Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 2, 2021, page 79-88

Time Series Modeling and Forecasting of Monthly Rainfall using Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA)

[1]Hasanah Mustapha and [2]Norazlina Ismail

[1,2]Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia.

e-mail: [1]hsanahh98@gmail.com, [2]i-norazlina@utm.my

Abstract This paper aims to model and forecast monthly rainfall time series by using two models followed Box-Jenkins methodology which are Autoregressive Integrated Moving Average (ARIMA) model and Seasonal Autoregressive Integrated Moving Average (SARIMA) model based on the monthly rainfall data from January 2000 to December 2019 at two rainfall stations in Peninsular Malaysia, namely Sungai Petai in Kelantan and station Pam Kubang Haji in Perak. Forecasting is carried out of both rainfall stations using the best fitted models and the Root Mean Square Error (RMSE) is used to evaluate the forecast accuracy between two models.

Keywords Monthly rainfall time series; Box-Jenkins methodology; autoregressive integrated moving average; seasonal autoregressive integrated moving average; root mean square error; forecasting

1 Introduction

Malaysia is located close to the equator and characterized as a country with a hot climate throughout the year [1]. Malaysia is divided into two parts: Peninsular Malaysia, which lies between Thailand and Singapore in the North and the two Borneo states of Sabah and Sarawak in the South. According to the Department of Malaysian Weather, there are four seasons in Malaysia such as Northeast monsoon (November to March), Inter-monsoon 1 (April), Southwest Monsoon (June to September) and Inter-monsoon 2 (October). These two Inter-monsoon seasons will result in heavy rainfall, usually in the form of convective rain. The West coast is usually wetter than the East coast during these seasons. Apart from that, rainfall forecasting is crucial for resolving several regional environmental issues of effective national water resource management, with implications for agriculture, climate change and natural hazards such as floods and droughts.

Thus, this study is conducted to determine the best model for forecasting of monthly rainfall time series in Peninsular Malaysia's East and West coast regions by using two models, namely Autoregressive Integrated Moving Average (ARIMA) model and Seasonal Autoregressive Moving Average (SARIMA) model. The main objective of this study is to develop the model and obtaining the forecast value for the next 4 years. The forecasting performance between these two models are evaluated by using Root Mean Square Error (RMSE) in order to find the best fitted model for forecasting monthly rainfall time series to help the policy and decision makers established strategies for proper planning in future.

2 Literature Review

2.1 Rainfall in Peninsular Malaysia

Over the years, researchers have carried out a variety of studies on rainfall in Malaysia. In a study conducted by Suhaila et al. [2], they highlighted that four seasons, namely two monsoon seasons and two inter-monsoon seasons, characterize the climate of Peninsular Malaysia. The season of the South-West Monsoon (SWM) is from May to August, while the season of the North-East Monsoon (NEM) is from November to February. The exposed areas on the Eastern part of the Peninsula would receive heavy rainfall during the NEM. The regions sheltered by the mountain ranges (Titiwangsa Range) are more or less free from its influence. For the whole country, the SWM period is a drier period, particularly for the other states on the West coast of the Peninsula.

Other than that, Chen et al. [3] in a study of the winter rainfall in Malaysia stated that Malaysia is geographically separated into Peninsular Malaysia and West Borneo. The maximum rainfall occurs during November to December in the former region, while it occurs during December to February in the latter region. The highest rainfall occurs during November to December in the former area, while it occurs during December to February in the latter region. The maximum rainfall occurs approximately one month earlier in the West than in the latter area.

2.2 Autoregressive Integrated Moving Average (ARIMA) Model

Autoregressive Integrated Moving Average (ARIMA) model is one mathematical approach and also known as Box Jenkins method to forecast time series. A research was conducted by Katimon et al. [4] on the ARIMA model regarding of modeling water quality and hydrological variables in Johor River, Malaysia. In their study, they used ARIMA model because it can handle missing value in the observed data. Some assumptions have been made, such as the regular distribution of the data, but it may be stationary (mean and variance are constant over time) or non-stationary. The transformation process for the non-stationary sequence requires a method of differentiation.

