IRAQI Academic Scientific Journals

العراقية المجلات الأكاديمية العلمية

Al-Kut University College Journal

ISSN (E) : 2616 - 7808 II ISSN (P) : 2414 - 7419
www.kutcollegejournal.alkutcollege.edu.iq
k.u.c.j.sci@alkutcollege.edu.iq

Special Issue for the Researches of the 5th Int. Sci. Conf. for Creativity for 13-14 December 2023

Rainfall Forecasting with Time Series Model in Karbala, Iraq

Manar Naji Ghayyib [1] , Ali Amer Abdulrazzaq [2] , Farah Alaa Adnan [3]

Abstract

Natural calamities like floods and droughts are most frequently caused by waterlogging. The organizations that needs to make decisions in the case of a disaster or incident prevention might benefit from using weather forecasts. The weather prediction includes a forecast for rainfall. The prediction of rainfall has been approached in a number of ways, including physical, statistical, and hybrid approaches. The objective of this study is to predict the amount of rainfall using the best forecasting model, and then to evaluate that model's performance. Karbala was chosen as the research area and the monthly rainfall data were taken from the Iraqi Meteorological Department and covered the years from January 2012 to December 2023. It was discovered that Karbala's rainfall pattern had a trend and seasonality throughout the year. To predict the rainfall data in this study, the Holt-Winters, Box-Jenkins, and hybrid methods were suggested. Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) are performance measures that were used to assess the models' performance. It was discovered that the Holt-Winters model yields the most precise result and can be used to predict future rainfall.

Keywords: Holt-Winters, Box-Jenkins, Mean Absolute Percentage Error, Time Series Model.

Affiliation of Authors

[1] Baghdad University, Collage of Science, Department of Geology, Iraq, Baghdad, 10001

[2] Baghdad University, Collage of science, Iraq, Baghdad, 10001

[3] Baghdad university, Collage of Science, Department of Mathematics, Iraq, Baghdad, 10001

[1]manar.naji@sc.uobaghdad.edu.iq
[2]ali.amer@sc.uobaghdad.edu.iq
[3]farah.alaa@sc.uobaghdad.edu.iq

[1] **Corresponding Author**

Paper Info.
Published: June 2024

التنبؤ بكميات الامطار في محافظه كربلاء باستخدام السلاسل الزمنية -العراق

منار ناجي غايب [1] ، علي عامر عبد الرزاق [2] ، فرح علاء عدنان [3]

المستخلص

لكوارث الطبيعية مثل الفيضانات والجفاف هي في أغلب الأحيان ناجمة عن كميات الامطار الساقطة. قد تستفيد المنظمات التي تحتاج إلى اتخاذ قرارات في حالة وقوع كارثة أو منع وقوع حادث من استخدام التنبؤات الجوية، يتضمن التنبؤ بالطقس توقعات هطول الأمطار. تم التعامل مع التنبؤ بهطول الأمطار بعدة طرق، بما في ذلك الأساليب الفيزيائية والإحصائية والهجينة. الهدف من هذه الدراسة هو التنبؤ بكمية الأمطار باستخدام أفضل نموذج للتنبؤ ومن ثم تقييم أداء هذا النموذج. تم اختيار محافظة كربلاء كمنطقة للبحث وتم أخذ بيانات هطول الأمطار الشهرية من دائرة الأرصاد الجوية العراقية وللسنوات من كانون الثاني 2012 إلى كانون الاول 2023. وقد اكتشف أن نمط هطول الأمطار في كربلاء له اتجاه وموسمية على مدار العام. للتنبؤ ببيانات هطول الأمطار في هذه الدراسة، تم اقتراح طرق مختلفة منها هولت-وينترز، بوكس-جينكينز، والطرق الهجينة. الجذر المتوسط الخطأ المربع (RMSE)، والمتوسط النسبة المئوية للخطأ المطلق (MAPE)، والمتوسط الخطأ المطلق (MAE) هي مقاييس الأداء التي تم استخدامها لتقييم أداء النماذج. تم اكتشاف أن نموذج هولت-وينترز يعطي النتيجة الأكثر دقة ويمكن استخدامه للتنبؤ بكميات الأمطار في المستقبل.

