American Journal of Innovative Research and Applied Sciences. ISSN 2429-5396 I www.american-jiras.com

ORIGINAL ARTICLE

BUILDING A FORECASTING MODEL FOR ANNUAL RAINFALL IN ALHWAIZ BASIN USING BOX-JENKINS METHODOLOGY

| Lama Nizar Shiban [1] | Ali Mohamad Alasaad [2] | and | Abbas Abdulrahman Abdulrahman [3] |

[1.] Tishreen University | department of Water Engineering and Irrigation | Lattakia | Syria |
[2.] Tishreen University | department of Water Engineering and Irrigation | Lattakia | Syria |
[3.] Tishreen University | department of Water Engineering and Irrigation | Lattakia | Syria |

| Received | 16 December 2018 | | Accepted 21 January 2019 | | Published 27 January 2019 | | ID Article | Shiban-ManuscriptRef.1-ajira160119-Seriya-1 |

ABSTRACT

Background: Water is a basic element of life and it has become increasingly needed nowadays, especially in areas with low rainfall, due to the great development in various aspects of life. **Objectives**: The objective of this research is to study the rainfall in ALHWAIZ basin to explain the behavior of the time series and applying one model of Box-Jenkins models to predict the future amounts. **Methods:** The Statistical analysis study was based on annual rainfall data for 59 years, and Through it we found the best model to represent the data after doing the wanted tests on model residuals. also, we dividing the data to 53 years to build the model and 6 years to test it, and finding the best forecasted model depending on the smallest of weighted mean of criteria RMSE, MAP, MAE. **Results:** This research finds that Autoregressive Integrated Moving Average Model (ARIMA Model) ARIMA (1,1,3) model is a good representation of the data, and ARIMA (2,1,0) model is the right model to forecast future rainfall which was decreasing amounted to 6.13 mm per year during the studied period. **Conclusions:** The study reveals that ARIMA model can be used as a good forecasting tool in forecasting the annual rainfall in ALHWAIZ basin.

Keywords: *ALHWAIZ basin, ARIMA Model, Rainfall.*

1. INTRODUCTION

The analysis of time series is one of the most important statistical methods, because it enables to know the changes in the studied phenomenon over time and forecasting the future values of it [1]. That important because it decrease the effects of extreme events such as floods and drought [2]. The time series can be stationary and non-stationary, and the series is considered stable if it has constant (mean and variance) [3]. Many researchers have under taken many studies to forecast rainfall using different techniques [4]. Because it is considered one of the most difficult elements of the hydrological cycle due to the complexity of the air operations that contribute to the formation of rainfall [5]. We will present in this research process of building a forecasting model for the rainfall of ALHWAIZ basin by using Box-Jenkins method, which is one of the best methods that give good results.

2. MATERIALS AND METHODS

2.1 Study site & data available

ALHWAIZ basin is located on the eastern coast of the Mediterranean Sea in north-west of Syria. The annual rainfall data from 1959 until 2017 is measured for three rain stations that are covering the whole basin (Skhaba, Hmemem, Ein sharqiyeh).

2.2 ARIMA Model: The Autoregressive Integrated Moving Average Model (ARIMA Model) is a time series model used for short-term forecasting. General form of ARIMA (p,d,q) model is following as Eq.(1). [6]:

$$\emptyset B\ (1-B)^d Z_t = \delta + \theta(B)a_t \qquad (1)$$

Where: $\emptyset B = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots \emptyset_p B^p$

: $\theta B = 1 - \theta_1 B - \theta_2 B^2 - \theta_q B^q$

$\emptyset_p, \theta_q$: are parameters; B: backward operator; d: order of differences; Z_t: the actual value; δ: model parameter; a_t: independent random variable represent the error term at time t; p: the order of autoregressive process; q: the order of moving average [6, 7].

2.3 Box-Jenkins methodology: It was used to analyze the rainfall time series by using Minitab software [8]. This method can identify the order of ARIMA model, estimation, diagnostic checking and forecasting [9].

