 65, no. 1, pp. 17–22, 2020, doi: 10.30954/0424-2513.1.2020.3.

[17] R. Faulina and Suhartono, "Hybrid ARIMA-ANFIS for Rainfall Prediction in Indonesia," *Int. J. Sci. Res.*, vol. 2, no. 2, pp. 159–162, 2013.

[18] A. L. S. Maia, F. D. A. T. De Carvalho, and T. B. Ludermir, "Forecasting models for interval-valued time series," *Neurocomputing*, vol. 71, no. 16–18, pp. 3344–3352, 2008, doi: 10.1016/j.neucom.2008.02.022.

[19] L. Ma, C. Hu, R. Lin, and Y. Han, "ARIMA model forecast based on EViews software," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 208, no. 1, 2018, doi: 10.1088/1755-1315/208/1/012017.

[20] F. S. Feroz, G. Leicht, J. Rauh, and C. Mulert, "The Time Course of Dorsal and Rostral-Ventral Anterior Cingulate Cortex Activity in the Emotional Stroop Experiment Reveals Valence and Arousal Aberrant Modulation in Patients with Schizophrenia," *Brain Topogr.*, vol. 32, no. 1, pp. 161–177, Jan. 2019, doi: 10.1007/s10548-018-0677-0.

[21] R. Rangamati and F. Ahmed, "Application of ARIMA Models in Forecasting Monthly Total Application of ARIMA Models in Forecasting Monthly Total Rainfall of Rangamati , Bangladesh," no. July 2017, 2018.

[22] L. Zhang *et al.*, "Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model," *Ecol. Indic.*, vol. 95, no. December 2017, pp. 702–710, 2018, doi: 10.1016/j.ecolind.2018.08.032.

[23] R. A. Ramlee, E. L. C. Yong, S. K. Subramaniam, A. Khmag, and A. S. F. Rahman, "Home switching using IoT system via telegram and web user interface," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 6, pp. 814–819, 2019, doi: 10.35940/ijrte.B1151.0782S619.

[24] M. Y. Lee, A. S. Azman, S. K. Subramaniam, and F. S. Feroz, "Performance Analysis of Linear Topology Wireless Sensor Network in Oil and Gas Industry," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 765, no. 1, 2020, doi: 10.1088/1757-899X/765/1/012070.

[25] A. S. A. Azman, M. Y. Lee, S. K. Subramaniam, and F. S. Feroz, "Performance Evaluation of Grid Wireless Sensor Network with Different Packet Size for Pipeline in Downstream of Oil and Gas Industry," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 765, no. 1, 2020, doi: 10.1088/1757-899X/765/1/012069.