الكلمات المفتاحية: هولت-وينترز، بوكس-جنكينز، متوسط النسبة المئوية للخطأ المطلق، نموذج السلسلة الزمنية

انتساب الباحثين

[1] جامعة بغداد، كلية العلوم، قسم علم الارض، العراق، بغداد، 10001

[2] جامعة بغداد، كلية العلوم، بغداد، العراق، 10001

[3] جامعة بغداد، كلية العلوم، قسم الرياضيات، بغداد، العراق، 10001

[1]manar.naji@sc.uobaghdad.edu.iq
[2]ali.amer@sc.uobaghdad.edu.iq
[3]farah.alaa@sc.uobaghdad.edu.iq

[1] **المؤلف المراسل**

معلومات البحث
تاريخ النشر: حزيران 2024

Introduction

Karbala is situated at longitude 44.01039 and latitude 32.60685 in the southwest of Baghdad, 105 kilometers away from it. 30 meters above sea level is where it is situated. Despite being at risk for both natural hazards like floods and forest fires as well as man-made disasters, Karbala has never experienced any significant natural disasters.

Karbala had witnessed a number of terrorist activities in the recent years, including burning of forests, rainfall forecasts are crucial for catchment management applications, especially flood warning systems in the aviation sector, severe rains can also lead to blurred vision, which affects the pilot's assessment of the runway's location and distance [1,2]. Rainfall accounted for 40% of all meteorological variables influencing flight safety, according to statistical study of all accidents involving all American airline firms from 1962 to 1984. Therefore, a trustworthy weather forecast is essential for both the wellbeing of the aviation industry's economy and the safety and security of pilots and passengers.

The process of forecasting involves looking back at past occurrences to make forecasts, making assumptions about the future, using historical data as indicators, and making predictions that can be relied on to come true in the future. The examination of sequences of data points gathered over a long period of time using a time series approach is a particular technique utilized in forecasting [3].

This study attempts to [4] use the most appropriate model to determine the pattern or trend of the amount of rainfall, [5] forecast the amount of rainfall using forecasting models such as the Holt-Winters model, SARIMA model, and a hybrid model and [6] assess the performance of forecasting models based on the performance pointers, which are Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Literature Review

Exponential smoothing procedure

One of the oldest and very popular forecasting techniques for time series data is exponential smoothing. In order to develop a prediction model for an inventory control system, Brown and Holt put out the exponential smoothing formulation in the late 1950s [7]. To put it simply, this technique estimates the future using the weighted averages from the past explanations [8].

Simple Exponential Smoothing (SES), Double Exponential Smoothing, and Triple Exponential Smoothing are the three primary forms of exponential smoothing algorithms [9]. When there is no trend or seasonality in a set of univariate data, simple exponential smoothing is utilized. Use the Double Exponential Smoothing approach, also known as Holt's approach], for observations that show patterns but no seasonal component [10,11]. For data containing trends and seasonality, the Triple Exponential Smoothing Method or the Holt-Winters Method is appropriate [9,12,13].

Box-Jenkins procedure

Time series modeling heavily relies on the Autoregressive Moving Average (ARMA) model. While they can model unknown processes with the fewest parameters when compared to autoregressive models (AR) and moving average models (MA), ARMA models offer the very effective linear models of fixed time series.

Forecast of rainfall is hard and challenging because of its nonlinear paradigm and wide changes in intensity [14]. Autoregressive Integrated Moving Average (ARIMA) modeling is one of the effective techniques [15]. The autoregressive moving average (ARMA) model, created by George Box and Gwilym Jenkins in the

1970s [16], is a generalization of the ARIMA model. Due to the fact that it encompasses a range of types, including season, absence of season, and stability The ARIMA model is promoted as a thorough statistical modeling approach for time series [17].