2.4 Modeling procedure: We plot the data and transform it if necessary, whereas Box-Jenkins method is only occasion for stationary time series data [10]. Models can be determined by the plots of Autocorrelation function (ACF) and Partial autocorrelation function (PACF), where it gives more information about behavior of time series [11]. We use Akaike's Information Criterion (AIC) in Eq.(2) to search for the best model [12]. After that we select min (AIC) as the best model that determines a model by how close its fitted values tend to be to the natural values in terms of a specific expected value [13].

$$AIC(m) = n * ln\ (\sigma^2) + 2 * m \qquad (2)$$

Where m: number of parameters in the model; n: number of observation; $\boldsymbol{\sigma}^2$: variation of residuals.

The most important part is to make sure that the residuals of the model random, and the estimated parameters are statistically significant [14]. So, we check the residuals from our chosen model by plotting (ACF) and (PACF) of the residuals and doing many tests to analysis model's residuals [15]. If the model is not suitable, the cycle repeats from the identification of the model [16]. It is not necessary that representation model is the same as the future forecasting model, so we depend on forecasting model to prediction [1].

Finally, we generate forecasts according to the final forecasting model, and depending on forecasting criteria which are: Root Mean Squared Error (RMSE) Eq.(3), Mean Absolute Deviation (MAD) Eq.(4), and Mean Absolute Percentage Error (MAPE) Eq.(5), and calculating the smallest of weighted mean of them which is given by Eq.(6) to Eq.(8) [2].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_t - \hat{Y})^2}{n}} \qquad (3)$$

$$MAD = \frac{1}{n}\sum_{t=1}^{n}|Y_t - \hat{Y}_t| \qquad (4)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \qquad (5)$$

Where: Y_t is the actual value; $\hat{Y}_t$: is the simulated value.

Weighted Mean 1 = (RMSE +4 MAPE+MAD)/6 (6)

Weighted Mean 2 = (4RMSE + MAPE+MAD)/6 (7)

Weighted Mean 3 = (RMSE + MAPE+4MAD)/6 (8)

3. RESULTS

3.1 Build Representation Model:

We used Minitab software to analyze annual rainfall (1959-2017). In figure 1 the graph trend of original data showed non-stationary data, because the data is decreasing along the axis of time, and it showed the value of decreasing=6.13mm/yr. the graph of original data of (ACF) showed non-seasonal and non-static data. In figure 2 we test stability of series in variance by applying normality test (Kolmogorov- Smirnov test) that showed the data is stable in variance. We take the first regular difference (d-1) to stabilize the variance in mean and remove the trend. A graph of time plot of annual rainfall after doing the difference shows stationary on the value of the data center, whereas the trend is decreasing at a fairly constant rate.

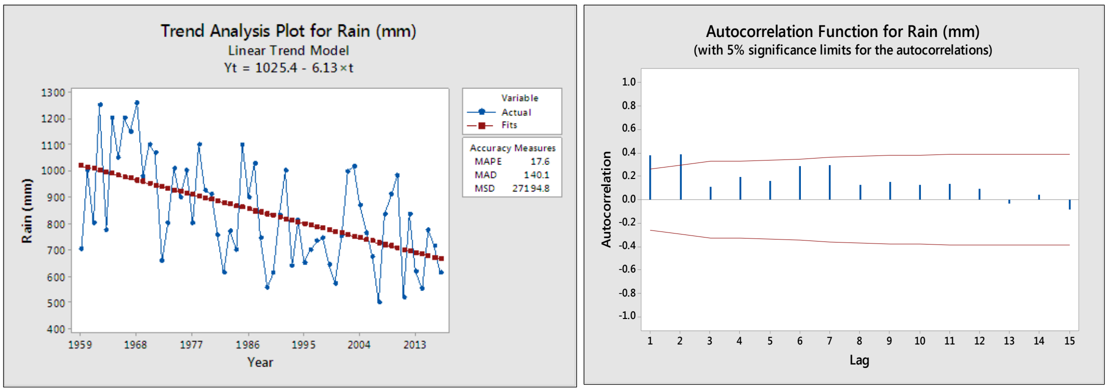

Figure 1: The figure presents trend of original data and the plot of Autocorrelation Function.

