ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse160942020.pdf
https://doi.org/10.30534/ijatcse/2020/160942020

Flood Prediction using ARIMA Model in Sungai Melaka, Malaysia

Wei Ming Wong[1], Siva Kumar Subramaniam[2], Farah Shahnaz Feroz[3], Indra Devi Subramaniam[4], Lew Ai Fen Rose[5]

[1,5]Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia, wongweiming5351@gmail.com

[2]Advance Sensors & Embedded Controls System (ASECS), Centre for Telecommunication Research & Innovation (CeTRI), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia

[3]Centre for Advanced Computing Technology (C-ACT), Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia

[4]Pusat Bahasa Dan Pembangunan Insan, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia

ABSTRACT

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.. The time series from 7 January 2020 12.00 am until 15 January 2020 8.00 am was used to check the stationarity by using the Augmented Dickey-Fuller (ADF) and differencing method to make a non-stationary time series stationary. The main methods used for model identification with autocorrelation (ACF) function and partial autocorrelation function (PACF) are visual observation of the series. The best ARIMA model was identified by the parameter Akaike Information Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The best ARIMA model for the Pengkalan Rama was ARIMA (2, 1, 2) with the AIC value 1297.5 and BIC value 1304.6. The time series had lead forecast up to 8 hours generated by using the ARIMA (2, 1, 2) model. The accuracy of the model was checked by comparing the original series and forecast series. The result of this research indicated that the ARIMA model is adequate for Sungai Melaka. In conclusion, ARIMA model is an adequate short term forecast of water level with the lead forecast of up to 8 hours. Hence, it is indubitable that the ARIMA model is suitable for river flood.

Key words : Flood, Prediction, Forecast, ARIMA, Melaka

1. INTRODUCTION

Flood is known to be world natural disaster that happen without prior warning, hence cause severe damage. They damage houses, crops, vehicles, schools, and everything that blocks their path. These include people as well as animal that before rescue efforts are made. While floods are unusual occurrence, they are now considered a life threat for humanity.

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[1]. Three forms of flood disasters that occur in Malaysia, include monsoon, mud, and flash flood. Eventually, monsoon floods can be represented as flooding due to wind, which produces a lot of rain. Many areas in Malaysia are generally affected by the monsoon flood, such in Johor, Melaka, Pahang, Kelantan, Terengganu, Sabah, and Sarawak[2]. Second, mud flood occurs when the mud flows as rainwater, hence causing the mud to be filled by water .In a situation like this there is shortage of trees to strengthen the soil, such an incident happen in Cameron Highlands in 2014[3]. In short, flash floods are sudden floods that are caused by heavy rains that occurr quickly in a short period of time. 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. Flash floods usually occur in urban areas such as in Selangor[4]. Flood able to devastate everything along its path. There were many ways used to prevent and monitor the flood like in the study [5] used floodgate to avoid the flood, while review [6] used a camera to monitor the flood. Unfortunately, it was difficult to avoid the flood. Thus it was essential to predict the incoming flood to prevent and get ready for it.

Flood prediction is one of the ways to mitigate flood risks and damage. Early flood predictions will warn people in flood-prone areas to evacuate themselves and their properties before the flood arrives. It will significantly minimize flood damage, and loss of human life, in particular, the approval for the implementation of the flood prediction program was obtained from the Malaysia Cabinet in 2001[7].
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 such as machine learning[8], SARIMA[9], NARX [10], and ARIMA[11].

With due consideration on the implementation of real-time series into the mathematical model structure, this paper discusses on ARIMA as the method used in flood forecasting. Many pat studies have used ARIMA for the prediction of floods. In case [12], the ARIMA model was used for the development of a flood forecasting system at the Segamat River. This scenario is similar to [7], that has use the ARIMA model for flood prediction at the Karkheh river basin. Furthermore, in [13], [14], the ARIMA method was used for traffic flow prediction. Besides, ARIMA has laso been found to be suitable for short-term forecasting and flash floods, as shown in study [15] that forecasted flash flood a day ahead. This is supported by study [16] that has used the ARIMA model for short-term forecasting at the Yangtze River.

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.

2. LITERATURE REVIEW

2.1 Study Area

The Sungai Melaka River or the Melaka River is located in Melaka, a historical landmark in western Malaysia. The total area of Melaka is 1720 KM2 with 0.93 million people in 2019[17]. Figure 1 shows the flow diagram of Sungai Melaka. Melaka's geographical topography is usually flat to undulate with vast coastal plains. Hills of the Titiwangsa, the largest mountain range from South Thailand to Northen Johor, to the south of Gunung Ledang, surround the tapered edge.

