Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia

Vol. 16, 2023, page 192 – 200

Forecasting Monthly Rainfall by Using Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average (Sarima) and Generalized Autoregressive Conditional Heteroskedasticity (Garch) Model

Low Ching Yi, Norazlina Ismail*
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: i-norazlina@utm.my

Abstract
In general, Subang experiences a tropical rainforest climate with high humidity and consistent rainfall throughout the year. On average, the annual rainfall in Subang ranges from 2000 to 2500 millimetres. However, there may be slight variations in rainfall intensity and patterns depending on the monsoon seasons and cause the phenomena of seasonality. As a results, forecasting rainfall is critical for water management and disaster preparedness by government. This study aims to model monthly rainfall in Subang using Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average (SARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. The models' performance is measured using Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The data of monthly rainfall in Subang from January 1980 to December 2009 were used as the training set to fit the forecasting model, while the 10 years data from 2009 used to access the forecasting accuracy of the models that have been implemented. The forecasting model with the lowest prediction error is chosen. This study shows that single exponential smoothing provides more accurate predictions than SARIMA and GARCH in forecasting monthly rainfall in Subang because the values of error measurement is the lowest.

Keywords: rainfall in Subang; exponential smoothing; Seasonal Autoregressive Integrated Moving Average; Generalized Autoregressive Conditional Heteroskedasticity

1. Introduction

The purpose of this study is forecast the monthly rainfall in Subang from January 2010 to December 2019. Rainfall in Subang Jaya is quite heavy throughout the year. The rainfall in Subang reaches a high peak during November which is the wettest month with 320mm of precipitation. On the other hand, June is the driest month with 126 mm of precipitation in Subang. Hence, the average amount of annual precipitation is 2630 mm in Subang.

Forecasting is a great tool for laying the groundwork for any action or policy before encountering any events. For instance, in the tropics, where there are only two seasons per year such as dry season and the rainy season, many nations, especially those that rely heavily on agricultural products, must forecast rainfall in order to determine when to plant their crops and how much they will yield (Kurniawan, 2020). Exponential smoothing can be expanded to support data with a systematic trend or seasonal component. Although the model explicitly utilises an exponentially decreasing weight for past observations, exponential smoothing forecasting approaches are similar in that a prediction is a weighted sum of past observations (Brownlee, 2018). Autoregressive Integrated Moving Average (ARIMA) is a forecast univariate time series data method which involves autoregressive and moving average elements. There is an extension of ARIMA that may support univariate time series with seasonality as known as SARIMA (Brownlee, 2018). Thus, the SARIMA model is popular, and many researchers use it to model rainfall. Time series data in rainfall often shows characteristics of instability. The Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized ARCH (GARCH) models, created by Engle (1982) and later expanded by Bollerslev, were two of the most well-known tools for capturing such changing variance (Ibrahim Lawal Kane, 2013).

The average amount of rainfall in Subang is about 7.2124 millimeters per month from January 1980 to December 2019. However, the amount of rainfall reached a higher peak during October to December. This phenomenon leads to frequent occurrence of floods within this period. Volatility also exists in the data. Hence, it is important for us to predict the monthly rainfall to prevent this issue in the future. The objectives of this study are to model rainfall using Exponential Smoothing, SARIMA and GARCH model, analyze the performance of Exponential Smoothing, SARIMA and GARCH model in rainfall and predict the accuracy of forecast monthly rainfall using Exponential Smoothing, SARIMA and GARCH model.

The scope of this research is to study the monthly rainfall in Subang. A time series data set regarding the topic is obtained. The modelling data is observed monthly from January 1980 until December 2019 to build a forecasting model. The methods that have been used in this study included the Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average (SARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. The software that will be implemented in this study is Microsoft Excel, Minitab and R.