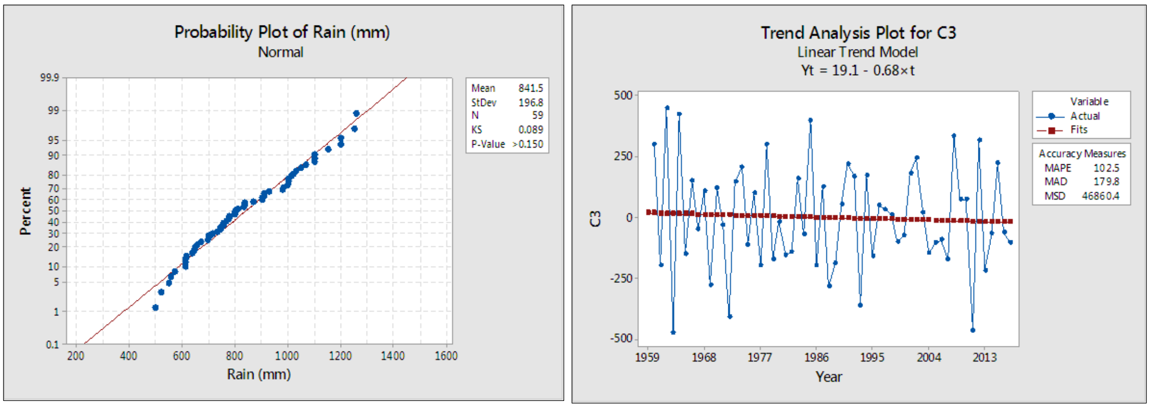

Figure 2: The figure presents normality test and the plot of trend data after in deference.

To determine the best model, we use min(AIC) and choose ARIMA (1,1,3) that shown in table (1), and considers it the best model due to the residuals passed the required tests.

Table 1: The table presents the AKAKI test results.

ARIMA Model	σ^2	m	AIC	ARIMA Model	σ^2	m	AIC
ARIMA (0,0,1)	35939.1	2	622.8853	ARIMA (0,1,1)	27255.8	2	606.5683
ARIMA (1,0,0)	33575.2	2	618.871	ARIMA (1,1,0)	35680.4	2	622.4591
ARIMA(2,0,0)	31189.0	3	616.5214	ARIMA (2,1,0)	36317.8	3	625.5037
ARIMA(3,0,0)	31085.6	4	618.3255	ARIMA (3,1,0)	33276.1	4	622.3431
ARIMA(4,0,0)	30892.0	5	619.9569	ARIMA (4,1,0)	32263.6	5	622.52
ARIMA(1,0,2)	31475.9	4	619.0617	ARIMA (1,1,2)	28125.8	4	612.4221
ARIMA(2,1,3)	23374.1	6	605.5036	ARIMA (1,1,3)	24130.5	5	605.3827
ARIMA(4,1,5)	24449.0	10	616.1563	ARIMA (5,2,3)	35170.6	9	635.61
ARIMA(5,1,1)	26198.5	7	614.234	ARIMA (5,2,4)	29880.9	10	627.9935
ARIMA(1,2,0)	63770.6	2	656.7198	ARIMA (4,2,1)	56264.9	6	657.3317
ARIMA(2,2,0)	59753.6	3	654.8811	ARIMA (4,2,2)	55696.2	7	658.7324
ARIMA(3,2,0)	54499.8	4	651.4512	ARIMA (4,2,3)	34413.1	8	632.3254

ARIMA Model: Autoregressive Integrated Moving Average Model.

In tables (2) and (3), we test the residuals of the selected model by using 1-Sample Z and Runs Test, whereas P-value >0.05. The figure 3 explained that the residuals look like white noise (WN) by observation the plots of (ACF, PACF), and we draw Histogram of residuals and apply normality test of them.

Table 2: The table presents the one-Sample Z results.

N	Mean	St Dev	SE Mean	P-Value
58	-3	150	3168	0.999

Table 3: The table presents the runs Test results.