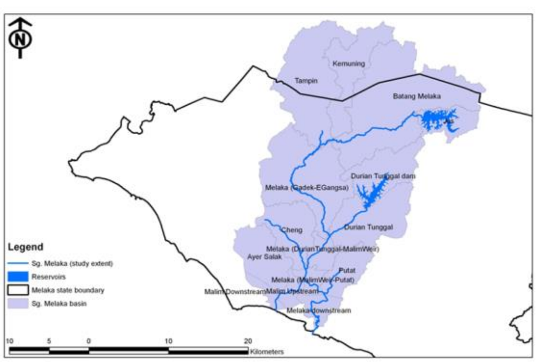

Figure 1: Map of Sungai Melaka and IFOS Location

In terms of the surroundings, the relatively urbanised nature of the country with population concentrations in Melaka City, the immediate hinterland has long since lost a large number of natural forests. It is currently one of the countries in which natural forest density is slightly higher than 3 per cent. Water shortages is one of the main problems in the state due to low forest density. Sungai Melaka now is clean enough to support boating and other water sports and is a key tourist attraction[18].

2.2 ARIMA Model

The ARIMA model is well known as the Box- Jenkins method developed by Box and Jenkins. ARIMA model can be applied to a class of time-domain models commonly used for time series fitting and forecasting with a temporal correlation. ARIMA model incorporates d difference into the ARMA model and is designed for stationary time series [19].

The ARIMA was implemented in the various fields for the forecast[20], e.g., monthly rainfall [21][11], streamflow[22], and water level[16]. The study [13] used an ARIMA model in the forecasting of real-time road traffic data prediction.

Three components encompassing of the general term for ARIMA (p, d, q) are used in ARIMA modelling. These three components are autoregressive (AR), integrated (Differencing), and moving average (MA) that are used in the respective order of p, d and q. Differencing is a function that converts the non-stationary series into a stationary series while the AR and MA terms are determined through the temporal correlation of time series[23], [24].

The four phases of ARIMA modelling include model identification, parameter estimation, diagnostic checking and prediction [25]. ARIMA is distinct from others. It has the potential to recognise complex trends in temporary datasets and is thus widely used for short-term predictions [16]. Study in [26] shows that the efficiency of ARIMA in predicting either a linear or a non-linear series of intervals is satisfying. It is also a good option for predicting inter-valued time series

2.3 Augmented Dickey-Fuller Test

Augmented Dicky-Fuller test (ADF) is used to check the null hypothesis that a unit root is present in the time series[27]. The function of this test is to check for the seasonal variation, variance and trend[28]. The alternate hypothesis (H_0) in ADF indicate that time series is trend-stationarity. Therefore, the null hypothesis (H_1) in ADF indicates that the series is non-stationary. The significance level used for the P-value is 5% or 0.05. Therefore, if the series is non-stationary, differencing is needed to convert the non-stationary into stationary series. Hence, no differencing is needed if the series is stationary and the series can be modelled using the ARMA model[11].

The ADF statistics are a negative number used in the analysis. The more negative the results, the greater will be the refusal of the null hypothesis that at some degree of confidence there is a unit root [29].

2.4 Ljung-Box Test

The Ljung-Box test is used to check the residual that is whether it is in a random sequence of numbers. The Ljung -Box test is a test based on the null hypothesis, H_0: The model does not exhibit a lack of fit, against the alternate hypothesis H_1: The model exhibits a lack of fit[30].

To classify the presence of any structure in the observed sequence, the model with a significant value of less than 5 per cent or 0.05 for P was considered. Hence, the model was not accounted for. Therefore, if the model has a significant P value, it shows that the model exhibit lack of fit. The residual of a model with P value more than 0.05 is indicative of white noise and is considered to be an adequate model[19].

2.5 AIC AND BIC

The Akaike Information Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [30] are standards for evaluating the accuracy and goodness of statistical model fitting and efficient methods for assessing the p and q orders [32].