Al Balasmeh et al. [5] attempted to forecast precipitation pattern in Wadi Shueib catchment area in Jordan by using ARIMA approach. They used daily precipitation data from five gauging stations. A data for a period of 34 years and 10 years were collected for model test data and model validated data respectively. They used t-test for the AIC or BIC to choose the best fit ARIMA model based on the p-value, which revealed that the ARIMA models are statistically significant within a 5% significance level. At the end of study, they found three ARIMA models to fit the majority of the monthly, average, and seasonal data well, with absolute errors ranging from 0.05 to 44.60 mm.

Somvanshi et al. [6] applied Artificial Neural Network and ARIMA techniques for rainfall modelling and prediction. They investigated that the first step is to decide if the time series is stationary or non-stationary in ARIMA (p, d, q) modelling. It is transformed into a stationary time series if it is non-stationary by applying the required degree of differentiation by choosing the proper value of d. The appropriate values of p and q are chosen by examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the time series.

2.3 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an extension of Autoregressive Integrated Moving Average (ARIMA) model that explicitly supports univariate time series data with a seasonal component. Fadhilah and Lawal Kane [7] conducted a study to model monthly rainfall time series in Kuantan and Malacca from January 1968 to December 2003 using ETS State Space and SARIMA models. They claimed that an accurate estimate of rainfall is needed for water resources applications, for instance in designing system and irrigation. By diagnostic checking, they found that both models to be adequate for forecasting because the residuals are uncorrelated, that is fits the series well.

Besides, Afrifa-Yamoah el al. [8] applied the SARIMA model to model and forecast of monthly rainfall in the Brong Ahafo (BA) Region of Ghana. They observed 420 data points which is from January 1975 to December 2009 data. The time series plot has a regular pattern of rise and fall, indicating a seasonality, therefore the seasonal differencing with period of 12 is needed. The best model for the data was identified, SARIMA $(0, 0, 0)(1, 1, 1)_{12}$ based on the featured portrayed by time series plot, ACF and PACF plots and also the minimum value of Akaike Information Criteria (AIC). They found that the forecasted values were very close to the actual points and it can be concluded that the model can forecast well.

Other than that, a study by Murthy et al. [9] has proposed a SARIMA process in modeling and forecasting rainfall patterns of South-West monsoon in North-East India. They used the seasonal rainfall data from 1951 and 2014. For parameter estimation, they used conditional least square estimation (CLS). The results show that SARIMA $(0, 1, 1)(1, 0, 1)_4$ is the best fitted model. However, the model does not model the extreme values well. Because of that, they suggested to consider X-12 ARIMA seasonal adjustment method for future study. Interestingly, the forecasting based on this model indicates that the rainfall patterns in India almost same in the next 3 years.

Furthermore, Fouli et al. [10] also used the Box-Jenkins approach to anticipate seasonal rainfall and runoff amounts in the Riyadh Region, Saudi Arabia. They focused on four gauges of Total Quarterly Rainfall (TQR) depth time series from 1964 to 2014. They identified three best-fit seasonal ARIMA (SARIMA) models for 10-year TQR forecasts from a collection of more than 20 models based on AIC and SBC criteria. It showed a good agreement in comparisons between the forecasted and the actual TQR. The most suitable model's accuracy assessment findings show that 58 percent of residuals are within 10 mm and 74 percent are within 20 mm for the most suitable model.

3 Materials and Methods

3.1 Data

The rainfall pattern of Peninsular Malaysia is influenced by two monsoon seasons namely Southwest monsoon which is from May to September and Northeast monsoon which is from November to March. The transition period between the two monsoons occurs on April and October which are called as inter-monsoon season. Two rainfall stations which are Sungai Petai, Kelantan station (S01) and Pam Kubang Haji, Perak (S02), located in the East and West coast region of Peninsular Malaysia respectively are chosen as the research area based on the completeness of the data and the length of records. The monthly rainfall time series data is obtained from Department of Drainage and Irrigation (DID) are divided into two sets, training and test data. The time series data from January 2000 to December 2015 are used as training data for modeling while the rest of data are used as test data for forecasting. The data is analysed with R software.

3.2 Autoregressive Integrated Moving Average (ARIMA) model

Autoregressive Integrated Moving Average (ARIMA) model is one of the most popular forecasting methods for univariate time series data forecasting. The ARIMA model consists of the following components called the order of autoregressive model (p), differencing order (d) and the order of moving average model (q).

