ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 6 No. 1 (2026) 467-477
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Forecasting of Rainfall in Kota Bharu Station Using Holt-Winters, SARIMA and Time Series Regression Model

Nur Ayu Nabila Rahim[1], Maria Elena Nor[1]*

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: maria@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.053

Article Info

Received: 4 January 2026
Accepted: 6 June 2026
Available online: 9 July 2026

Keywords

Rainfall Forecasting, Rainfall Amount, Holt Winters, SARIMA, Time Series Regression, Forecast Accuracy

Abstract

Rain plays a vital role as a component of the hydrologic cycle in various aspects of life. However, continuous heavy rainfall can exacerbate other natural disaster problems, and it is essential to accurately forecast rainfall amounts, especially in areas that consistently experience heavy rain. This study focuses on comparing the forecast accuracy of three methods such as Holt Winters Exponential Smoothing (HWES), Seasonal Autoregressive Integrated Moving Average (SARIMA) and Time Series Regression (TSR) models, for monthly rainfall amount data at Kota Bharu Station. Based on the results, the SARIMA model was chosen as the best model and the most suitable method for forecasting rainfall amount compared to the HW and TSR models, as it presents lower values of Mean Absolute Deviation (MAD), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). The inconsistent trend of the data made it difficult for HW to adjust the level, trend and seasonal components effectively. Even though the TSR model was able to capture the seasonality patterns by using dummy variables, the linearity of the model and the inconsistent fluctuations of the data make the model less flexible in capturing sudden changes or extreme rainfall values. By incorporating seasonal differencing, seasonal Autoregressive (AR) and seasonal Moving Average (MA) components, the SARIMA model can capture both the seasonal patterns and fluctuations in rainfall amount over time. With a MAPE value of 40.5, the SARIMA model also exhibits reasonable forecasting performance in terms of MAPE interpretation and outperforms other methods. Given the limitations of this study, more robust error metrics, such as Mean Arctangent Absolute Percentage Error (MAAPE) or Mean Absolute Scaled Error (MASE) are recommended for future research especially when rainfall data contain minimal values near 0. Additionally, including other climate-related factors such as temperature, humidity, and wind speed, can further improve forecast accuracy.

1. Introduction

Rain is a water droplet that has been produced from a continuous evaporation and condensation process of the clouds to form a raindrop that will fall to the surface of the earth. Rain has many uses that benefit our lives, such as providing a source of water, promoting agricultural crop growth, facilitating transportation, mitigating disaster and supporting the economy, among others. However, continuous heavy rainfall can also have a significant impact on other natural disaster problems such as the risk of flooding and landslides. To mitigate the risks associated with continuous heavy rainfall, one effective way to assess the issues is through weather forecasting, especially in specific locations that have been affected. In Malaysia, flooding has become the most damaging natural calamity the country has ever experienced, and one cause of the flooding is continuous heavy rainfall that occurs in specific areas. This will create a challenging situation for the community due to the natural disasters resulting from constant heavy rain and flooding in their area [1].

Malaysia's position which is exposed to the northeast and southwest monsoons can also cause some areas in the country to experience a continuous pattern of heavy rainfall. According to [2], the northeast monsoon is caused by winds blowing from the northeast across the South China Sea, bringing heavy rain to nearby states and the east coast of Peninsular Malaysia, such as Kelantan, Terengganu and Pahang. The worst flood occurred in December 2014 in Kelantan, with flood levels reaching up to 5 to 10 meters [3]. In Kota Bharu, it consists of several stations and records varying amounts of rainfall depending on their location and geography. This indicates that the flood in Kota Bharu is caused by the combined contribution of high rainfall amounts from multiple stations within the district and possibly by the overflow of nearby rivers. To mitigate these problems, accurate rainfall forecasting is essential as it enables us to plan our daily activities and prepare for significant weather events such as storms, droughts, and floods [4].

In agricultural planning, water supply management, and flood prevention, rainfall forecasting is crucial in addressing these issues. Time series data modelling and forecasting are essential to many real-world applications [5]. This study explores and compares traditional techniques including HWES, SARIMA and TSR to forecast monthly rainfall amount data at Kota Bharu Station, Kelantan. However, the randomness of rainfall data influenced by other meteorological factors is one of the reasons why it is difficult to predict. Thus, some researchers have found that Machine Learning (ML) is an effective way to predict future rainfall due to the non-linear pattern of rainfall. [6] proposed some of the ML methods like random forest, logistic regression models, decision tree and neural network to predict the rainfall data from a variety of meteorological department in Iraq and Cyprus. A comparative study is being conducted by [7], [8], [9], and [10], which focuses on comparing the performance of ML and traditional methods in forecasting to identify the processes that yield better results in weather forecasting. The results of these studies indicate that ML outperforms traditional methods in forecasting rainfall data.

However, other studies still employ traditional forecasting techniques as their primary method of research. Although it may seem simple, it can provide intense competition to ML and produce better results without requiring support from advanced technology. This statement can be supported by the study of [11], which compares and discusses the reviews of other studies that have utilised ML and traditional techniques in forecasting. The results obtained from [11] show that in real-world scenarios, the traditional methods frequently yield better performance compared to ML in terms of forecast accuracy. The findings of [12], [13] and [14] indicate that the traditional technique is intensely competitive with ML. [13] found that the Autoregressive Integrated Moving Average (ARIMA) model shows a competitive performance by producing a low MSE, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values, underscoring its dependability in predicting rainfall. The findings of this study will help identify the most effective forecasting method for the data enabling farmers, local government organisations, or disaster response teams to make timely and informed decisions. In addition to this, by analysing the effectiveness of time series analysis using actual meteorological data, this study supports national initiatives to enhance early warning systems and climate resilience plans.