The accuracy of the model can be determined by using the term Mean Absolute Percentage Error (MAPE).

$$AIC = 2k - 2\ln(L)$$
$$BIC = -2\ln(L) + k\ln(n)$$
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_{i-}\bar{x}_i}{x_i}\right| * 100 \qquad (1)$$

(1) shows the equation for the AIC, BIC equation and MAPE. In the equation AIC and BIC, the L is the log-likelihood in the maximum value of the model, n is the sample size of series and k is the number of parameters that are calculated in the model [33]. Whereas, for the equation MAPE, x_i is the actual values of the i-th ,and ¯x_i is the forecast value of the i-th [24].

3. METHODOLOGY

3.1 Data

The strength of this study includes having a collection of data from the Internet of Things (IoT) Flood Observation System (IFOS), which is designed by Universiti Teknikal Malaysia Melaka (UTeM). IFOS function as the flood warning system and also the water level monitoring system. Figure 2 illustrates the location of data obtained at the Pengkalan Rama Jetty, Sungai Melaka with the coordinate of 2° 12’30.3 “N 102° 15’02.8 “E. For this study, data is taken from 7 January 2020 12.00 am until 15 January 2020 8.00 am with a 1-hour interval.

The daily water data in centimetre (cm) was collected and pre-processed to prepare for the ARIMA model development. The pre-processing started with transforming the conversion epoch UNIX time to convert the original data time and dates into the time-and-date format in excel.

Figure 2: IFOS Location in Satellite View

The modelling part of the analysis is divided into training and validation. A total of 168 data samples that were used for the forecast starting from 12.00 am on the 07 January 2020 till on 14 January 2020 with a one hour interval. Meanwhile, the data samples were used in the validation part, starting from 1.00am on 14 January 2020 till 8.00am on 15 January 2020.

3.2 Plotting The Time Series: ACF and PACF

The main methods used for model identification with autocorrelation (ACF) function and partial autocorrelation function (PACF) are visual observation of the series.[7]. The ACF and PACF were then accessed the series behaviour and stationarity. It was evident that, ACF and PACF were not significant and fell within the confidence band. The observations indicated that they were independent. The time series is a white noise process. In such a situation, no modelling could be carried out. A time series is stationary if it has a rapidly declining ACF. ACF slowly decay means that the series need to be undergo differencing since it is non-stationary. Further tests to confirm the non-stationary existence should be carried out [11].

The identification of autoregressive (p) and moving-average (q) orders were also based on the physical observation of the ACF and PACF plots. If the series has autoregressive terms, the ACF plot dies down slowly and the PACF is abruptly cut off after p lags. In that case, the autoregressive term p was considered. 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. The mixed model was considered to be used if both the autoregressive and the moving average terms slowly fell after a few lags.[22].

3.3 Stationarity Test

The ARIMA modelling required the time series to be stationary. In case of the non-stationary time series, to transform the time series into stationary, differencing is needed before the ARIMA modelling [11].

The Augmented Dickey-Fuller (ADF) Test could identify the stationary 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, the data became stationary, and it was confirmed by using ADF to prove that the test rejected the null hypothesis with the P-value less than 5% or 0.05, which indicates that the data is stationary. Therefore, in the scenario time series that contains both trend and seasonality, both non-seasonal and seasonal differencing need to be applied as two successive operations in either order. Differencing to go beyond two differencing is not advisable since over-difference can lead to unnecessary levels of dependence on time series data [14].

3.4 ARIMA Model Identification

This step is to define the possible ARIMA model that represents the behaviour of the time series. The series behaviour was analysed by the ACF (autocorrelation function) and partial autocorrelation function (PACF). The ACF and PACF were used for the evaluation of the model order. ACF and PACF knowledge was useful in deciding the type of models to be constructed. The final model was then selected using the Akaike information criterion (AIC) and Bayesian Information Criterion (BIC). Such parameters help sort models since the models with the lowest criterion value are the best.

After identifying the appropriate model as an initial step, the determination of model parameters was achieved. The predicted values were calculated using maximum likelihood for the AR and MA parts of the model. The AR and MA criteria were examined to ensure that they are statistically significant or not. The related parameters, such as standard error of estimates and associated t-values, are also calculated. The AR (p), integration (d), and MA (q) are three parameters used in summarising an ARIMA model. The order or AR and MA indicate from the parameter p and q, while the parameter d means the order of differencing to make the time series stationary [24]. To calculate the number of AR and MA lags from the ARIMA(p,1,q) model, the number of p and q varies with the number of p=1, 2, 3 and q = 1,2,3 were calculated. The lowest AIC and BIC value specifies the model that has the best fit{Formatting Citation}.