2.1 Rainfall

Malaysia has a tropical climate. The nature of the climate in Malaysia is hot and humid throughout the year. There is no discernible difference between the seasons. Nonetheless, the two monsoon seasons, northeast and southwest, have an impact on rainfall characteristics. During these monsoon seasons, the nation experiences between 150 and 200 rainy days with an annual rainfall of 2000 to 4000 millimeters (mm). A mean annual total of 3001.9 mm from 1991 to 2020 indicates that rainfall was also plentiful throughout the year. The average monthly precipitation can also be said to be constant, ranging between approximately 200 mm during June and July and 350 mm in November and December (World Bank Climate Change Knowledge Portal, 2020). Selangor had the most yearly rainfall (3673.2mm), according to the Department of Statistics Malaysia. Labuan, W.P. Labuan, came in second (3433.6mm), while Bintulu, Sarawak, came in third (3316.6mm) (Department of Statistics Malaysia Official Portal, 2022).

2.2 Exponential Smoothing

A univariate time series forecasting method called exponential smoothing can be used to support data that includes a seasonal or systemic pattern. Due to its ease of use and forecast accuracy, this technique is often used in predictive analysis. (Maia & de Carvalho, 2011). Dhamodharavadhani and Rathipriya (2018) conduct a study about rainfall prediction by using MapReduce-based exponential smoothing techniques. In this paper, Simple Exponential Smoothing, Holt's Linear, and Holt-Winter's Exponential Smoothing methods are proposed with MapReduce computing model to predict region-wise rainfall with two different datasets. MSE accuracy measures are calculated to show accuracy for rainfall predictions. Therefore, Holt-Winter's smoothing method is the best model in this paper.

Dewi Darma Pertiwi (2021) applied exponential smoothing Holt-Winter method for rainfall forecast in Mataram City. Holt Winters' approach was utilized in this investigation because the data were time series data with a single variable. In this study, the amount of rainfall is exclusively examined using historical monthly rainfall data for Mataram. The forecasting of rainfall using the Holt-Winters Exponential Smoothing method for the next 12 periods or the following year in the city of Mataram has increased and decreased as a result.

2.3 SARIMA model

The three iterative processes of identification, estimation and diagnostic checking were introduced by Box and Jenkins which related to the SARIMA modelling procedure for seasonal patterns (Zhang et al., 2013). The next step is to ensure that the data are stationary by achieving performance with an appropriate seasonal difference in addition to the regular difference of the ARIMA model. Therefore, Augmented Dickey-Fuller (ADF) is the method to test stationarity. Once the transformed data is stationary, the autocorrelation function (ACF) and partial autocorrection functions (PACF) of the transformed data help determine seasonal and non-seasonal orders.

Fadhilah and Lawal Kane (2012) used Exponential Smoothing State Space model (ETS) and SARIMA model to conduct a study about modeling monthly rainfall in Malacca and Kuantan from January 1968 to December 2003. As a result, the models' suitability is put through diagnostic testing. Although this information was not suitable for predicting the precise monthly average rainfall, it may provide data that could aid in the establishment of strategies for effective planning of agriculture, drainage systems, and other applications of water resources in Malacca and Kuantan.

Jahan et al. (2018) conducted a study about modeling and forecasting monthly rainfall from January 1960 to December 2015 in Bangladesh by using SARIMA model. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test with a *p*-value of 0.1 and Augmented Dickey- Fuller (ADF) test with *p*-value of 0.01 proved the stationarity of Rainfall data after taking first difference original data was transformed to stationary. The Auto Regressive parameter was found to be statistically insignificant and SARIMA (0,0,0)×(2,1,1)_12 model that best fit and was used to forecast the data.

2.4 GARCH model

GARCH was introduced in 1986 by Dr. Tim Bollerslev. The GARCH algorithm is suitable for time series data where the error term's variance is serially autocorrelated after an autoregressive moving average process. Ibrahim Lawal Kane (2013) studied modeling the daily rainfall of Ipoh and Alor Setar by using hybrid ARIMA-GARCH modeling since Ipoh and Alor Setar are affected by nonlinear characteristics of the variance as volatility. Hybrid ARIMA-GARCH models were developed to take into account both the serial dependence and volatility in the daily rainfall series of Ipoh and Alor Setar.

Pantelidis (2014) studied forecasting volatility with a GARCH (1,1) model providing some new analytical formula. They examined the forecast performance of GARCH (1,1) model and showed the difference between the MSEs of the forecasts of the two models which are GARCH and homoscedastic models. Based on the comparison of the volatility forecasts of GARCH (1,1) to the volatility forecasts of a model that assumes homoscedasticity, the evaluation of the forecasting accuracy of GARCH (1,1) is processed. As a result, the accuracy volatility of the GARCH (1,1) model over the homoscedastic model.