N	Number of observations> Mean	Number of observations< Mean	P-Value
58	31	27	0.818

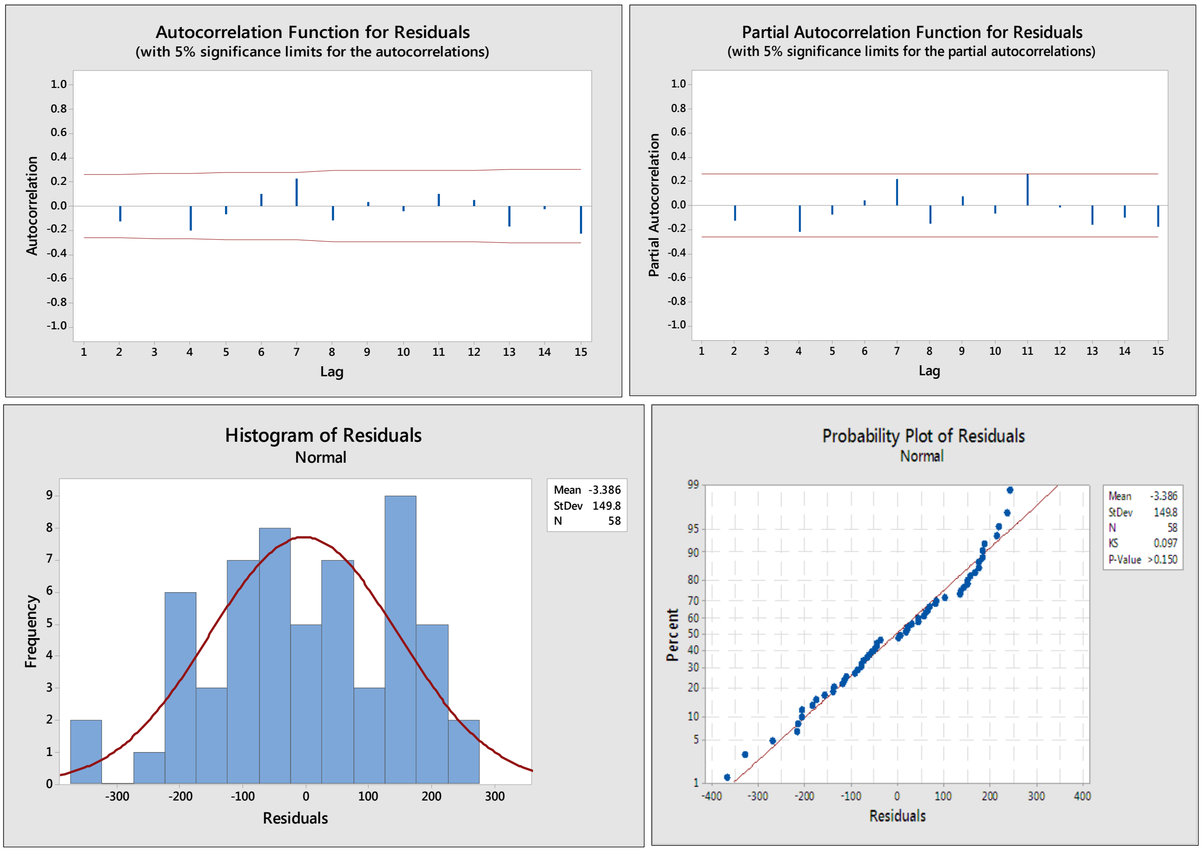

Figure 3: The figure presents the plots of (ACF, PACF) of residuals, Histogram of residuals and normality test of them.

By equal 1, we write the final shape of representation model that shown in Eq.(9).

$$z_t = (1 - 0.737)Z_{t-1} + 0.737Z_{t-2} - 13.809 + a_t - 0.134 * a_{t-1} - 0.55 * a_{t-2} - 0.416 * a_{t-3} \quad : $$
$$a_t \approx WN(0{,}24130.5) \qquad (9)$$

3.1 Build a Forecasting Model:

We used the data of annual rainfall (1959-2011) to build a forecasting model, and the data from 2012 to 2017 to verify the quality of model. The figure 4 shown the comparing between measured values and forecasted values after proposing several models to find the best model. Calculating of forecasting criteria as in table (4). Table (5) contains a conclusion of weighted Mean values of criteria to find ARIMA (2,1,0) as the best forecasting model that have the least values (6.41, 21.408, 10.627) and it gives us forecasted rainfall for (2018-2019-2020) with (651.18-644.1- 644) mm values. Table (6) shows the error ratios between original data and forecasting values of ARIMA model.