3.5 Diagnostic Checks

Next phase of building the ARIMA model is diagnostic checking. This process requires testing the appropriateness of the chosen model. Diagnostic statistical tools such as residual plots were analysed to identify whether the residuals are associated with white noise. The Ljung-Box method is used by analysing the residuals to determine the accuracy of the chosen model.

The best model that passes the diagnostic test would have a set of synthetic time series to compare with the original time series. It determines the degree to which the synthetic series resembled the original data set. If the synthetic series pattern is identical to the original series pattern, then the model can be said to be in a good fit.

3.6 Series Comparison and Forecasting

The last step of ARIMA modelling is to develop the forecasting model[35]. Therefore, the calculated model parameters of the ARIMA model will be compiled and used to predict future time series intervals. The parameter), root mean squared error (RMSE), Mean absolute percentage error (MAPE), and determination coefficient (R2) is used in this analysis to compare the observed values to the forecast values. This technique and parameter for finding the best ARIMA model are also used in [22].

4. RESULTS

The results for each analysis are going to be presented here. The water level time series are presented in Figure 3. In the figure, it can be observed that the plot for the Pengkalan Rama Jetty exhibits a relatively consistent trend for each day.

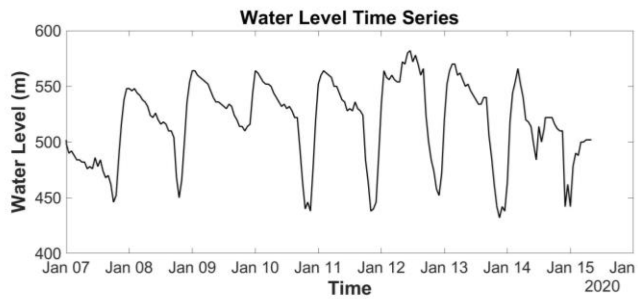

Figure 3: IFOS Water Level Time Series Data

ARIMA modelling only works with the stationary time series. Hence, it is important to prove that the time series is stationary from the ACF and PACF plot.

4.1 Stationary Test

Figure 4 and Figure 5 show the ACF and PACF plot of the water level series. The ACF plot indicates the slow decay, and this means the possibility of non-stationarity of the data.

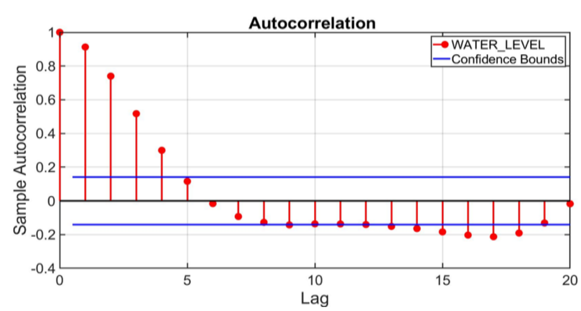

Figure.4: ACF of Water Level Series

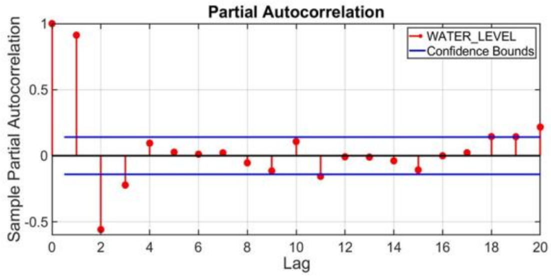

Figure.5: PACF of Water Level Series

From the illustration of the ACF and PACF plot, it is unconfirmed that the water level series is stationary. Thus, the Augmented Dickey-Fuller (ADF) test was carried out to confirm the initial presumption that the time series was stationary.

Table 1: ADF Result

P-Value	Test Statistic	Critical Value	Remark
0.12074	-3.055	-3.4342	Non-Stationary

Results presented in Table 1 show that the data is not stationary. The ADF result, it indicate that the P-value is more than 0.05, hence implying the acceptance of the null hypothesis. The test also confirmed that the unit root contained in the data and is non-stationary.

4.2 Differencing The Series

The differencing method was applied to the water level series data twice to obtain the optimum value of d. Figure 6 illustrates the output of the first-order differenced (d =1) and Figure 7 illustrates the second-order differenced (d = 2) methods. Therefore, the standard deviations of the original series and differenced series are shown in Table 2. From the table, it was discovered that the minimum standard deviations obtained was first-order differenced (d = 1) model with the value of 15.82 compared to second-order differenced (d = 2) model with the value of 16.77 and original (d = 0) with value 37.06.