In a study by Momani [18], the ARIMA model was employed to forecast rainfall data, and it was asserted that the forecast's outcome was favorable. Ponnamperuma and Rajapakse [19] also discovered that ARIMA might be used to predict rainfall over the short term. Although both ARIMA and SARIMA approaches were recommended for use in forecasting future primary energy demand, Ediger and Akar [20] had come to the conclusion that ARIMA forecasts of Turkey's short-term primary energy demand are more accurate.

Hybrid method

A hybrid model combines traditional and/or contemporary models. To enhance time series forecasting, hybrid approaches that integrate the best features of statistics and machine learning have been proposed [21]. The two approaches that are most frequently employed in time series analysis are exponential smoothing and ARIMA models [22].

Grigonyt and Butkeviit [6] propose hybrid models to improve the accuracy of wind speed forecasts. In a case study, Kamisan et al. [23] created a hybrid model of ARIMA and SES, and the hybrid model keeps a favorable outcome. Sanusi and Safi investigated hybrid time series models [17] in order to predict COVID-19 worldwide. A hybrid design was created.

According to the aforementioned study, the hybrid model of ARIMA plus exponential smoothing outperforms the majority of time series models in a variety of industries. However, scholarly studies do not sufficiently support the hybrid model of SARIMA and Holt-Winters to predict rainfall; as a result, the hybrid model was explored further in this study [24].

Research Methodology

Hybrid approach

A development of Simple Exponential Smoothing is a model acknowledged as Double Exponential Smoothing, or Holt's approach. To detect trends, a second exponential smoothing model is added to the original one. The following equations were employed [25]:

Level: $k_t: \alpha y_t + (1-\alpha)(s_{t-1} + b_{t-1})$

Trend: $b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$

Forecast: $z_t = k_t + nb_t$

To find the initial values in Holt-Winters Additive method, the formula for the level s_t, the growth rate b_1 and seasonal factors $d_1, d_2, \ldots$ is as shown as follows:

$$s_1 = y_1 \quad (1)$$

$$b_1 = \frac{y_4 - y_1}{3} \quad (2)$$

Seasonal autoregressive integrated moving average (SARIMA)

Additional seasonal components are incorporated to create the seasonal ARIMA model. It is basically stated as SARIMA (p, d, q) (P, D, Q), where S is the seasonal interval [16]. The non-seasonal part of the model is called lowercase notation, and the seasonal part is called uppercase notation. Similar expressions are used in the seasonal ingredient of the model as in the non-seasonal ingredient, but there is a seasonal rearward shift. The following is the equation [26]:

$$\boldsymbol{\delta_h(P) \times \delta_h(P^S)(1-B)^d\left(1-B^S\right)^D Y_t = \delta_q(P)\vartheta_q(P^S)} \quad (3)$$

With

$$\boldsymbol{\delta_h(P) = 1 - \delta_1 B - \cdots - \delta_h B^h} \quad (4)$$

$$\boldsymbol{\vartheta_h(P) = 1 - \vartheta_h B^s - \cdots - \vartheta_h B^{hk}} \quad (5)$$

where y_t is the actual data; δ_h and ϑ_h represents AR component coefficient and MA component coefficient respectively; c is the constant value; μ is the mean value of the series and ε_t is the random error, which also known as white noise. B represents the non-seasonal backshift operators and d is the non-seasonal differencing order. For seasonal part, ϑ is the seasonal AR component coefficients while ϑ_h is the seasonal MA component coefficients. B- is the seasonal backshift operators and D representing the seasonal differencing order.

Test statistical

$$\boldsymbol{w^2 = -n - s} \quad (6)$$

$$w^{*2} = w^2\left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right) \quad (7)$$

Mean absolute error (MAE)

The discrepancy between the predicted and actual values is calculated in absolute terms as the absolute error. The following formula can be used to determine MAE [27, 28].

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - z_t| \quad (8)$$

Mean absolute percentage error (MAPE)

To compute this precision as a percentage, MAPE is computed as the average absolute percent error for every time minus real values split by actual values. The resulting is the formula [27, 9]:

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - z_t}{y_t}\right| \quad (9)$$

Root mean square error (RMSE)

The absolute fit of the model to the information, or how approaching the actual information points match the values predicted by the model, is symbolized by RMSE. It can be computed by the following formula [29]:

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n}\frac{(y_t - z_t)^2}{n}} \quad (10)$$

Result and Discussion

Time series of rainfall data in Karbala

The monthly rainfall statistics for Karbala were used from the Iraqi Meteorological for this study. The time frame for gathering this data was from January 2012 to December 2022.