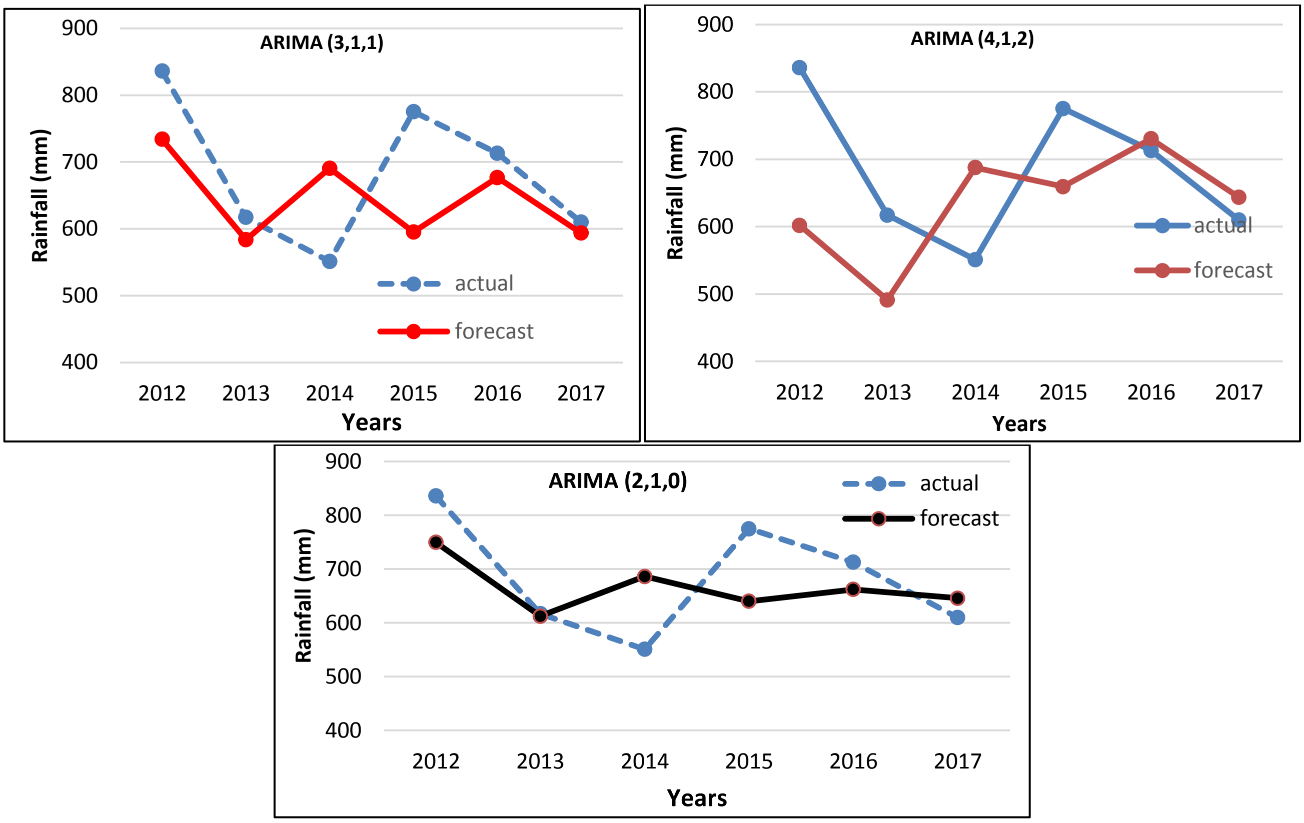

Figure 4: The figure presents the comparing between actual and forecasted values for many models.

Table 4: The table presents the MAPE, RMSE, MAD Criteria.

Model	MAPE	RMSE	MAD
ARIMA (3,1,1)	0.0139	35.027	9.581
ARIMA (4,1,2)	0.018	44.323	12.514
ARIMA (2,1,0)	0.012	30	8.438

ARIMA Model: Autoregressive Integrated Moving Average Model; **RMSE**: Root Mean Squared Error; **MAD**: Mean Absolute Deviation; **MAPE**: Mean Absolute Percentage.