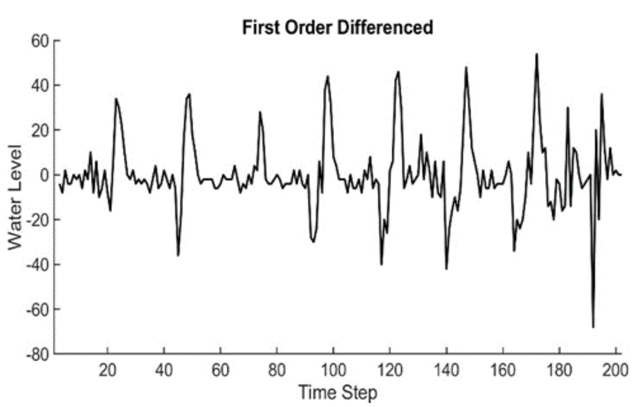

Figure.6: First Order Differenced Residual

Figure.7: Second Order Differenced Residual

Table 2: Standard Deviations of Original and Differenced Series

Order, d	Standard Deviation
0	37.06
1	15.82
2	16.77

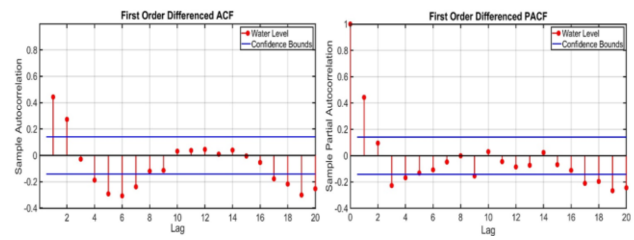

Figure.8: First Order Differencing ACF and PACF

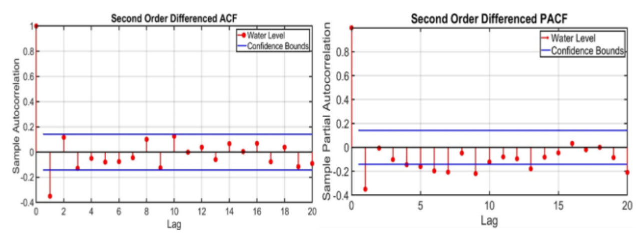

Figure.9: Second Order Differencing ACF and PACF

In the ACF and PACF plot as in Figures 8 and Figure 9 the first-order differenced (d = 1) and second-order differenced (d = 2) models. Both were compared at the lag 1. The physical visual observation of the figure indicates that the second-order differenced model was negative value and lower than -0.4, hence indicating that the second-order differenced model was over-differenced [11]. Hence, by comparing these three levels of differencing methods, the optimum level of differencing was the first order differenced model, and the value for d was one.

A further test was required for the confirmation of the stationarity of the water level series. Hence, the stationary ADF test was conducted again for confirmation. The results as shown in Table 3 confirms that the P-value is lower than 0.005. This result indicates that the water level series is stationary after first-order differencing (d = 1).

Table 3: Standard Deviations of Original and Differenced Series

P-Value	Test Statistic	Critical Value	Remark
0.001	-8.7252	-3.4343	Stationary

The order of autoregressive term (p) and moving average term (q) parameters must be determined after having selected the best differences (d). The ACF and PACF of the differenced series help identify the order of p and q. Several p and q values were also recommended to get the best model.

4.3 ARIMA Modelling and Diagnostic Checking

The ARIMA model recognises complex trends in the temporal dataset and is thus commonly used for short-term predictions. To determine the parameter of AR (p) and MA (q) term for ARIMA (p, 1, q) parameter, the numbers were varying for p = 1, 2 and q = 1,2 for the AIC and BIC calculation. The best-fitting model is thus determined by the lowest AIC and BIC. The AIC and BIC value are presented in Table 4. The table shows that the best fitting model is ARIMA (2,1,2) with the lowest AIC 1297.511 and BIC 1304.553.

Table 4: AIC and BIC of ARIMA Model

ARIMA (p,d,q)	AIC	BIC
(1,1,0)	1606.056	1613.099
(0,1,0)	1398.579	1405.622
(0,1,1)	1343.956	1350.999
(0,1,2)	1323.768	1330.811
(1,1,0)	1326.674	1333.717
(1,1,1)	1325.499	1332.542
(1,1,2)	1321.337	1328.38
(2,1,0)	1324.626	1331.669
(2,1,1)	1297.871	1304.914
(2,1,2)	1297.511	1304.553

4.4 Goodness of Fit

The ARIMA (2,1,2) model was implemented into the water level series data. The good fit of the original and forecast data is illustrated in Figure 10.