The data for this study is split into two categories. The goal of the in-sample data, which spans starting January 2012 toward December 2022, is to identify the forecasting model combinations that work best.

The performance indicator is used to evaluate the data beginning January 2023 toward December 2023 then choose the top forecasting models.

There is a pattern in the data, as shown by the time series plot in Figure (1), which appears a little upward movement of the data throughout the given period of time. as shown in Figure (1).

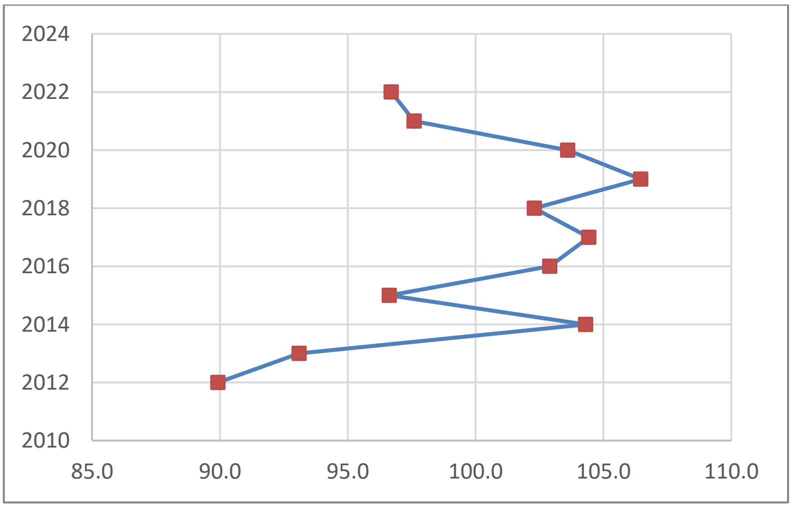

Figure (1): Time series plot of yearly data

Furthermore, the data do show a regular pattern of up and down fluctuations over the course of a year. As a result, it was determined that the data exhibits annual seasonality.

The dataset's normality was examined using the Anderson Darling (AD) Test. The AD-Test's p-value, which is more than 0.06, is 0.406. In light of this, it may be said that the data followed a normal distribution and are appropriate for use as time series data.

The Figure (2) show the probability of rain percentage for yearly data by taken monthly mean for every year from 2012 to 2022. as shown in Figure (2).

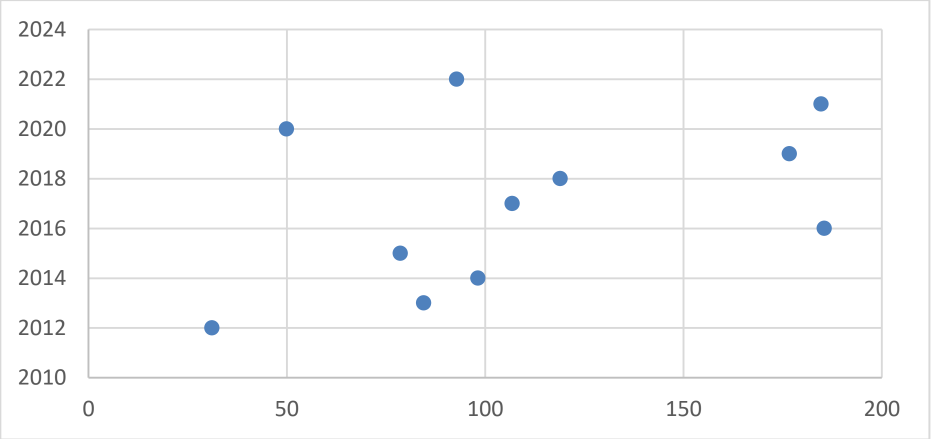

Figure (2): Probability plot of yearly data

SARIMA Model

In order to complete a Box-Jenkins analysis, the procedure is split it into five steps, which are checking for stationarity of data, model identification, coefficients and parameters estimations, model diagnostics, and model prediction. The process is carried out with the aid of Minitab [30, 31].