The continuous heavy rain accompanied by large amounts of rainfall not only leads to other natural disasters but also has a significant impact on the surrounding nature, the country's economy and the welfare of the local population. Therefore, accurate forecasts are crucial for enhancing people's welfare and enabling the timely provision of initial supplies to address natural disaster phenomena [15]. With unexpected rainfall patterns resulting from climate change, accurate rainfall forecasts are crucial for informed future planning. Inaccurate rain forecasts can lead to a failure in disaster management to deal with unpredictable situations such as continuous heavy rainfall which can bring about additional major problems [16]. This is why many forecasting techniques are used to estimate and predict future rainfall amounts. This study aims to investigate the suitability of various forecasting techniques including HWES, SARIMA and TSR models to enhance the precision of rainfall forecasts, particularly in areas highly susceptible to climate fluctuations such as Kelantan.

2. Methodology

This section discusses the methodology used in this study such as data description, data adjustment, HWES method, Box-Jenkins (BJ) method, TSR model and the measure of accuracy.

2.1 Data Description

The 132 observations range from January 2014 until December 2024 of the monthly rainfall amount in Kota Bharu Station, Kelantan are used in this study. The data used were specially request from the Malaysian Meteorological Department website accessible at https://mymetdata.met.gov.my/ and contain one variable which is the monthly rainfall amount in the mm. The data are separated into the training and testing sets. There are 120 records in the training set spanning from January 2014 to December 2023 and 12 records in the testing set covering the period from January 2024 to December 2024. The training set will be used to capture the features of the data while estimating the parameters and minimising the error. The testing set will be used to evaluate the model's performance. At the end of the analysis, the best model will be chosen based on its low MAD, MAPE, and MSE which indicate the model's performance.

2.2 Data Adjustment

However, the monthly rainfall data have several small observations that were close to 0 with minimum value of 0.2. With the minimum value appearing, it can affect the stability of percentage-based forecast accuracy measures such as MAPE. According to [17] and [18], the challenges of measures based on percentage error consist of having an extreme value if any observation is near to 0 and being infinite or undefined if any observation is equal to 0. This shows that even though the absolute forecast error is small, the percentage error will be overly large and misleading when the actual values are small because the denominator in the MAPE formula becomes small. To reduce the misleading caused by nearly 0 values in percentage-based error measures, the adjustment was applied by adding a constant value of 100 to the original monthly rainfall amount data and a similar approach also has been demonstrated by [19]. This adjustment was only made to stabilize the denominator, and the relative comparison of forecasting models is unaffected by this change. It can be considered as one of the applicable data adjustments or preprocessing methods like the adjustment applied in [20] and [21] which to improve the forecast accuracy. The forecast values will be generated using the adjusted data and it will be transformed back to the original scale solely for model comparison with the actual observations.

2.3 Holt-Winters Exponential Smoothing Method

According to [22], random fluctuations can be smoothed using an exponentially weighted moving average which has advantageous characteristics including that older data is given a decreasing weight, is simple to calculate, and requires the least amount of data. To obtain a new average value, a weighted average of two variables need to be computed which involves the current value of the variable and the average from the previous period. This theory serves as the foundation for the HW model, which builds upon it by incorporating trend and seasonal components. Both the additive and multiplicative HW methods were considered. Equation (1) showed the equation for the additive HW method, while Equation (2) showed the equation for the multiplicative HW method:

Additive Holt-Winters: $$y_t = (\beta_0 + \beta_1 t) + SN_t + IR_t \tag{1}$$
Multiplicative Holt-Winters: $$y_t = (\beta_0 + \beta_1 t) \times SN_t \times IR_t \tag{2}$$

where, $(\beta_0 + \beta_1 t)$ is the level, SN_t is seasonal patterns and IR_t is the irregular component.

The equations for each smoothing constant in additive HW are given in Equation (3) through Equation (6) while for multiplicative HW, the equation for each smoothing were given in Equation (7) until Equation (10).

Level Smoothing: $$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{3}$$
Trend Smoothing: $$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \tag{4}$$
Seasonal Smoothing: $$s_t = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{t-m} \tag{5}$$
Forecasting: $$\hat{y}_{t+h}(t) = \ell_t + hb_t + s_{t+h-m} \tag{6}$$

Level Smoothing: $$\ell_t = \alpha\left(\frac{y_t}{s_{t-m}}\right) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{7}$$
Trend Smoothing: $$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \tag{8}$$
Seasonal Smoothing: $$s_t = \gamma^*\left(\frac{y_t}{\ell_t}\right) + (1-\gamma^*)s_{t-m} \tag{9}$$
Forecasting: $$\hat{y}_{t+h}(t) = (\ell_t + hb_t)s_{t+h-m} \tag{10}$$

where h = 1, 2, …., n and m is the frequency of seasonality, ℓ_t is the level, b_t is the trend and s_t is the seasonal component at time t with the corresponding smoothing parameters α for the level, β^* for the trend and $\gamma^*$for seasonal factor.