However, the ARIMA model can also be modelled by constant or without constant. Table 5 shows the ARIMA parameter without constant, and table 6 shows the parameter with constant. Findings from table 6 demonstrates that the P-value is more than 0.005. The result indicates that the ARIMA with constant is not significant. Thus, the most suitable model for the ARIMA model is without the constant.

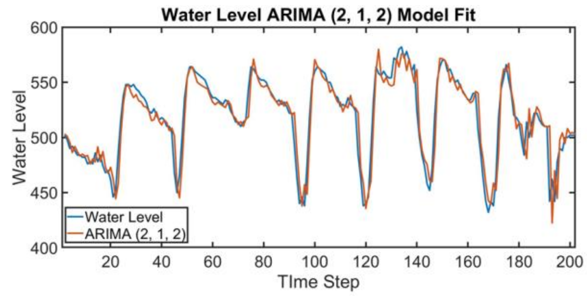

Figure.10: Model Fit of the observed Water Level and predicted ARIMA (2,1,2)

Table 5: ARIMA (2, 1, 2) Without Constant

Parameter	Value	Standard Error	t Statistic	P- Value
Constant	0	0		
AR{1}	1.6128	0.093392	17.2688	8.0875e-67
AR{2}	-0.7429	0.088158	-8.4269	3.5496e-17
MA{1}	-1.3093	0.11139	-11.7539	6.7414e-32
MA{2}	0.33543	0.11156	3.0067	0.002641
Variance	169.495	10.0085	16.9351	2.4775e-64

Table 6: ARIMA (2, 1, 2) With Constant

Parameter	Value	Standard Error	t Statistic	P-Value
Constant	0.0061896	0.032769	0.18889	0.85018
AR{1}	1.6131	0.094401	17.0877	1.8325e-65
AR{2}	-0.74284	0.087899	-8.4511	2.8864e-17
MA{1}	-1.31	0.111	-11.8014	3.8392e-32
MA{2}	0.33533	0.11119	3.0158	0.0025629
Variance	169.4488	11.3607	14.9154	2.6174e-50

The residuals were homoscedastic, which means that the variances were constants. Homoscedasticity was significant for the residuals, as it determined if the model was consistent in predicting variable values. A model with heteroscedastic residuals cannot produce reliable results, and transformation of data is necessary. The residual histograms were plotted to reflect their distributions visually.

Figures 11 depiction the residual histogram whereby the residual appears to be normally distributed. The trend of the residual as shown in Figures 12 is the residual Q-Q plot. The physical visual observation is indicate of the fact that the plot is usually distributed. To achieve a satisfactory confidence interval, the normality of the residual distribution is essential.

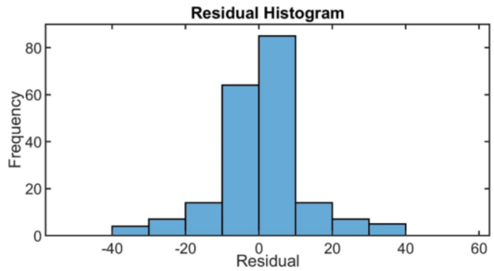

Figure.11: Residual Histogram

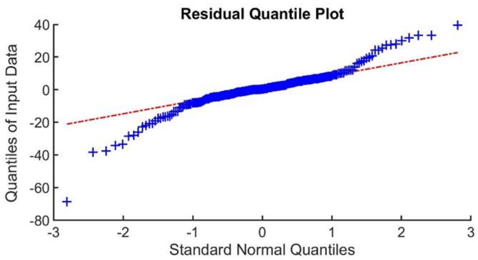

Figure.12: Residual Q-Q Plot

Table 7: L-Jung Box Test

P-Value	Test Statistic	Critical Value
0.51179	19.1547	31.4104

Lastly, the good fit of the model between actual and forecast was tested with the L-Jung Box test to confirm the good fit of the original and forecast water level series. These results are demonstrated in table 7. The table shows that the P-value for the test is more than 0.005, thus indicate that the test is significant. This brings to evidence that the ARIMA model is a good fit between the original and forecast of the water level series data.