Figure (3) 's ACF plot and PACF plot for the original data show that there is a high association at the first season lag (lag 1) and that the correlation weakens over several seasons. as shown in Figure (3).

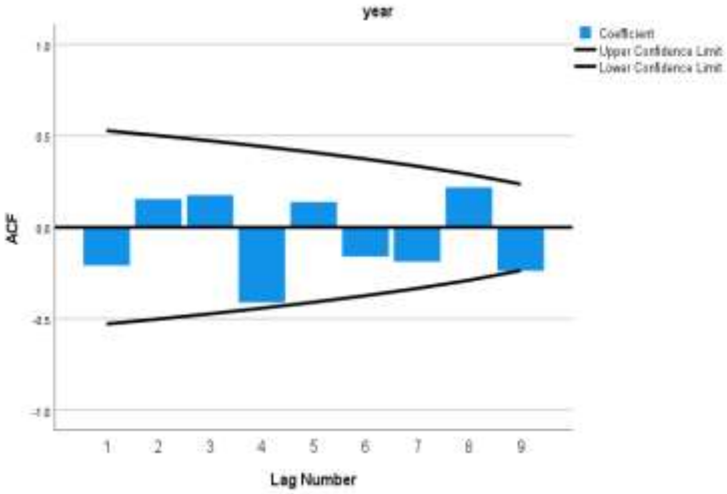

Figure (3): Autocorrelation of original data

The seasonal length should be used as the lag when comparing the data, which is 1. The dataset is in a stationary state after applying seasonal differencing, and the plots, the tail off pattern is not clearly shown as indicated in Figure (4). Both the significant spike at lag 1 in the PACF and the seasonal MA (1) components are indicated by the significant spike in the PACF. Since AR (1) and MA (1) are the simplest estimations of the model for stationary time series data, they are taken into consideration for the non-seasonal AR or MA portion. . as shown in Figure (4).

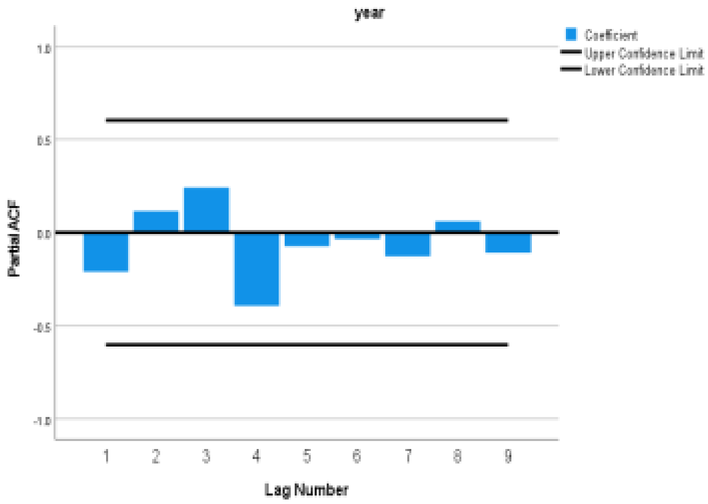

Figure (4): Partial autocorrelation of original data

A few models were suggested based on the outcomes of the model identification charts, and the values for AIC and BIC were determined to punish the models and prevent over-fitting.

The SARIMA (0,0,0) (0,1,1) 12 model was chosen among these models because, when compared to other models, it has the least AIC and BIC values.

According to the ACF plot and PACF plot of the residuals in Figures (5) and (6), none of the spikes are significant and are therefore not autocorrelated because they are all inside the boundaries. All four of these graphs for the residuals support the notion that the variance is constant and the mean of the residuals is zero. As a result, the remainders data almost follows a normal distribution. Consequently, it can be said that this model is suitable for use, and the anticipated value is shown in the center red line of the time series plot. as shown in Figures (5) and (6).