3. Methodology

3.1 Data and Sources

This study employs forecasting models using the time series of the rainfall in Subang. The data for rainfall in Subang is obtained from January 1980 until December 2019 and the data is given by day. This study divides the data into two parts, firstly the data from January 1980 to December 2009 was analysed to find appropriate forecasting models. The second part uses the data from January 2010 to December 2019 to compare with the forecasted values derived from the employed forecasting models in the study. A total of 480 data will be used for the analysis in the next chapter using Exponential Smoothing, SARIMA and GARCH.

3.2 Exponential Smoothing

The Simple Exponential Smoothing method used for time series has no trend (stationary) and the mean (or level) of the time series y_t is slowly changing over time.

$$F_{t+1} = \alpha y_t + (1-\alpha)F_t \tag{1}$$

where
F_{t+1}= forecast for the next period
α = smoothing constant
y_t= observed value of series in period t
F_t= old forecast for period t

3.3 SARIMA model

The general form of multiplicative seasonal model $SARIMA\ (p,d,q) \times (P,D,Q)_s$ is given by

$$\Phi_P(B^s)\phi_P(B)\nabla^D{}_s\nabla^d x_t = \mu + \Theta_Q(B^s)\theta_q(B)a_t \tag{2}$$

where
$\Phi_P(B^s) = 1 - \Phi B^s - \Phi_2 B^{2s} - LL\Phi_P B^{Ps}$ is the seasonal autoregressive operator of order P
$\phi_P(B) = 1 - \phi_1 B - \phi_2 B^2 - LL{-}\phi_p B^p$ is the regular autoregressive operator of order p
$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - LL - \Theta_Q B^{Qs}$ is the seasonal moving average operator of order Q
$\theta_q(B) = 1 - \theta B - \theta_2 B^2 - LL{-}\theta_q B^q$ is the regular moving average operator of order q
μ is the intercept term or mean term $\nabla^d = (1-B)^d$; $\nabla^D{}_s = (1-B^s)^D$; $B^k x_t = x_{t-k}$
a_t is the non-stationary time series and it is the usual Gaussian white noise process
S is the period of the time series
B is the backshift operator

Augmented Dicker Fuller (ADF) test can be used to determine if the time series is stationary. The model identification is based on ACF and PACF graph.

Table 1: Properties of ACF and PACF for SARIMA

	$\boldsymbol{SAR(P)_S}$	$\boldsymbol{SMA(Q)_S}$	$\boldsymbol{SARMA(P,Q)_S}$
ACF	Tails off to 0 at lags k for $k = 1,2,\ldots$	Cuts off to 0 after lags Q_S	Tails off to 0 after lags Q_S
PACF	Cuts off to 0 after lags P_S	Tails off to 0 at lags k for $k = 1,2,\ldots$	Tails off to 0 after lags P_S

Besides that, the parameter estimation can be obtained through the estimation for constant and the coefficients of the parameter. The significance of parameters is tested using standard t-test. AIC and BIC criteria can be used to choose the best model among all possible models. The values of AIC and BIC should be as small as possible. In this diagnostic checking, the model must be checked for adequacy by using Ljung-Box Q statistics.

3.4 GARCH model

In the Generalized ARCH (GARCH) model, which Bollerslev proposed, conditional variance is also a linear function of its own lags, and its evolution is described by the following two equations:

$$y_t = \mu + \varepsilon_t \tag{3}$$

$$\varepsilon_t = \sigma_t \xi_t \tag{4}$$

$$\sigma^2{}_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}{}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}{}^2 \tag{5}$$

where the constraints parameters of above equation are:

$\omega > 0$, $\alpha_i \geq 0, i = 1,2,\ldots,q$ and $\beta_j \geq 0, j = 1,2,\ldots,p$, to ensure that the conditional variance ${\sigma^2}_t$ is nonnegative.

Checking for heteroscedasticity in a time series model is another way to test for ARCH effects or disorders. Check for ARCH disorders using the Lagrange Multiple (LM) test.

3.5 Forecasting Comparison

In the part of comparison, we use Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to compare the performance and accuracy of the model.