2.4 Box-Jenkins Method

Using a methodology developed by George Box and Gwilym Jenkins, the ARIMA model of BJ is one kind of model that does take autocorrelation into consideration [23]. The ARIMA model is a time series modelling technique that combines an autoregressive AR(p), differencing, d and moving average MA(q). A crucial requirement for time series forecasting is that the series remains stationary after the differencing term eliminates seasonality or trends.

The current value dependence on lagged values is considered by the AR(p) term whereas a value dependence on lagged error terms is considered by the MA(q) term. Overall, the ARIMA (p, d, q) model provides a general framework for evaluating the linear relationships in time series data.

Equation (11) is the expanded form for the ARIMA model:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_Q(B^S)\alpha_t \quad (11)$$

where p denotes as the order of autoregressive (AR), d denotes as the differencing order, q denotes as the order of moving average (MA), P denotes as the seasonality order of AR, D denotes as the seasonality differencing order, Q denotes as the seasonality order of MA, B denotes as the backshift operator, y_t denotes as the observation for the time t and a_t denotes as the distribution of time t of non-seasonal operator

2.4.1 Model Selection

Model will be evaluated through a diagnostic checking which are including parameter significance, Ljung-Box test and the Residual Sum of Squares (RSS). In addition to diagnostic checking, the best model will be selected based on lowest values in certain criteria such as Akaike's Information Criterion (AIC), Corrected AIC (AICc) and Bayesian Information Criteria (BIC). Equation (12) represents the equation for AIC, Equation (13) for AICc and Equation (14) for BIC:

$$AIC = -2log(L) + 2(p+q+k+1) \quad (12)$$

$$AICc = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2} \quad (13)$$

$$BIC = AIC + [log(T) - 2](p+q+k+1) \quad (14)$$

where L is the likelihood of the data, $k = 1$if $c \neq 0$ and $k = 0$if $c = 0$

2.5 Time Series Regression Model

Regression analysis examines the relationship between one or more independent variables and dependent variables [17]. Linear Regression (LR) is a type of regression model used to study the linear relationship between one or more independent variables and a dependent variable. However, this method can be used to forecast the dependent variable and the TSR model is referred to when time related components such as trends or seasonal effects are combined in model [24]. LR can be expanded by adding dummy variables to reflect the seasonal impacts [25]. In regression, all the independent variables are treated as numerical. Dummy variables are artificial variables designed to represent an attribute with two or more different categories [26]. The only two possible values for the dummy variable are 1 and 0. Dummy variables enables the model to identify recurring trends within time frames such as months or quarters. Equation (15) shows the TSR model that included seasonal dummy variables and January was treated as the baseline:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 M_{Feb} + \beta_3 M_{Mar} + \cdots + \beta_{12} M_{Dec} + \epsilon_t \quad (15)$$

where Y_t is the dependent variable at time t, β_0 is the intercept, β_t is the coefficient of time t, M_{Feb} until M_{Dec} are dummy variable and ϵ_t is the error term.

2.6 Measure of Accuracy

Next, to evaluate the forecast performance of the applied methods, MAD, MAPE, and MSE need to be computed at the end of the analysis. To calculate this error, the actual and the predicted values need to be compared to obtain the forecast error. The best model will be chosen based on the lower error and is expected to have better performance in forecasting the following year's rainfall amount at Kota Bharu Station, Kelantan. Equation (16), (17) and (18) shows the formulas for the MAD, MAPE and MSE:

$$\text{MAD} = \frac{\sum_{t=1}^{n}|(y_t - \hat{y}_t)|}{n} \quad (16)$$

$$\text{MAPE} = \frac{\sum_{t=1}^{n}\left|\frac{(y_t - \hat{y}_t)}{y_t}\right|}{n}(100) \quad (17)$$

Enhanced Knowledge in Sciences and Technology Vol. 6 No. 1 (2026) 467-477

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2 \tag{18}$$

where y_t is the actual value, the $\hat{y}_t$ is the predicted or forecast value and n is the number of observations.

3. Results and Discussions

This section presents the results obtained by using HW, SARIMA, TSR and the model evaluation using accuracy metrics such as MAD, MAPE and MSE.

3.1 Time Series Plot

Fig. 1 shows the time series plot of monthly rainfall in Kota Bharu Station from January 2014 until December 2024. The plot showed that the data has a seasonal pattern by month with the rainfall amount presenting a high value in November and December. The data contains the minimum and maximum observation values of 0.2 mm and 1392.6 mm respectively. The data were separated into a training set and a testing set which the training set was included from January 2014 until December 2023 while the testing set was included from January 2024 until December 2024. Since the original data have some several values near 0, the adjustment was applied by adding 100 to the data. The new rescaling of the original data was referred to as the adjusted monthly rainfall amount data. Fig. 2 shows a time series plot of adjusted monthly rainfall amount at Kota Bharu Station from January 2014 until December 2024.