The general form of the model ARIMA (p, d, q) is given by

$$(1-\phi_1 B-\cdots-\phi_p B^p)(1-B)^d y_t=\delta+(1-\theta_1 B-\cdots-\theta_q B^q) e_t \tag{1}$$

Equation (1) can be written as

$$\phi_p(B)(1-B)^d y_t=\delta+\theta_q(B) e_t \tag{2}$$

3.3 Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

ARIMA model can be extended to handle the seasonal components of a data series. The seasonal ARIMA model, SARIMA $(p, d, q)(P, D, Q)_s$ can be defined as

$$\begin{aligned}(1-\phi_1 B-\phi_2 B^2-\cdots & \\ & -\phi_p B^p)(1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots \\ & -\Phi_P B^{Ps})(1-B^s)^D(1-B)^d x_t \\ & =\delta+(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q)(1-\Theta_1 B^s-\Theta_2 B^{2s} \\ & -\cdots-\Theta_Q B^{Qs}) \varepsilon_t\end{aligned} \tag{3}$$

Equation (3) can be written as

$$\phi_p(B) \Phi_P(B^s)(1-B)^d(1-B^s)^D y_t=\delta+\theta_q(B) \Theta_Q(B^s) \varepsilon_t \tag{4}$$

where

$\phi(B)$: Autoregressive component of order p, AR (p)
$\theta(B)$: Moving average component of order q, MA (q)
$\Phi_P(B^s)$: Seasonal autoregressive component of order P, SAR (P)
$\Theta_Q(B^s)$: Seasonal moving average component of order Q, SMA (Q)
$(1-B)^d$: Difference component of order d, I (d)
$(1-B^s)^D$: Seasonal difference component of order D, I (D)
S : seasonal period

3.4 Box-Jenkins Algorithm

To build the ARIMA and SARIMA models, the Box-Jenkins Algorithm are followed. There are four stages involved which are model identification, parameters estimation, diagnostic checking and forecasting.

4 Results and Discussion

4.1 ARIMA Model Development

In model identification, the first step is to check for the stationarity by plotting the time series. The time series plot for the both stations are presented in Figure 1. The series is non-stationary for S01 since there are no fluctuation around the constant mean and while the fluctuation around the constant mean for S02 is an indication of stationarity. Hence, the differencing is only needed for S01 and one order difference is enough for the monthly rainfall time series data to achieve stationarity as shown in Figure 2.