4. Results and discussion

4.1 Time Series Analysis

The monthly rainfall in Subang from January 1980 to December 2009 is visualized.

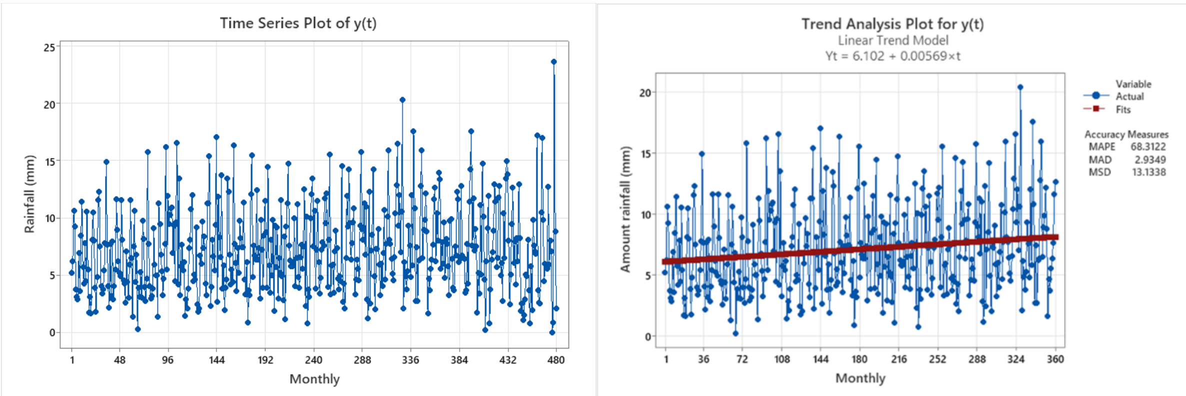

Figure 1 Time Series Plot and Trend Analysis for Monthly Rainfall in Subang

Trend is one of the essential factors in time series. The stationarity of the series represents the existence of a trend in the series. Based on figure, the graph shows a series of trends. There is a slightly increasing trend from January 1980 until December 2009.To clarify, the Augmented Dickey-Fuller (ADF) test was applied. The result shows a p-value of 0.001, which is smaller than the 0.05 significance level. Thus, H_0 is rejected at 5% significance level. Hence, we can conclude that there is sufficient evidence to show the data is stationary.

4.2 Single Exponential Smoothing

The monthly rainfall in Subang from January 1980 to December 2009 are used in training by using the single exponential smoothing. The different values of α are used by comparing their SSE and MSE. Hence, the best parameter is $\alpha = 0.9$.The equation of single exponential smoothing model can be expressed by using the general formula

$$F_{t+1} = 0.9y_t + (1-0.9)F_t \tag{6}$$

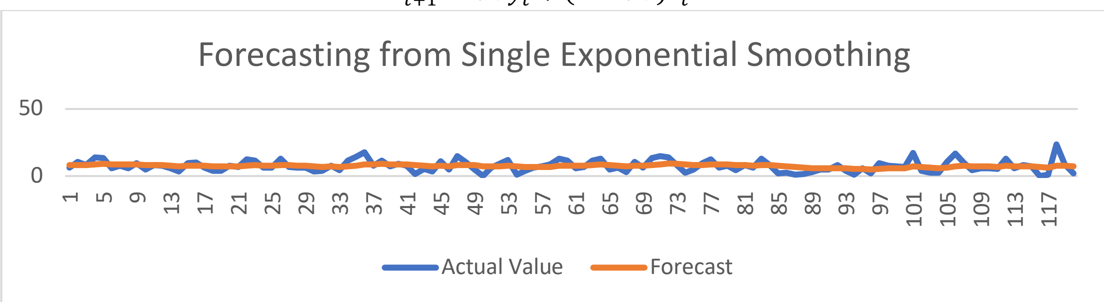

Figure 2 Forecasting from Single Exponential Smoothing

4.3 SARIMA model

The monthly In this section, the Seasonal Autoregressive Integrated Moving Average (SARIMA) is used to model the monthly rainfall in Subang from January 1980 to December 2009. ACF graph and PACF graph to determine the SARIMA model and parameters of the SARIMA model.