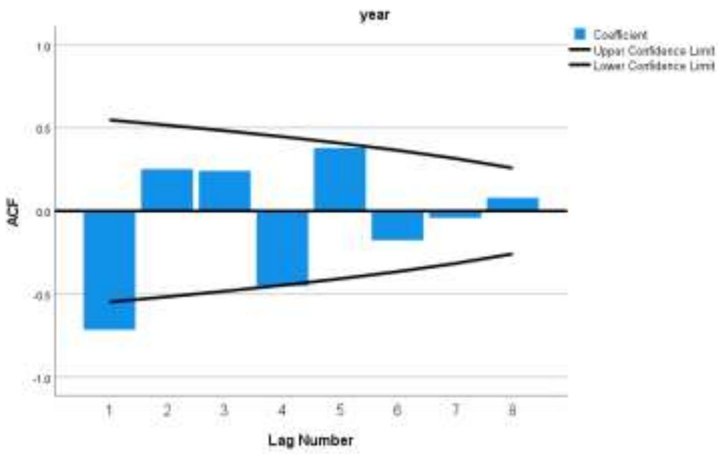

Figure (5): Autocorrelation of for Data After Seasonal Differencing (D=1)

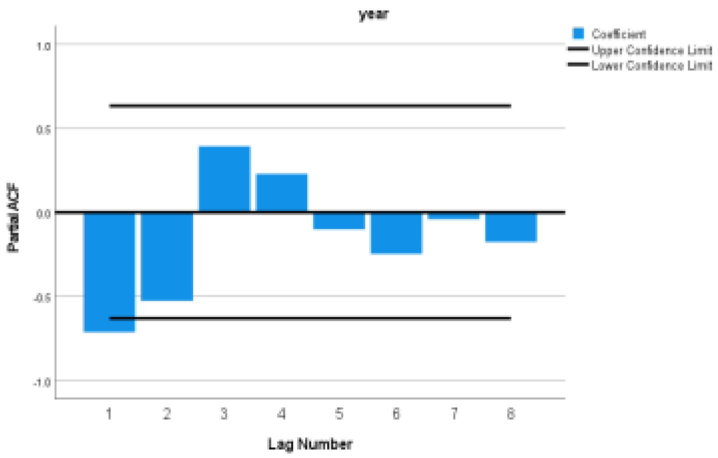

Figure (6): Partial autocorrelation of for Data after Seasonal Differencing (D=1)

Hybrid Model

SARIMA (0,0,0) $(0,1,1)_{12}$ model is chosen to build the hybrid model because it has the lowest AIC and BIC values, as explained in Table 1. The hybrid model is used to predict the rainfall data based on the strategy suggested in hybrid method. as shown in table (1)

Table (1) : Overfitting Models of SARIMA $(p, d, q)\ (P, D. Q)_{12}$

Model	significant	r	AIC	BIC
SARIMA (0,0,0)$(\mathbf{1,1,1})_{12}$	**NO**	**4**	**3.887**	**4.301**
SARIMA (0,0,0)$(\mathbf{1,1,0})_{12}$	**YES**	**3**	**4.312**	**4.462**
SARIMA (0,0,0)$(\mathbf{0,1,1})_{12}$	**YES**	**3**	**3.987**	**4.298**
SARIMA (1,0,1)$(\mathbf{1,1,1})_{12}$	**NO**	**6**	**3.923**	**4.451**
SARIMA (1,0,0)$(\mathbf{1,1,1})_{12}$	**NO**	**5**	**3.901**	**4.512**
SARIMA (1,0,0)$(\mathbf{1,1,0})_{12}$	**NO**	**4**	**4.054**	**4.582**

Model Evaluation

In Table 2 Taking the Mean absolute percentage error for ’SARIMA (0,0,0) Model $(\mathbf{0},\ \mathbf{1},\ \mathbf{1})\mathbf{12}$ Model, Hybrid Holt-Winters’ model. It implies that the Hybrid Holt-Winters’ model from table note that, Holt. Winters' is the best method in the prediction of rainfall. as shown in table (2)

Table (2): The Comparison of The Performance of Prediction Models

Holt-Winters’ Model	SARIMA (0,0,0) $(0,1,1)_{12}$	Hybrid model	Holt-Winters’ Model
RMSE	91.922	97.652	104.087
MAPE	66.645	83.142	62.741
MAE	80.543	80.784	87.999

The prediction model's accuracy increases with decreasing MAE, MAPE, and RMSE values. Holt-Winters model outperformed the other two models in terms of performance since it has the two lowest readings for the metrics, followed by the hybrid model and the SARIMA model as depicted as shown in Figure (7).