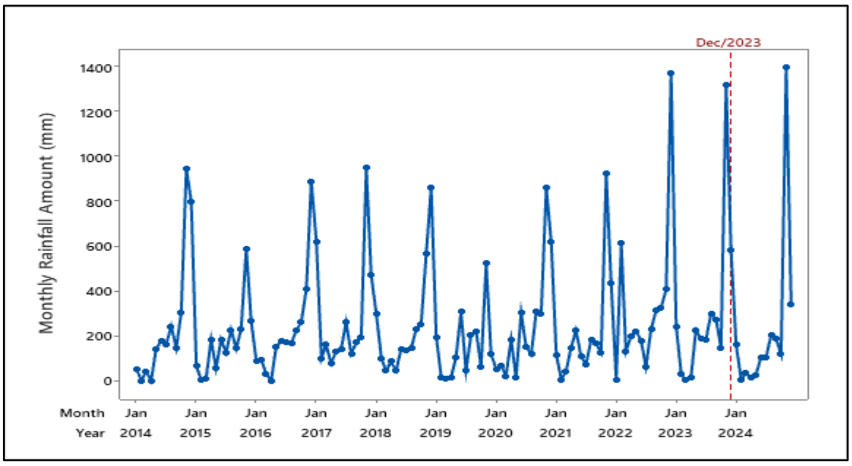

Fig. 1 Time series plot of Monthly Rainfall Amount (mm) in Kota Bharu Station from 2014 until 2024

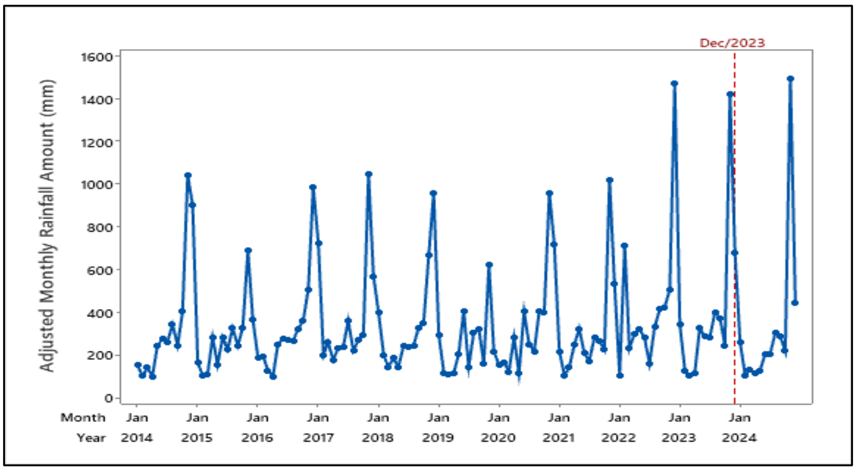

Fig. 2 Time series plot of Adjusted Monthly Rainfall Amount (mm) in Kota Bharu Station from 2014 until 2024

3.2 Holt-Winters Exponential Smoothing

Microsoft Excel was used to analyse the data, and the Solver was used to find the best combination of smoothing parameters which minimises the MAD. The combination of smoothing parameters was selected as the final parameters for the model and the value of each parameter fell between 0 and 1. For both additive and multiplicative HW, the smoothing parameters for level, trend and seasonality were computed as 0, 0 and 0 respectively. This indicates that the HW model does not give much weight to new data, especially when changing the level, trend and seasonal parts. The 0 value in the smoothing parameters indicates that the HW model considers the historical data stable and does not provide an estimated value based on new data. Thus, the rainfall data does not change significantly over time and does not exhibit a strong seasonal or trend pattern that aligns with the latest data in terms of minimising MAD.

Table 1 shows the forecast accuracy for both HW methods in the training and testing sets. Based on the results, the Multiplicative HW method shows lower values for MAD, MAPE and MSE in both sets compared to the Additive HW method. This suggests that, although the multiplicative model may not fit the new data as well as its historical data, it performs better forecasting than additive model. Thus, Multiplicative HW demonstrates better performance and was the chosen HW method for this study.

Table 1 The MAD, MAPE and MSE for both Holt-Winters Methods

Measurement	**Additive**		**Multiplicative**	
	Training	**Testing**	**Training**	**Testing**
MAD	101.1	184.2	99.9	167.9
MAPE	34.5	76.6	33.8	66.9
MSE	23866.0	53558.0	23589.6	46755.1

3.3 Box-Jenkins Method

In this study, Minitab software was used to analyse the BJ model. This method involves three important steps such as model identification, parameter estimation and diagnostic checking to get the best model. In the model identification step, the data must have a stable variance and be stationary before proceeding to the next step. The variance of the data can be checked using the Box-Cox Plot, where the given λ value indicates whether the data requires a transformation to stabilize the variance. The given rounded values determine what transformation needs to be applied. If the confidence interval for the optimal λ value includes 1, no transformation is necessary because the data has a stable variance and the original data are acceptable for the model estimation.

Fig. 3 shows the result of the adjusted data after the application of the reciprocal square root transformation. The transformation was applied because the lambda, λ value of the original data is -0.5. Based on Fig. 3, both the estimated value and the rounded value of the λ obtained for the data are 0.62 and 0.5 with a 95% confidence interval ranging from 0.08 to 1.21. The result shows that the confidence interval for the optimal λ is includes 1 indicating that the data already has a stable variance and the data is stationary.