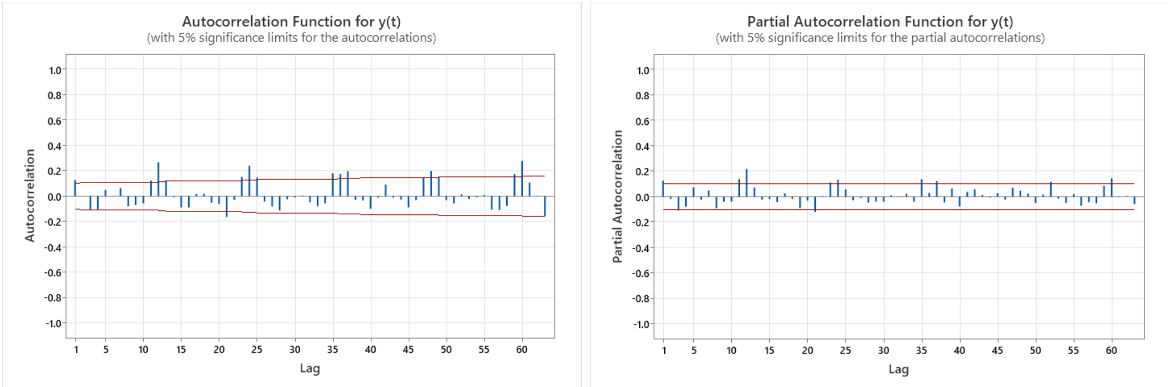

Figure 3 ACF and PACF of Monthly Rainfall

Based on t-test and their AIC and BIC values, the equation of best model which is $SARIMA(0,0,0)(2,0,0)_{12}$is

$$(1 - 0.2290B^{12} - 0.2183B^{24})y_t = 3.943 + e_t \tag{7}$$

After estimate the parameter of SARIMA model, we proceed to diagnostic checking by using Ljung-Box Q statistics and ACF and PACF of residuals plots.

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	6.74	21.26	31.75	49.71
DF	9	21	33	45
P-Value	0.664	0.443	0.529	0.291

Figure 4 Ljung-Box Q test

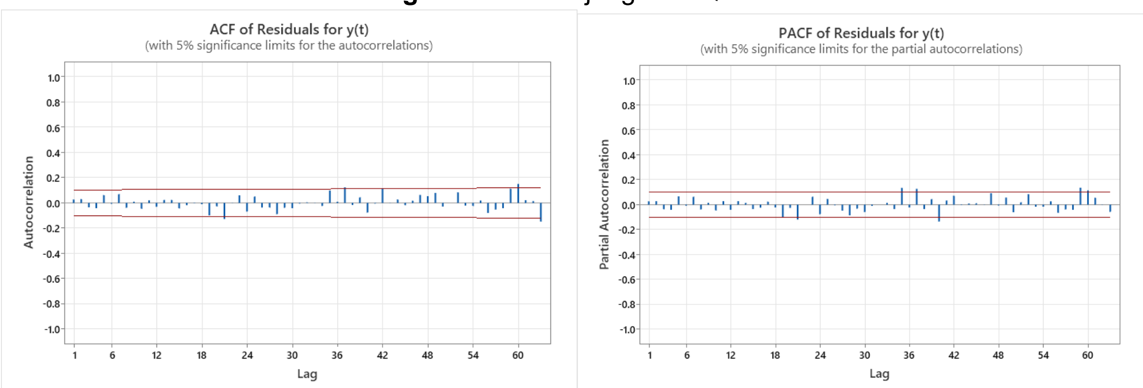

Figure 5 ACF and PACF of Residuals

Based on the figure 5, the residuals of $SARIMA(0,0,0)(2,0,0)_{12}$are not autocorrelated and the training data fit the model well. Based on the ACF and PACF plot of residual for the chosen model, there are few outliers of ACF and PACF located below the lower boundary of 95% confidence interval. However, most of the training sample autocorrelations lie within the 95% confidence interval. Besides that, the ACF and PACF of residuals for $SARIMA(0,0,0)(2,0,0)_{12}$is well within their two standard error limits indicating residuals are white noise.Therefore, it is an adequate model.