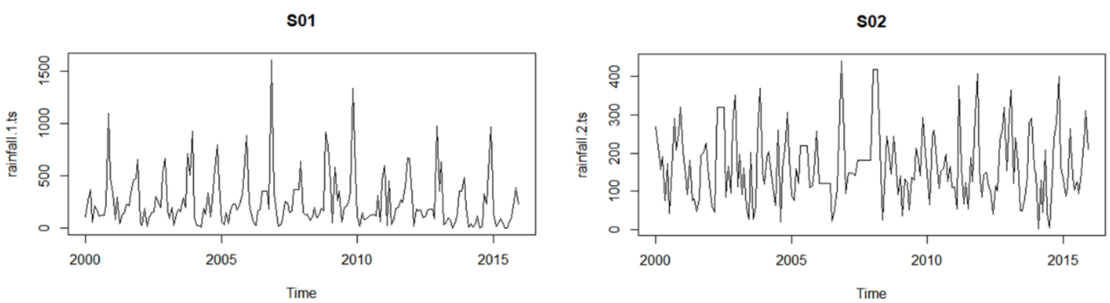

Figure 1 Plots of monthly rainfall time series

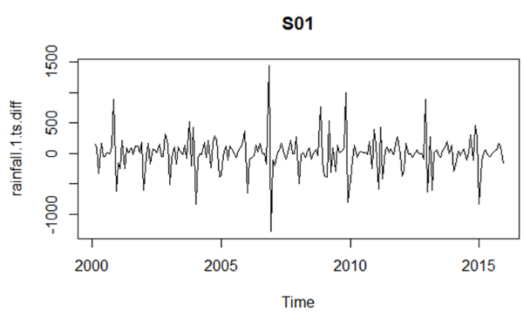

Figure 2 Plot of first differenced monthly rainfall time series

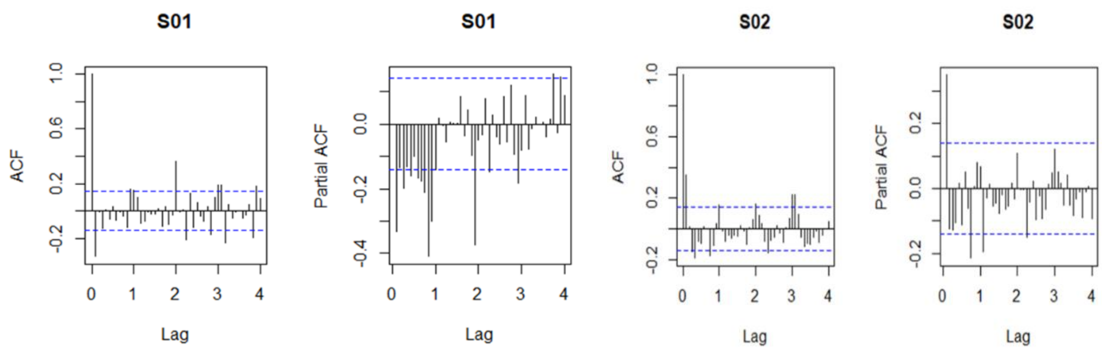

Figure 3 ACF and PACF plots of differenced monthly rainfall time series

Once the series is stationary, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs as presented in Figure 3 are examined to identified the tentative models. ARIMA (1, 1, 1) model is selected as the best model among five tentative models for S01 while ARMA (0, 1) model for S02 among five tentative models. The models are selected based on the minimum value of Akaike Information Criterion (AIC). The estimated parameters ARIMA (1, 1, 1) and ARMA (0, 1) are significant from Figure 4. In diagnostic checking, the residual plots and the distribution of the residuals for the both best ARIMA models as presented in Figure 5 does not show a white noise. Hence, it can be concluded that the model is a not an adequate model for forecasting time series. In the stage of forecasting, the actual and forecast values by ARIMA (1, 1, 1) for S01 and ARMA (0, 1) for S02 are presented in Figure 7. For the both stations, the forecast series not closely resembles the actual series, showing that the ARIMA models are a not good for forecasting.

```
Call:
arima(x = rainfall.1.ts, order = c(1, 1, 1))

Coefficients:
         ar1      ma1
      0.3172  -1.0000
s.e.  0.0689   0.0177

sigma^2 estimated as 57830:  log likelihood = -1320.5,  aic = 2647.01
```

```
Call:
arima(x = rainfall.2.ts, order = c(0, 0, 1))

Coefficients:
         ma1  intercept
      0.3603   167.9689
s.e.  0.0641     8.3679

sigma^2 estimated as 7285:  log likelihood = -1126.29,  aic = 2258.58
```

Figure 4 Parameter estimates for ARIMA (1, 1, 1) and ARMA (0, 1)