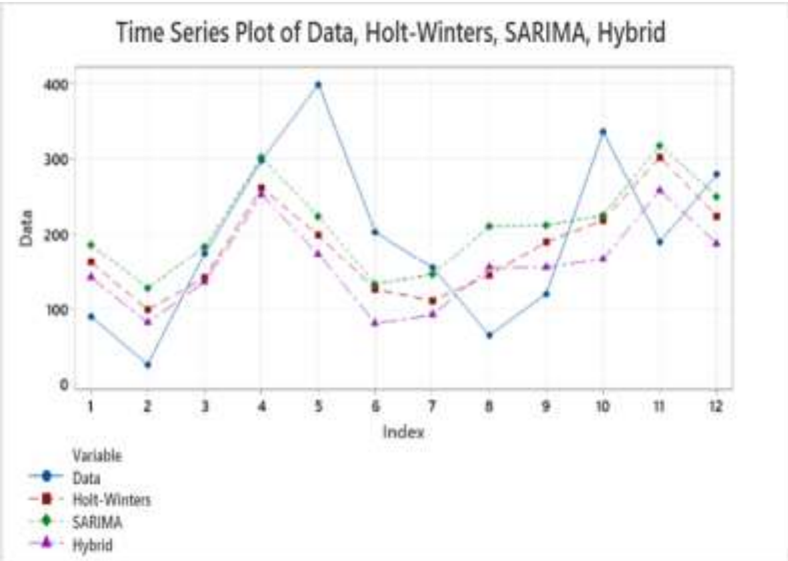

Figure (7): Comparison of forecasted out-sample results for all models

Conclusions

To forecast the rainfall data in this paper, Holt-Winters, SARIMA, and a hybrid model were suggested and applied. The Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) were applied to assess every model's performance. After data analysis, the Holt-Winters model has the least values for RMSE and MAE, proving that it is more accurate than other models that have been suggested. Actual data and predicted data appear different as a result of the research. For further study, a different forecasting technique is required so that the outcomes can accurately reflect the rainfall data.

References

[1] Farhan DA, Fakhir MR. Forecasting the Exchange Rates of the Iraqi Dinar against the US Dollar using the Time Series model

(ARIMA). International Journal of Engineering and Management Research. 2019;9(6):50-55.

[2] Shadab A, Said S, Ahmad S. Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study. International Journal of Energy and Water Resources. 2019;3(4):305-18.

[3] Dhamodharavadhani S, Rathipriya R, editors. Region-Wise Rainfall Prediction Using MapReduce-Based Exponential Smoothing Techniques. Advances in Big Data and Cloud Computing; 2019; Singapore: Springer.

[4] Celestial J. Worst flooding in 50 years leaves 6 dead, 50 000 displaced in Malaysia: The Watchers - Daily News Service 2021 [Available from: https://watchers.news/2021/01/11/malaysia-flood-january-2021.

[5] Ghayyib MN, Al-Rubaie AIF, Adnan FA. A Statistical Analysis of the Effects of Afforestation on the Environment in Iraq (Northern Iraq). IOP Conference Series: Earth and Environmental Science. 2023;1215(1):012039.

[6] Hadwan M, Al-Maqaleh BM, Al-Badani FN, Khan RU, Al-Hagery MA. A hybrid neural network and box-jenkins models for time series forecasting. CMC-Comput Mater Contin. 2022;70:4829-45.

[7] Abdul-Aziz A, Anokye M, Kwame A, Munyakazi L, Nsowah-Nuamah N. Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: The case of Ashanti region. International Journal of Humanities and Social Science. 2013;3(3):224-33.