Table 5: The table presents the Weighted Mean of (MAPE, RMSE, MAD) Criteria.

Model	Weighted Mean 1	Weighted Mean 2	Weighted Mean 3
ARIMA (3,1,1)	7.44	24.95	12.227
ARIMA (4,1,2)	9.48	31.637	15.732
ARIMA (2,1,0)	6.41	21.408	10.627

ARIMA Model: Autoregressive Integrated Moving Average Model; **RMSE**: Root Mean Squared Error; **MAD**: Mean Absolute Deviation; **MAPE**: Mean Absolute Percentage.

Table 6: The table presents the error ratios between original and forecasting data.

Original (mm)	Forecast (mm)	Error Ratio (%)	Year
836	749.71	10.32	2012
617	612.64	0.7	2013
551	686.27	-24.5	2014
775.26	640.4	17.39	2015
713	662.39	7.09	2016
610	645.85	-5.87	2017

4. DISCUSSION

We use Box-Jenkins method to predict annual rainfall for 2018-2019 and 2020 by analyzing the annual rainfall series from 1959 to 2017 which was annually decreasing during the studied period by 6.13 mm. We leaving the last 6 years as a model-testing set. We found Autoregressive Integrated Moving Average Model (ARIMA Model) (1,1,3) model as the best model to represent the data after applying the least value of AIC and doing the required tests on the model residuals and passing them successfully, whereas the autocorrelation does not exist between series residuals, and P>P-value (random residuals) and taken the natural shape and following the normal distribution.

Also, we concluded that ARIMA (2,1,0) is the best forecasted model depending on the smallest of weighted mean of forecasting criteria which were equal: RMSE=30, MAP=0.012 and MAD=8.438 for the chosen model. The forecasted values were very close to actual values; therefore, we recorded the rainfall values for years (2018 to 2020) that we obtained from the model whereas these predictions are reliable and ready to use.

5. CONCLUSION

There are many methods for forecast time series, but Box-Jenkins models is adopting to analysis the time series data, due to its high accuracy, and their compatibility with the characteristics of each region and therefore cannot be directly applied to a different area. Especially there is a lot of important effect for the natural characteristics of the areas in changing the amount of forecasting rainfall from year to year. If we used the model that we finding it, we always had to update it, because ARIMA is a short-term forecast method.

6. REFERENCES

1. Hayek S., Hammad M. Predict of the monthly water volumes incoming in AL-ROOS River in the Syrian Coast by using the time series analysis. *Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series*. 2016;38 (5):45-64. Available on : https://shamra.sy/academia/show/59f358a1231f3.

2. Ammar G., Hayek S., and Hamdan N.M. Building a forecasting model for annual rainfall in Husn Suleiman station using mathematical modeling. *Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series*. 2016;38 (4):491-508. Available on : https://shamra.sy/academia/show/5989ae787c60f.

3. Osarumwense O.I. Applicability of Box Jenkins SARIMA Model in Rainfall Forecasting: A Case Study of Port-Harcourt south south Nigeria. *Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine*. 2013;4 (1):1-4. Available on : https://www.researchgate.net/publication/280822361_Applicability_of_Box_Jenkins_SARIMA_Model_in_Rainfall_Forecasting_A_Case_Study_of_Port-Harcourt_south_south_Nigeria.

4. Ammar G.A., Haidar B.Y., and Aldarwish A.Q. AN ARTIFICIAL NEURAL NETWORK MODEL FOR MONTHLY PRECIPITATION FORECASTING INHOMS STATION, SYRIA. *American Journal of Innovative Research and Applied Sciences*. 2017;4 (6):240-246. Available on : https://www.american-jiras.com/Ghatfan-ManuscriptRef.1-ajira160517.pdf.

5. Hammad M. Building a forecasting model for monthly rainfall in Slonfe station located in northern Syrian coast. *Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series*. 2017;39 (5):77-96. Available on : http://journal.tishreen.edu.sy/index.php/engscnc/article/download/4185/3963http://journal.tishreen.edu.sy/index.php/engscnc/article/download/4185/3963.