4.5 Forecast result

The time series generated from the ARIMA model which is the forecast result was used to contrast with the original series to check for the accuracy. Both Figures 13 and Figures 14 show the water level series and forecast series generated by using ARIMA (2,1,2). Figures 4-12 illustrate the comparison between the forecast series with the original series starting from 15 January 2020. The statistical accuracy of the prediction is indicated by the R-squared, MAPE and RMSE.

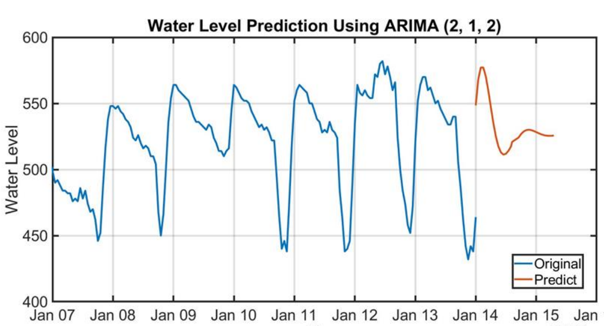

Figure.13: Forecast Series

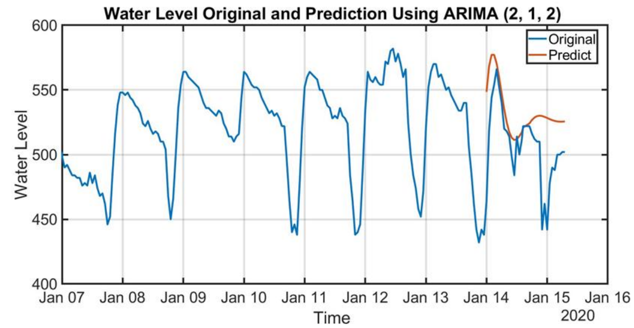

Figure.14: Original Series and Forecast Series

Results in tables 8 indicate the mean absolute percentage error (MAPE). The MAPE result from this model is significantly low. Table 9 shows the R square of the series, thus indicating multiple lead time from 2 hours until 8 hours.

In this study, the forecast required longer lead time with reasonable accuracy. Hence, in the result, the lead time of 6 hours with an accuracy of 94% was indicated. The RMSE of the model depicted the value as 13.0172.

Table 8: MAPE

Time	MAPE
2 hour	4.833354
4 hour	3.961335
6 hour	4.093351
8 hour	3.785191

Table 9: R-Squared

Time	R2	Percentage
2 hour	0.998338	99.8%
4 hour	0.935298	93.5%
6 hour	0.940688	94.1%
8 hour	0.879217	87.9%

5. CONCLUSION

The autoregressive integrated moving average (ARIMA) method was used to carry out an effective statistical modelling on the study of Sungai Melaka. The model also developed a forecast series to offer sequences of future stage and water level values. The purpose of this forecast is to predict the incoming flood as well as to enable the response team to have time for preparation. The ARIMA model is ideal for short-term forecasting as it could predict very well for short term. This short-term forecast can be used in the scenario of flash flooding.

The ARIMA modelling from the Box-Jenkins method was found to be acceptable and suitable for Sungai Melaka river in Pengkalan Rama Jetty. The accuracy of the model has been found to decrease as the forecast period increases. The value of the MAPE was close starting from 2 hours until 8 hours. The flood forecast with the lead time of eight hours shows that the forecast values were 87.9 per cent of best fit. The forecast value were identical to the observation that was last recorded.

The current model will provide a foundation for future hydrological process studies of Sungai Melaka. The ARIMA model's drawbacks can be overcome by adding other algorithms such as Kalman Filtering, taking into consideration of the nonlinearity and complexity of most time series prediction problems. This ARIMA simulation will be incorporated into the framework of future work to analyse the real-time data series with the real-time forecast.

ACKNOWLEDGEMENT

This work was conducted as part of W.M. Wong's postgraduate fellowship from Universiti Teknikal Malaysia Melaka (UTeM). The authors would like to thank and acknowledge the financial support from Kementerian Pengajian Tinggi (KPT) and the educational guidance from Universiti Teknikal Malaysia Melaka (UTeM).