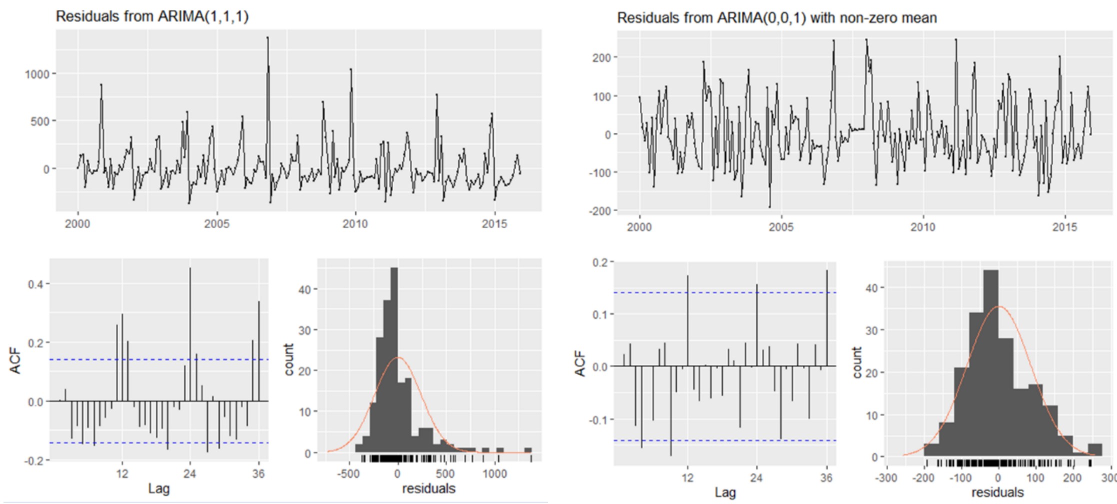

Figure 5 Residual plots and distribution of residuals

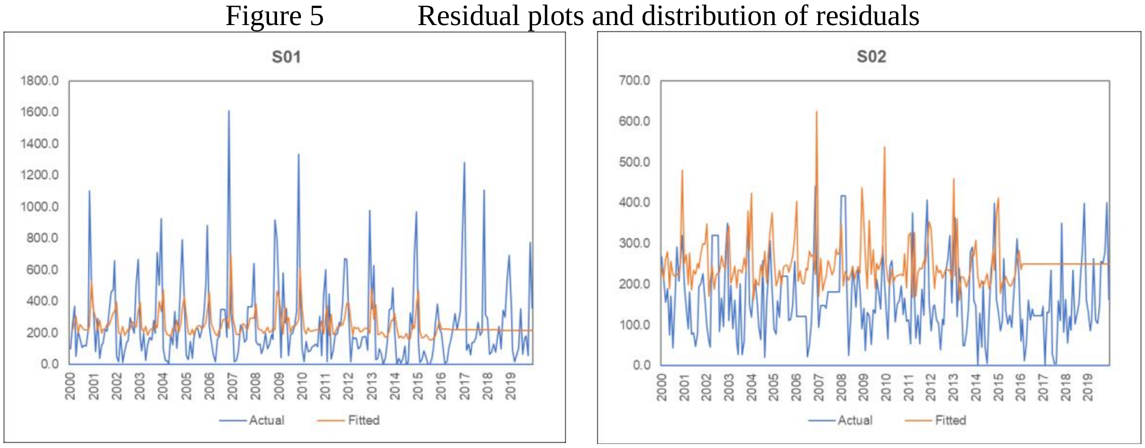

Figure 6 Plots of actual value and forecast value by ARIMA (1, 1, 1) and ARMA (0, 1)

4.2 SARIMA Model Development

The first step in developing SARIMA model is to transform a non-stationary monthly rainfall time series to stationary time series. Based on Figure 1, only the series for S01 is non-stationary, hence after the first seasonal differencing as shown in Figure 8, the series fluctuates around the zero mean and has become stationary. Based on ACF and PACF graphs in Figure 3 and Figure 7, several tentative SARIMA models are identified and SARIMA (0, 0, 1)(2, 1, 4)$_{12}$ and SARMA (1, 0)(2, 3)$_{12}$ are selected as the best fit model for S01 and S02 based on the minimum value of AIC.

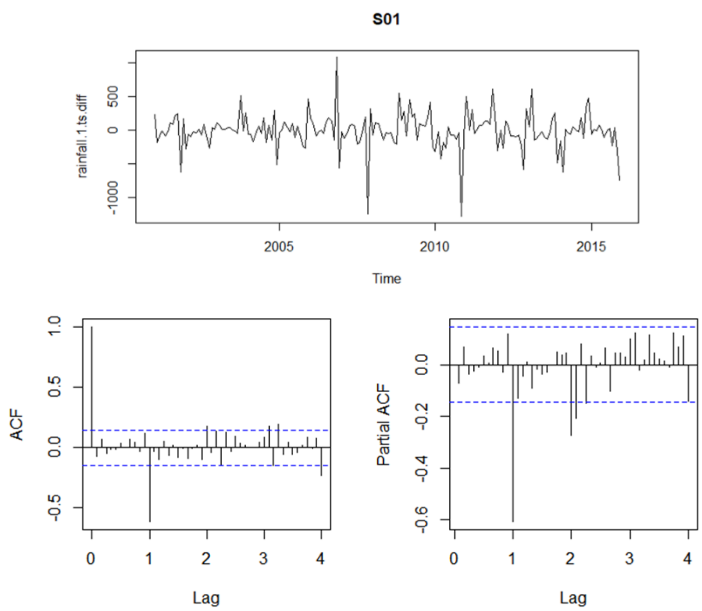

Figure 7 Time series, ACF and PACF plots for seasonally differenced monthly rainfall time series