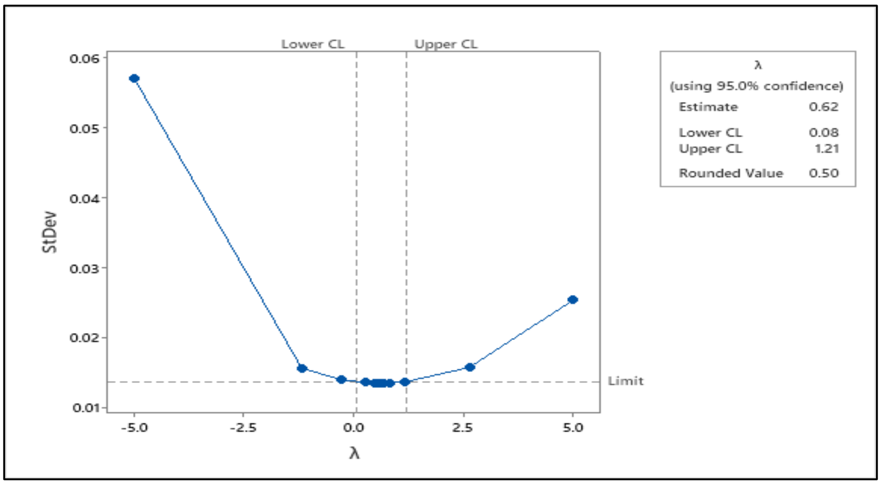

Fig. 3 Box-Cox Plot of Transformed Data

From the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots for the data after transformations are applied, a cut off is observed at lag 1 in both plots for the non-seasonal data while for the seasonal data, there is no significant lag shown in the PACF. However, there is a cut off at the seasonal lag in the ACF plot which is at lags 12 and 24. Regarding the seasonal lags, the data exhibits a strong seasonal pattern and applying seasonal differencing is necessary to stabilise the mean. Fig. 4 and Fig. 5 show an ACF and PACF plot for the Transformed Data after seasonal differencing has been applied. There is only one significant lag which is lag

12 for both the ACF and PACF. This result indicates that no considerable lag is observed in the ACF and PACF plots for the non-seasonal data after applying seasonal differencing. Hence, the data is satisfactory enough to select the best SARIMA model and estimate its parameters.

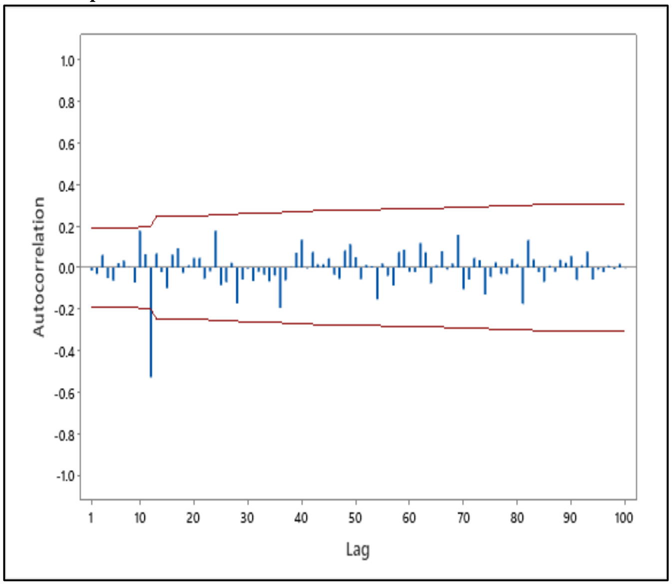

Fig. 4 The ACF of Seasonal Differencing Transformed Data

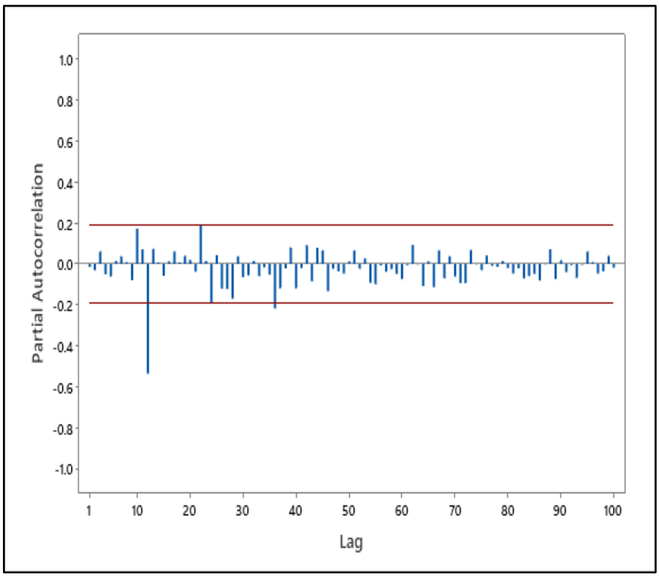

Fig. 5 The PACF of Seasonal Differencing Transformed Data

In parameter estimation, based on the ACF and PACF, there is no non-seasonal differencing applied, and no significant lag is observed in the ACF and PACF plot for the non-seasonal. However, for the seasonal component, a considerable lag was observed in both the ACF and PACF plots after applying seasonal differencing at lag 12. The parameter value was selected based on SARIMA (p,d,q) $(P, D, Q)_{12}$ model. Hence, the model chosen will be SARIMA (0,0,0) $(0,1,1)_{12}$ and SARIMA (0,0,0) $(1,1,0)_{12}$.

Table 2 presents the model parameters, Ljung-Box Chi-Square statistics and model criterion for two different SARIMA models. The p-value of the term in the model must be less than 0.05 to say that there is no association existing between the term and the response. Next, the value from the Ljung Box Chi-Square Statistics determines whether the model meets the assumption that the residuals are independent. The p-value needs to be greater than 0.05 to indicate that the residuals are independent. Since all the terms in these two models were less than 0.05 and the p-value of the Ljung Box was greater than 0.05, these two models still can be considered. However, based on the value of Residual Sum of Squares, SARIMA (0,0,0) $(0,1,1)_{12}$ has a lower Sum Square Error (SSE) and MSE value compared to SARIMA (0,0,0) $(1,1,0)_{12}$ model with the value of 0.0177 and 0.00017 respectively. The AIC, AICc and BIC for SARIMA (0,0,0) $(0,1,1)_{12}$ model were also lower than SARIMA (0,0,0) $(1,1,0)_{12}$ with the values of

-609.632, -609.401 and -601.586 respectively. Therefore, the better model for this data was SARIMA (0,0,0)(0,1,1)$_{12}$.