[8] Mahsin MD. Modeling rainfall in Dhaka division of Bangladesh using time series analysis. Journal of mathematical modelling and application. 2011;1(5):67-73.

[9] Suhartono S. Time series forecasting by using seasonal autoregressive integrated moving average: Subset, multiplicative or additive model. J. Math. Stat. 2011;7:20-27.

[10] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015.

[11] Mado I, Soeprijanto A, Suhartono S. Applying of double seasonal ARIMA model for electrical power demand forecasting at PT. PLN Gresik Indonesia. International Journal of Electrical and Computer Engineering. 2018 Dec 1;8(6):4892.

[12] Humaira SP, Nursuprianah I, Darwan D. Forecasting of the number of schizophrenia disorder by using the Box-Jenkins of time series analysis. Journal of Robotics and Control (JRC). 2020;1(6):213-9.

[13] Stephanie. ARIMA (Box-Jenkins Models): Autoregressive Integrated Moving Average: Statistics How To; 2016 [Available from: https://www.statisticshowto.com/arima/.

[14] Ostertagova E, Ostertag O. Forecasting using simple exponential smoothing method. Acta Electrotechnica et Informatica. 2012;12(3):62.

[15] Zhang Z, Moore J. Autoregressive moving average models. Mathematical Physical Fundamentals of Climate Change. Boston, MA, USA: Elsevier; 2015. p. 239–90.

[16] Ediger VŞ, Akar S. ARIMA forecasting of primary energy demand by fuel in Turkey. Energy policy. 2007;35(3):1701-8.

[17] Mahmud I, Bari SH, Rahman M. Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average

method. Environmental Engineering Research. 2017;22(2):162-8.

[18] Proietti T, Lütkepohl H. Does the Box–Cox transformation help in forecasting macroeconomic time series? International Journal of Forecasting. 2013;29(1):88-99.

[19] Momani P, Naill P. Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. American Journal of Environmental Sciences. 2009;5(5):599.

[20] Ponnamperuma N, Rajapakse L, editors. Comparison of time series forecast models for rainfall and drought prediction. Moratuwa Engineering Research Conference (MERCon); 2021: IEEE.

[21] Grigonytė E, Butkevičiūtė E. Short-term wind speed forecasting using ARIMA model. Energetika. 2016;62(1-2):45-55.

[22] Kamisan NA, Lee MH, Hassan SF, Norrulashikin SM, Nor ME, Rahman NH. Forecasting Wind Speed Data by Using a Combination of ARIMA Model with Single Exponential Smoothing. Mathematical Modelling of Engineering Problems. 2021;8(2): 207-212.

[23] Karmaker C, Halder P, Sarker E. A study of time series model for predicting jute yarn demand: case study. Journal of industrial engineering. 2017;2017:8.

[24] Liu Q, Liu X, Jiang B, Yang W. Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC infectious diseases. 2011;11(1):1-7.

[25] Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015.

[26] Diodato N, Bellocchi G. Using historical precipitation patterns to forecast daily extremes of rainfall for the coming decades in Naples (Italy). Geosciences. 2018;8(8):293.

[27] Berberich D. Hybrid Methods for Time Series Forecasting: inovex; 2020 [Available from: https://www.inovex.de/de/blog/hybrid-time-series-forecasting.

[28] Safi SK, Sanusi OI. A hybrid of artificial neural network, exponential smoothing, and ARIMA models for COVID-19 time series forecasting. Model Assisted Statistics and Applications. 2021;16(1):25-35.

[29] Iraq Mete. Organ. Records of meteorological data for the period 1980-2012. 2012.

[30] Xie Y, Jin M, Zou Z, Xu G, Feng D, Liu W, Long D. Real-time prediction of docker container resource load based on a hybrid model of ARIMA and triple exponential smoothing. IEEE Transactions on Cloud Computing. 2020;10(2):1386-401.

[31] Bisgaard S, Kulahci M. Time series analysis and forecasting by example. John Wiley & Sons; 2011.