The estimated parameters of SARIMA (0, 0, 1)(2, 1, 4)$_{12}$ and SARMA (1, 0)(2, 3)$_{12}$ are significant from Figure 8 and Figure 9. In diagnostic checking, the residual plots and the distribution of the residuals for the both best fit ARIMA model as presented in Figure 10 shows a white noise. Hence, it can be concluded that the model is a suitable model for forecasting time series. In the stage of forecasting, the actual and forecast values by SARIMA (0, 0, 1)(2, 1, 4)$_{12}$ and SARMA (1, 0)(2, 3)$_{12}$ are presented in Figure 11. The forecast series closely resembles the actual series, indicating that both stations, S01 and S02, have a good forecast.

```
Coefficients:
          ma1      sar1      sar2      sma1     sma2      sma3      sma4
      -0.0702   -0.7632   -0.6616   -0.2333   0.2750   -0.6498   -0.1381
s.e.   0.0723    0.2859    0.3096    0.2735   0.5988    0.4663    0.2234
```

Figure 8 Parameter estimates of SARIMA (0, 0, 1)(2, 1, 4)$_{12}$ for S01

```
Coefficients:
          ar1    sar1    sar2   sma1     sma2    sma3  intercept
       0.3119  0.0906  0.7956  0.104  -0.8229  0.0721   168.7421
s.e.   0.0714  0.1170  0.1353  0.282   0.2948  0.1200    16.5880
```

Figure 9 Parameter estimates of SARMA (1, 0)(2, 3)$_{12}$ for S02

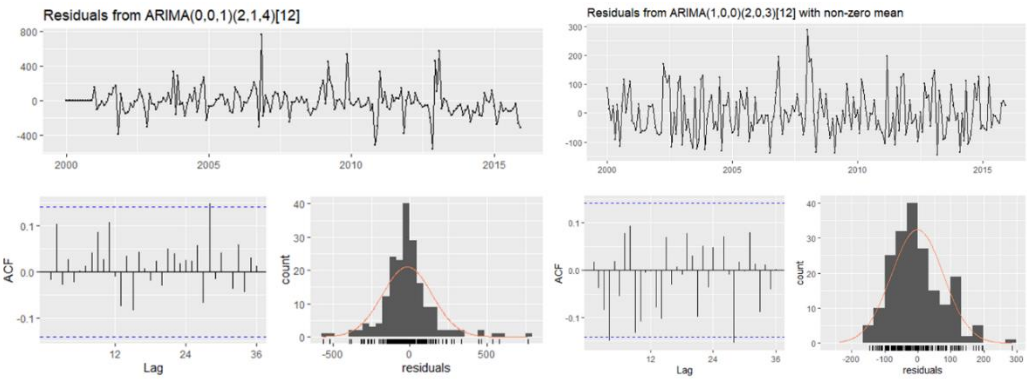

Figure 10 Residual plots and distribution of residuals

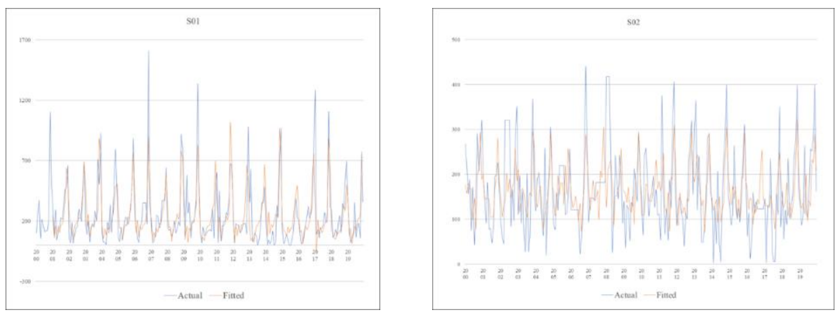

Figure 11 Plot of actual value and forecast value by SARIMA $(0, 0, 1)(2, 1, 4)_{12}$ and SARMA $(1, 0)(2, 3)_{12}$

4.3 Performance Comparison Between ARIMA and SARIMA Models

Forecast performance can be evaluated by measuring the Root Mean Square Error (RMSE) criterion. From the RMSE values presented in Table 1, the performance of SARIMA model is better than ARIMA since for S01 and S02, the RMSE of SARIMA models are 161.4366 and 78.21133 respectively which is smaller than the RMSE of ARIMA models. Hence, it can be concluded that SARIMA model is more precise compared to ARIMA model.