Table 2 ARIMA Model Parameters, Ljung Box Chi-Square Statistic and model criterion

Model Parameter	Parameter Significant (*p*-value)	Modified Box-Pierce (Ljung-Box) Chi-square statistic (*p*-value)	RSS		AIC	AICc	BIC
			Sum Square (SS)	Mean Square (MS)			
(0,0,0) (1,1,0)$_{12}$	0.000	0.231 0.323 0.062 0.124	0.0228	0.00022	-597.7	-597.4	-589.6
(0,0,0) (0,1,1)$_{12}$	0.000	0.284 0.791 0.274 0.540	0.0177	0.00017	-609.6	-609.4	-601.6

3.4 Time Series Regression Model

The TSR for the adjusted Monthly Rainfall Amount is given by the Equation (19).

$$Y_t = 237.2 + 0.646\,\text{t} - 71.7\,\text{February} - 126.0\,\text{March} - 85.0\,\text{April} - 45.0\,\text{May} + 13.3\,\text{June} - 40.8\,\text{July} + 15.3\,\text{August} + 40.3\,\text{eptember} + 39.6\,\text{October} + 567.7\,\text{November} + 459.5\,\text{December} \tag{19}$$

In the TSR method, January was selected as the baseline and the intercept value of 237.2 represented the average rainfall amount in January. The positive value of the trend coefficient showed an average upward trend of 0.646 mm per month. The negative coefficients for February, March, April, May and July indicate that the rainfall amounts in these months were lower than those in January while the positive coefficients for June, August, September, October, November, and December showed higher rainfall than January. For example, the rainfall amount in March was 126 mm lower than in January and the rainfall amount in September was 40.3 mm higher than in January. However, the values of coefficients that exceed 400 mm in November and December indicate a strong seasonal effect due to the monsoon season at the end of each year.

3.5 Comparing Forecast Performance

From Table 3, it can be seen that the forecast accuracy for three methods applied in the study which are HW, SARIMA and TSR. For the MAD, MAPE and MSE, the SARIMA model yields the lowest values compared to the other two methods with the values of 124.6, 40.5 and 40801.9 respectively. Additionally, MAPE values for HW and TSR exceed 50% which means that both methods show inaccurate forecasting based on the MAPE interpretation. With a MAPE value of 40.5, SARIMA exhibits reasonable forecasting performance in terms of MAPE interpretation and outperforms other methods. Therefore, based on the comparison of forecast accuracy measures, the SARIMA model was selected as the best forecasting method for predicting the adjusted monthly rainfall amounts for the year 2025. This selection was supported by the consistently lower MAD, MAPE, and MSE values produced by the SARIMA model, indicating better overall forecasting performance.

Table 3 The Forecast Accuracy for Holt-Winters, SARIMA and Time Series Regression Model

Measurement	Forecast Accuracy		
	Holt-Winters	SARIMA	Time Series Regression
MAD	167.9	124.6	159.8
MAPE	66.9	40.5	61.8
MSE	46755.1	40801.9	49238.8

Fig. 6 presents the visual representation of the actual values along with the forecast value for each method applied in the study to assess how well each technique captures the component of the data. From the plot, the SARIMA method forecasts closely on the original data even though the model might not adequately capture the relative fluctuations in some months.

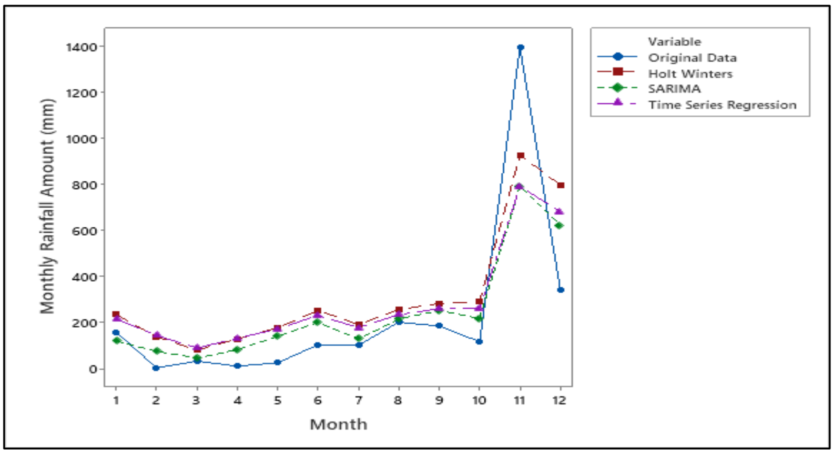

Fig. 6 The Original Data versus Forecast Values

3.6 Forecast

Based on the forecast performance of HW, SARIMA and TSR model, the SARIMA yields the lowest values of MAD, MAPE and MSE. Since the SARIMA (0,0,0) $(0,1,1)_{12}$ model yields the lowest value, it has been selected as the best method for forecasting the following year's monthly rainfall amount data at Kota Bharu Station. Table 4 presents the 1 year ahead which is year 2025 of monthly rainfall amount using SARIMA (0,0,0) $(0,1,1)_{12}$ while Fig. 7 shows the time series plot of monthly rainfall amounts (mm) with the forecast values. Based on the plot, it indicates that the forecast data in 2025 shows the repeated seasonal pattern which gives the consistent seasonal trend across time.