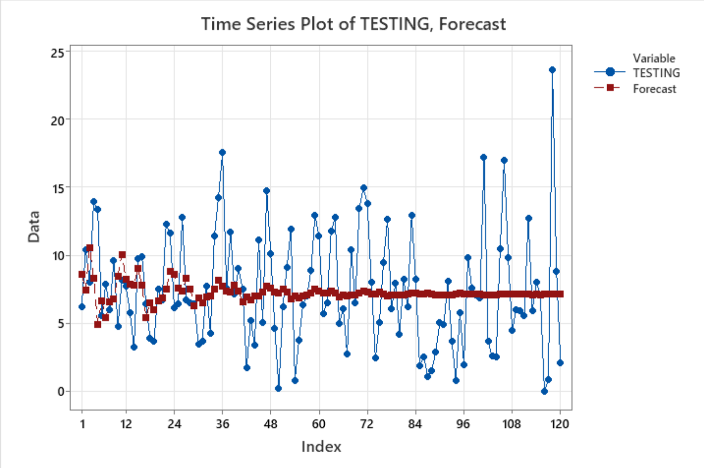

Figure 6 Forecasting from $SARIMA(0,0,0)(2,0,0)_{12}$ model

4.4 GARCH model

Lagrange multiplier (LM) test for autoregressive conditional heteroskedasticity (ARCH) of Engle in 1982 is frequently employed as a specification test. The result of ARCH-LM test for the data is shown in Figure below.

```
        ARCH LM-test; Null hypothesis: no ARCH effects

data:  timeseries
Chi-squared = 31.27, df = 12, p-value = 0.001792
```

Figure 7 LM test for ARCH effect

In this research, the GARCH model is used to model the data which is GARCH(1,1) because previous researchers have shown that the GARCH(1,1) model is the most suitable among all the GARCH model. Hence, GARCH(1,1) will be used to model the monthly rainfall data in Subang. GARCH(1,1) model can be written as

$$y_t = 7.116683 + \varepsilon_t \quad (8)$$

$$\sigma_t{}^2 = 0.0017 + 0.9990\sigma_{t-1}{}^2 \quad (9)$$

Diagnostic checking can be done by using weighted ARCH-LM test.

```
Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]     1.296 0.500 2.000  0.2550
ARCH Lag[5]     4.984 1.440 1.667  0.1036
ARCH Lag[7]     5.492 2.315 1.543  0.1793
```

Figure 8 Weighted ARCH LM test

Null hypothesis will be rejected if *p*-value <0.05. Based on the figure above, the *p*-value is more than 0.05, it means that null hypothesis is not rejected. Hence, there is sufficient evidence to conclude that there is no ARCH effect that exists at 5% level of significance.

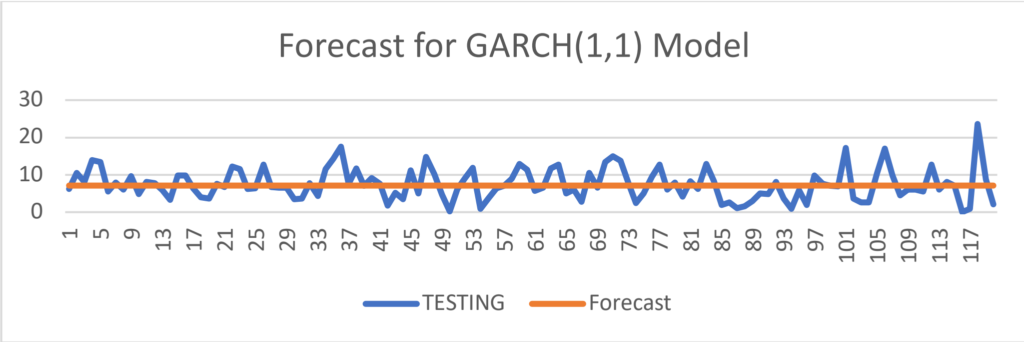

Figure 9 Forecasting from $GARCH(1,1)$ model

4.5 Forecasting Performance Evaluation

In this study, the method of single exponential smoothing, SARIMA and GARCH have been employed in modelling the monthly rainfall in Subang from year 1980 to 2009. The monthly forecasts from year 2010 until 2019 are completed by using these models. After forecasting the time series data with three different model, the values of error measurement are evaluated by using actual data and forecast data. Hence, the values of error measurement are summarized as table as below:

Table2: Error measurement of three models

	MAPE	MSE	RMSE
Single Exponential Smoothing	435.1904	14.8048	3.8477
$SARIMA(0,0,0)(2,0,0)_{12}$	464.6001	17.1053	4.1359
$GARCH(1,1)$	463.2794	17.4634	4.1789

Based on Table 2, single exponential smoothing is the best model among three models because the values of error measurement are the smallest such as MAPE is 435.1904, MSE is 14.8048 and RMSE is 3.8477. However, the values of MAPE of three model is large since it may be due the high variability in the data. If the data being forecasted has high variability or extreme values, it can lead to larger percentage errors. MAPE is calculated by taking the absolute percentage difference between the forecasted values and the actual values, so larger fluctuations in the data can result in larger errors.

Conclusion

In summary, forecasting monthly rainfall in Subang is important and it can influence water resource management, flood prediction and mitigation, climate monitoring and research and so on. The performances of the forecasting for these models were evaluated by using error measurements such as MAPE, MSE and RMSE. The model with forecasting accuracy with lowest values of error measurement can be conclude it is the best models among three types of models. In this study, single exponential smoothing is the best predictive of monthly rainfall in Subang. However, the data can be forecasted until 2019 only due to the COVID-19 outbreak from 2020.

Based on the analysis in this study, there are several forecasting approaches that can be considered for further research such as ARIMA, Artificial Neural Network (ANN) Machine Learning Models. Moreover, the application of hybrid forecasting, which is the combination of two methods could be a good practice in predictive analysis. The hybrid methodology can be used to overcome the drawbacks of the individual methods.

Acknowledgement
We would like to thank Mohd Eizam bin Yusof from the Department of Hydrology Management for giving access to SPRHiN system to apply the data for analysis in this project.

References

[1] Brownlee, J. (2019, April 24). A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python. Machine Learning Mastery. https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/

[2] DOSM. (2023). Department of Statistics Malaysia. Dosm.gov.my. https://www.dosm.gov.my/portal-main/landingv2

[3] Deluar, M., Moloy, J., & Kumar Mondal, S. (n.d.). Using SARIMA Approach to Modeling and Forecasting Monthly Rainfall in Bangladesh.

[4] Dewi Darma Pertiwi. (2021). Applied Exponential Smoothing Holt-Winter Method for Predict Rainfall in Mataram City. Journal of Intelligent Computing and Health Informatics (JICHI), 1(2), 46–49.

[5] Dhamodharavadhani, S., & Rathipriya, R. (2018). Region-Wise Rainfall Prediction Using MapReduce-Based Exponential Smoothing Techniques. Advances in Intelligent Systems and Computing, 229–239. https://doi.org/10.1007/978-981-13-1882-5_21

[6] Kurniawan, D. (2020, May 17). Rainfall Time Series Analysis and Forecasting - Towards Data Science. Medium; Towards Data Science. https://towardsdatascience.com/rainfall-time-series-analysis-and-forecasting-87a29316494e

[7] Lawal, I., Umaru, K., & Yar', M. (2014). Hybrid of ARIMA-GARCH Modeling in Rainfall Time Series.

[8] Maia, A. L. S., & de Carvalho, F. de A. T. (2011). Holt's exponential smoothing and neural network models for forecasting interval-valued time series. International Journal of Forecasting, 27(3), 740–759.

[9] Pantelidis, T. (2014). Forecasting Volatility with a GARCH (1, 1) Model: Some New Analytical and Monte Carlo Results.

[10] World Bank Climate Change Knowledge Portal. (2020). Worldbank.org. https://climateknowledgeportal.worldbank.org/country/malaysia/climate-data-historical

[11] Yusof, F. (n.d.). MODELING MONTHLY RAINFALL TIME SERIES USING ETS STATE SPACE AND SARIMA MODELS.

[12] Zhang, X., Liu, Y., Yang, M., Zhang, T., Young, A. A., & Li, X. (2013). Comparative Study of Four Time Series Methods in Forecasting Typhoid Fever Incidence in China. 8(5), e63116–e63116. https://doi.org/10.1371/journal.pone.0063116