Table 1 Performance comparison based on RMSE value

	RMSE	
	S01	S02
Forecast by ARIMA	239.8522	85.35241
Forecast by SARIMA	161.4366	78.21133

5 Conclusion

In this study, two models of univariate time-series analysis, which is ARIMA model and SARIMA model were applied to model and forecast the monthly time series rainfall of two stations in Peninsular Malaysia which are station Sungai Petai in Kelantan, located at East coast region and station Pam Kubang Haji in Perak, located at West coast region. By applying Box-Jenkins methodology, we were found that SARIMA $(0, 0, 1)(2, 1, 4)_{12}$ are the best fit model at station Sungai Petai, Kelantan. Meanwhile, at station Pam Kubang Haji in Perak, SARIMA $(1, 0)(2, 3)_{12}$ are chosen as a best fit model to forecast monthly rainfall time series. The model fitted the time series data well except for some extreme values. In addition, the forecasts accuracy was measured used RMSE and it is proven that the best fit SARIMA model at the both stations are to be successful compared to ARIMA model due to the smaller error. Hence, it can be concluded that the SARIMA model is a good model for forecasting monthly rainfall time series.

6 References

[1] Wong, C., Venneker, R., Uhlenbrook, S., Jamil, A. and Zhou, Y. (2009). Variability of rainfall in Peninsular Malaysia. doi: 10.5194/hessd-6-5471-2009

[2] Jamaludin, S., Mohd Deni, S., Wan Zin, W. Z. and Jemain, A. A. (2010). Trends in Peninsular Malaysia rainfall data during the southwest monsoon and northeast monsoons seasons. *Sains Malaysiana*, 39 (4), 533-542.

[3] Chen, T., Tsay, J., Yen, M. and Matsumoto, J. (2013). The Winter Rainfall of Malaysia. *Journal of Climate*, 26(3), 936-958. doi: 10.1175/jcli-d-12-00174.1

[4] Katimon, A., Shahid, S. and Mohsenipour, M. (2017). Modeling water quality and hydrological variables using ARIMA. *Sustainable Water Resources Management*, 4(3), p. 101007.

[5] Al Balasmeh, O., Babbar, R., and Karmaker, T. (2019). Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan. *Arabian Journal of Geosciences*, 12(2). doi: 10.1007/s12517-018-4205-z

[6] Ray, S., Das, S., Mishra, P. and Al Khatib, A. (2021). Time series SARIMA modelling and forecasting of monthly rainfall and temperature in the South Asian countries. *Earth Systems and Environment*. doi: 10.1007/s41748-021-00205-w

[7] Yusof, F. and Lawal Kane, L. (2012). Modeling monthly rainfall time series using ETS State Space and SARIMA models. *International Journal of Current Research*, 4(), 155-200.

[8] Afrifa-Yamoah, Ebenezer & Saeed, Bashiru & Azumah, Karim. (2016). Sarima Modelling and Forecasting of Monthly Rainfall in the Brong Ahafo Region of Ghana. World Environment. 6. 1-9. 10.5923/j.env.20160601.01.

[9] Narasimha Murthy, K.V., Saravana, R. and Vijaya Kumar, K. (2018). Modeling and forecasting rainfall patterns of southwest monsoons in North–East India as a SARIMA process. *Meteorol Atmos Phys*, 130(), 99-106.

[10] Fouli, H., Fouli, R., Bashir, B., & Loni, O. (2017). Seasonal Forecasting of Rainfall and Runoff Volumes in Riyadh Region, KSA. *KSCE Journal of Civil Engineering*, 22(7), 2637-2647. doi: 10.1007/s12205-017-0726-3