Table 4 1 Year Ahead Forecast of Monthly Rainfall Amount(mm) using SARIMA (0,0,0) $(0,1,1)_{12}$

Period	Month	Forecast
133	January	139.7
134	February	51.7
135	March	37.5
136	April	59.1
137	May	120.0
138	June	185.4
139	July	134.5
140	August	214.7
141	September	235.4
142	October	195.2
143	November	867.2
144	December	549.0

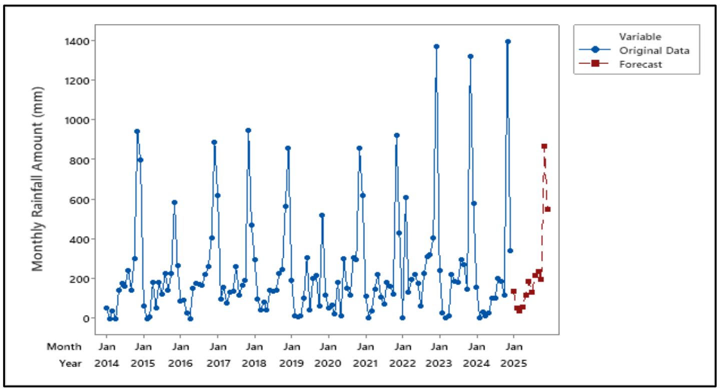

Fig. 7 Time series plot of Monthly Rainfall Amount (mm) with the Forecast Value of SARIMA (0,0,0) $(0,1,1)_{12}$

4. Conclusions and Recommendations

This study primarily focuses on the rainfall data from one station in Malaysia, specifically Kota Bharu Station, Kelantan and all the objectives of this study were successfully achieved. The analysis showed that the data have a strong seasonality by month and inconsistent trend over time. Based on the findings, it is evident that HW and TSR are less effective or less suitable for forecasting rainfall data compared to the SARIMA model. The inconsistent trend of the data made it difficult for HW to adjust the level, trend and seasonal components effectively. As a result, the HW model did not produce meaningful forecasts, and the SARIMA model was chosen instead for predicting future rainfall. The TSR model was able to capture the seasonality patterns by using dummy variables.

The high positive coefficients in November and December indicate that they have a strong seasonal effect due to the monsoon season at the end of each year. However, since the model was linear and the data exhibits inconsistent fluctuations, it was less flexible in capturing the sudden changes or extreme rainfall values. The SARIMA model produced lower values for MAD, MAPE and MSE with values of 124.6, 40.5 and 40801.9 respectively compared to other methods, HW and TSR. By incorporating seasonal differencing, seasonal AR and seasonal MA, the SARIMA model can capture both the seasonal patterns and fluctuations in rainfall amount over time. Additionally, with a MAPE value of 40.5, SARIMA demonstrates reasonable forecasting in terms of MAPE interpretation and outperforms other methods. Based on the comparison of forecast accuracy, the SARIMA model is selected as the best forecasting method for predicting rainfall amounts for the next year, 2025.

By comparing the multiple models and selecting the most reliable ones based on the error metrics, this study contributes to the field of rainfall forecasting in terms of addressing the rainfall forecasting challenges like handling the small values near 0 and evaluating the model performance. The challenges and findings provide guidance for selecting the most reliable method which can be used for forecasting the future rainfall and informing practical decision making for local authorities.

This study has several limitations where the rainfall data is inconsistent due to climate change. This becomes the main problem it contained an extremely small value that near to 0 especially in dry condition and large values during the heavy rainfall periods. Next, this study only relied on historical rainfall data of one station in Kota Bharu which has limits spatial coverage and may not represent rainfall variability across the region. This study also only focuses on rainfall forecasting and does not directly model floods limiting its direct application to flood prediction. So, it is recommended to use another alternative error metrics that are more robust for small values such as Mean Arctangent Absolute Percentage Error (MAAPE) or Mean Absolute Scaled Error (MASE) if the data have an observation value near to 0. Next, this study only relied on historical rainfall data, but rainfall mostly is influenced by another climate-related factors. For more accurate forecasts, the include of another factor like temperature, humidity and wind speed are recommended for the future research by increasing an overall accuracy.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology at Universiti Tun Hussein Onn Malaysia, for its support.

Conflict of Interest

The authors declare no conflict of interest in the paper’s publication

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Nur Ayu Nabila Rahim;* ***solve the governing equation:*** *Nur Ayu Nabila Rahim, Maria Elena Nor;* ***data collection:*** *Nur Ayu Nabila Rahim;* ***analysis and interpretation of results:*** *Nur Ayu Nabila Rahim, Maria Elena Nor;* ***draft manuscript preparation:*** *Nur Ayu Nabila Rahim, Maria Elena Nor. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Diya, S. G., BarzaniGasim, M., EkhwanToriman, M., & Abdullahi, M. G. (2012). Floods in Malaysia: Historical Reviews, Causes, Effects and Mitigations Approach. *International Journal of Interdiciplinary Research and Innovations*, *2*(4).