REFERENCES

1. A. Kean Hua, "**Monsoon Flood Disaster in Kota Bharu, Kelantan Case Study: A Comprehensive Review**," *Int. J. Sci. Eng. Res.* , vol. 3, no. 9, pp. 2347–3878, 2014.
2. N. H. Rodzi, "**Floods hit several states in Malaysia; thousands evacuated**," *The Straits Time*, 16-Dec-2019.
3. A. Y. L. Ivan Loh, Chan Li Leen, "**Heavy rains cause mud flood in Camerons, one dead**," *The Star*, 06-Nov-2014.
4. T. P. Ying, "**Flash floods in Klang Valley after afternoon storm**," *New Straits Times*, 15-Aug-2019.
5. Rachmat and D. N. Utama, "**Water-flow-like algorithm based decision support model for flood control operation system**," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 3, pp. 3611–3618, 2020. https://doi.org/10.30534/ijatcse/2020/168932020
6. A. S. Alon, J. N. Mindoro, C. D. Casuat, M. A. F. Malbog, and J. A. B. Susa, "**An inference approach of flood level detection and alert system: Flood-induced simulated environment,**" Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 2, pp. 2259–2264, 2020.decision support model for flood control operation system," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3611–3618, 2020. https://doi.org/10.30534/ijatcse/2020/205922020
7. 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.
8. 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.
9. K. B. Tadesse and M. O. Dinka, "**Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa**," *J. Water L. Dev.*, vol. 35, no. 1, pp. 229–236, 2017.
10. A. Faruq, S. S. Abdullah, A. Marto, M. A. A. Bakar, S. F. M. Hussein, and C. M. C. Razali, "**The use of radial basis function and non-linear autoregressive exogenous neural networks to forecast multi-step ahead of time flood water level**," *Int. J. Adv. Intell. Informatics*, vol. 5, no. 1, pp. 1–10, 2019.
11. 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.
12. 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 autoregressive integrated moving average modelling**," *J. Flood Risk Manag.*, vol. 11, no. 2, pp. S794–S804, 2018.
13. 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. https://doi.org/10.1631/FITEE.1500381
14. S. V. Kumar and L. Vanajakshi, "**Short-term traffic flow prediction using seasonal ARIMA model with limited input data**," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, 2015.
15. H. Galavi, M. Mirzaei, L. T. Shui, and N. Valizadeh, "**Klang River-level forecasting using ARIMA and ANFIS models**," *J. Am. Water Works Assoc.*, vol. 105, no. 9, pp. 81–82, 2013.
16. 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.
17. **Melaka @ a Glance**," Melaka, 2020.
18. "**Review of The National Water Resources Study (2000-2050) and Formulation of National Water Resources Policy-Volume 16-Melaka**."
19. 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.
20. G. M. L. Box, George E. P.Gwilym M. Jenkins, Gregory C. Reinsel, ***Time Series Analysis : Forecasting and Control***, 5th ed. Canada: John Wiley & Sons, Inc., Hoboken, New Jersey., 2016.
21. 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.
22. 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.
23. M. K. Douglas C. Montomery, Cheryl L. Jennings, ***Introduction to Time Series Analysis and Forecasting***. John Wiley & Sons, Inc., Hoboken, New Jersey., 2008.
24. 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.
25. R. Faulina and Suhartono, "H**ybrid ARIMA-ANFIS for Rainfall Prediction in Indonesia**," *Int. J. Sci. Res.*, vol. 2, no. 2, pp. 159–162, 2013.
26. 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.
27. D. A. Dickey and W. A. Fuller, "**Distribution of the Estimators for Autoregressive Time Series With a Unit Root**," *J. Am. Stat. Assoc.*, vol. 74, no. 366, p. 427, 1979.

https://doi.org/10.2307/2286348

28. 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.
29. 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.
30. S. A. Sarkodie, "**Estimating Ghana's electricity consumption by 2030: An ARIMA forecast**," *Energy Sources, Part B Econ. Plan. Policy*, vol. 12, no. 10, pp. 936–944, 2017.
31. H. Akaike, "**A New Look at the Statistical Model Identification**," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.
32. K. P. Burnham and D. R. Anderson, "**Multimodel inference: Understanding AIC and BIC in model selection**," *Sociol. Methods Res.*, vol. 33, no. 2, pp. 261–304, 2004.
33. J. L. Ng, S. Abd Aziz, Y. F. Huang, A. Wayayok, and M. Rowshon, "**Generation of a stochastic precipitation model for the tropical climate**," *Theor. Appl. Climatol.*, vol. 133, no. 1–2, pp. 489–509, 2018.
34. A. Stitou, "**SARIMA short to medium-term forecasting and stochastic simulation of streamflow** , water levels and sediments time series from the HYDAT database," 2019.
35. 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. https://doi.org/10.1016/j.ecolind.2018.08.032