6. Ohyver M., Pudjihastuti H. Arima Model for Forecasting the Price of Medium Quality Rice to Anticipate Price Fluctuations. *Elsevier, Science Direct*. 2018;135 :707-711. Available on : https://www.sciencedirect.com/science/article/pii/S1877050918315072/pdf?md5=b235ebfc5f05e20f56faa16f3d80cda0&pid=1-s2.0-S1877050918315072-main.pdf.

7. Zakaria S., Alansari N., Knutsson S., and Albadrany T. ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq. *Journal of Earth Sciences and Geotechnical Engineering*. 2012;2 (3):25-55. Available on : http://www.diva-portal.org/smash/get/diva2:979693/FULLTEXT01.pdf.

8. Hayek S., Ammar G. Rainfall Prediction in Tartous Station Located in the Southern Part of the Syrian Coast. *Tishreen University Journal for Research and Scientific Studies - Engineering Sciences Series*. 2015;37 (2):74-90. Available on : https://shamra.sy/academia/show/579b76b0a9100.

9. Suhermi N., Suhartono., Prastyo D.D., and Ali B. Roll motion prediction using a hybrid deep learning and ARIMA model. *Elsevier, Science Direct*. 2018;144 :251-258. Available on : https://www.sciencedirect.com/science/article/pii/S187705091832235X/pdf?md5=07f347074bbfe7412d4811e243e82316&pid=1-s2.0-S187705091832235X-main.pdf.

10. Arumugam P., Saranya R. Outlier Detection and Missing Value in Seasonal ARIMA Model Using Rainfall Data. *Elsevier, Science Direct*. 2018;5 :1791-1799. Available on: https://www.sciencedirect.com/science/article/pii/S2214785317325452/pdf?md5=7be4eb9c236af895da2a7e8b22a9b2d8&pid=1-s2.0-S2214785317325452-main.pdf.

11. Ghalhari G.A.F., Bayatani F., and Fahiminezhad E. Comparing the Forecasting Accuracy of the Box–Jenkins Models in Modeling Seasonal Precipitation (Case Study: The South of Kerman Province, Iran). *Journal of Applied Environmental and Biological Sciences*. 2015;5 (12):64-78. Available on: https://pdfs.semanticscholar.org/a26c/f228711e438bca809ca3993a839139e7a6db.pdf.

12. Awchi T.A., Hasan I.F. Low flow modeling for Greater Zab and Lesser Zab rivers in Northern Iraq. *Al-Rafadain Engineering Journal*. 2014;22 (3):108-120. Available on : https://www.iasj.net/iasj?func=fulltext&aId=88206.

13. Sarpong S.A. Modeling and Forecasting Maternal Mortality; an Application of ARIMA Models. *International Journal of Applied Science and Technology*. 2013;3 (1):19-28. Available on : https://www.ijastnet.com/journals/Vol_3_No_1_January_2013/3.pdf.

14. Abdulaziz A.R., Anokye M., Kwame A., Munyakazi L., and Nuamah N.N.N.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):224-233. Available on: http://www.ijhssnet.com/journals/Vol_3_No_3_February_2013/22.pdf.

15. Reddy J.C.R., Ganesh T., Venkateswaran M., and Reddy P.R.S. Forecasting of Monthly Mean Rainfall in Coastal Andhra. *International Journal of Statistics and Applications*. 2017;7 (4):197-204. Available on : https://www.researchgate.net/publication/319135842_Forecasting_of_Monthly_Mean_Rainfall_in_Coastal_Andhra.

16. Junior P.R., Salomon F.L.R., and Pamplona E.O. ARIMA: An Applied Time Series Forecasting Model for the Bovespa Stock Index. *Applied Mathematics*. 2014;5 :3383-3391. Available on : https://www.scirp.org/journal/PaperDownload.aspx?paperID=52119.

Cite this article: Lama, Nizar Shiban, Ali Mohamad Alasaad, and Abbas Abdulrahman Abdulrahman. BUILDING A FORECASTING MODEL FOR ANNUAL RAINFALL IN ALHWAIZ BASIN USING BOX-JENKINS METHODOLOGY. *American Journal of Innovative Research and Applied Sciences*. 2019; 8(1): 18-23.