[2] Weather Phenomena. (n.d.). Malaysian Meteorological Department. Retrieved from https://www.met.gov.my/en/pendidikan/fenomena-cuaca/

[3] Alias, N. E., Mohamad, H., Chin, W. Y., & Yusop, Z. (2016). Rainfall analysis of the Kelantan big yellow flood 2014. *Jurnal Teknologi (Sciences & Engineering), 78(9-4).*

[4] Benziane, S. (2024). Survey: Rainfall prediction precipitation, review of statistical methods. *WSEAS Transactions on Systems, 23, 47-59*.

[5] Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*. https://doi.org/10.4-8550/arXiv.1302.6613
[6] Gupta, A., Mall, H. K., & Janarthanan, S. (2022). Rainfall Prediction Using Machine Learning. 2022 1st International Conference on Artificial Intelligence Trends and Pattern Recognition, ICAITPR 2022. https://doi.org/10.1109/ICAITPR-51569.2022.9844203
[7] Aini, N. N., Iriany, A., Nugroho, W. H., & Wibowo, F. L. (2022). Comparison of Adaptive Holt-Winters Exponential Smoothing and Recurrent Neural Network Model for Forecasting Rainfall in Malang City. ComTech: Computer, Mathematics and Engineering Applications, 13(2). https://doi.org/10.21512/comtech.v13i2.7570
[8] Kassem, Y., Gökçekuş, H., Çamur, H., & Esenel, E. (2021). Application of artificial neural network, multiple linear regression, and response surface regression models in the estimation of monthly rainfall in Northern Cyprus. Desalination and Water Treatment, 215. https://doi.org/10.5004/dwt.2021.26525
[9] Dani, A. T. R. (2025). Navigating Samarinda's climate: A comparative analysis of rainfall forecasting models. MethodsX, 14, 103080. https://doi.org/10.1016-/j.mex.2024.103080
[10] Abdullah, A. S., Ruchjana, B. N., Jaya, I. G. N. M., & Soemartini. (2021). Comparison of SARIMA and SVM model for rainfall forecasting in Bogor city, Indonesia. Journal of Physics: Conference Series, 1722(1). https://doi.org/10.1088/1742-6596/1722/1/012061
[11] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE, 13(3). https://doi.org/10.1371/journal.pone.0194889
[12] Ponnamperuma, N., & Rajapakse, L. (2021). Comparison of time series forecast models for rainfall and drought prediction. MERCon 2021 - 7th International Multidisciplinary Moratuwa Engineering Research Conference, Proceedings. https://doi.org/10.1109/MERCon52712.2021.9525690
[13] Mungale, N., & Shinde, J. (2024). Rainfall Forecasting: A Comparative Analysis of Deep Learning and Machine Learning Models with Application to Environmental Data. International Journal of Intelligent Systems and Applications in Engineering, 12(12s).
[14] Akbar, A. A., Darmawan, Y., Wibowo, A., & Rahmat, H. K. (2024). Accuracy Assessment of Monthly Rainfall Predictions using Seasonal ARIMA and Long Short-Term Memory (LSTM). In JCSE Journal of Computer Science an Engineering (Vol. 5, Issue 2). http://icsejournal.com/index.php/
[15] Ritchie, H. (2024, March 12). Weather Forecasts Have Become Much More Accurate; We Now Need to Make Them Available to Everyone. Our World Data. Retrieved from https://ourworldindata.org/weather-forecasts
[16] Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., Domeisen, D. I. V., Reserve, F., Koszalka, I., Kreibich, H., Pantillon, F., Parolai, S., Pinto, J. G., Hunger, H. J., Rivalta, E., Schroter, K., Strehlow, K., Weisse, R., & Wurpts, A. (2020). Impact Forecasting to Support Emergency Management of Natural Hazards. Reviews of Geophyscics, 58(4). https://doi.org/10.1029/2020-RG000704
[17] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice, 2nd edition. OTexts. Retrieved from https://otexts.com/fpp2/
[18] Kim, S., & Kim, H. (2016). A new matric of absolute percentage error for intermittent demand forecasts. International Journal of Forecasting 32, 669-679. http://dx.doi.org/10.1016/j.ijforecast.2015.12.003
[19] Minitab Blog Editor. (2016). Model Fit: Don't be Blinded by Numerical Fundamentalism. Minitab. Retrieved from https://blog.minitab.com/en/blog/statistics-and-quality-data-analysis/model-fit-dont-be-blinded-by-numerical-fundamentalism
[20] Imhoff, R., Verbrugge, M., & Goulart, H. M. D. (2025). Rainfall downscalling and adjustment methods. Deltares. Retrieved from https://www.deltares.nl/expertise/publicaties/rainfall-downscaling-and-adjustment-methods
[21] Nguyen, T. A., & Ha, C. K. (2025). Effect of Data Preprocessing on Short-Term Load Forecasting: A Convolutional Neural Network Approach. *Journal of Robotics and Control. 6(5).*
[22] Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting, 20(1), 5-10.
[23] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
[24] Wei, W. W. S. (2011). Time Series Regression. International Encyclopedia of Statistical Science. https://doi.org/10.1007/978-3-642-04898-2_596
[25] Usman, A. U., Hassan, S. A., & Tukur, K. (2015). Application of dummy variables in multiple regression analysis. International Journal of Recent Scientific Research, 7(11), 7440-2.
[26] Bobbitt, Z. (2021, May 31). How to Use Dummy Variables in Regression Analysis. Statutory. Retrieved from https://www.statology.org/dummy-variables-regression/