Time Series Analysis of Baghdad Rainfall Using ARIMA Method

Sawsan M. Ali*
Department of Geology, College of Science, University of Baghdad, Baghdad, Iraq

Abstract
Monthly rainfall data of Baghdad meteorological station were taken to study the time behavior of these data series. Significant fluctuation,very slight increasing trend and significant seasonality were noticed. Several ARIMA models were tested and the best one were checked for the adequacy. It is found that the SEASONAL ARIMA model of the orders SARIMA(2,1,3)x(0,1,1) is the best model where the residual of this model exhibits white noise property, uncorrelateness and they are normally distributed. According to this model, rainfall forecast for four years was also achieved and showing similar trend and extent of the original data.

Keywords: ARIMA, Time Series, Baghdad

تحليل السلاسل الزمنية لامطار بغداد باستخدام تقنية ARIMA

سوسن مجيد علي
قسم علوم الارض، كلية العلوم، جامعة بغداد، بغداد، العراق.

الخلاصة :
استخدمت البيانات الشهرية للامطار في محطة بغداد للأنواء الجوية لدراسة السلوكية الزمنية لها حيث تم ملاحظة تذبذبات واضحة فيها مع وجود نمط تزايدي طفيف جدا اضافة الى الدورية الواضحة فيها. تم اختبار نماذج متعددة لل ARIMA واخضاع النموذج الافضل لاختبار الوائمة. لوحظ ان افضل نموذج كان ذا المعالم SARIMA(2,1,3)x(0,1,1) اذ تعكس بواقي هذا النموذج خصائص سلسلة الضجة البيضاء. تم استخدام هذا النموذج للتنبؤ بقيم الامطار لفترة اربع سنوات قادمة بعد التحقق من موائمته لتمثيل المعلومات الاصلية

Introduction:-

Time series forecast is one of the most important tools in the water resources management field. Various methods have been used for forecasting purposes. The shorten variations in the time series can be studied by either autoregressive (AR) and or Moring average (AM) approaches. It is necessary, when generating synthetic data for a hydrologic variable, to take into consideration the data that are similar to the observed data, [1]. The most important time series model is the Box- Jenkins approach. This approach known also as ARIMA autoregressive integrated moving averages and it has been widely used for the simulation of many hydrologic and meteorological variables worldwide. If the time series shows seasonal component, ARIMA model should be expended to include this component and then called seasonal ARIMA, SARIMA.

Iraq has suffered from water deficit since the last years where significant decrease in the amount of rainfall were recorded and the prediction of rainfall was the main subject of several works. For the purpose of the present study,, data of rainfall of Baghdad station for the

*Email: sawsanjaff@hotmail.com

period 1980-2012 were used to analyze the rainfall behavior using ARIMA technique. These data have been provided by Iraqi Meteorological Organization. [2]

Methodology:-

Box–Jenkins, 1976, proposed a method for analyzing time series data consists of four steps, i.e, i.) model identification ii) estimation of model parameters, iii) diagnostic checking for the identified model appropriateness for modeling and v) application of the model in forecasting purposes.[3]

ARIMA model has the parameters of p, d, q which represent the order of the autoregressive part (AR) ; degree of differencing involved and order of the moving average part (MA) respectively [4].

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

$$Y_t = C + ØY_{t-1} + Ø_2 Y_{t-2} + \cdots Ø_p Y_{t-p} + e_t - θ_1 e_{t-1} - θ_2 e_{t-2} - \cdots θ_p e_{t-p} \ldots\ldots(1)$$

Or in backshift notation,

$$(1 - Ø_1 B - Ø_2 B^2 - \cdots Ø_p B^p)\, Y = C + (1 - θ_1 B - θ_2 B^2 - \cdots θq B^q)\, e_t \ldots\ldots\ldots.(2)$$

Where :

C=constant term, $Ø_i$=j^{th} autoregressive parameter, $θ_j$=j^{th} moving average parameter, e_t= error term at time t, and $β_k$=k^{th} order backward shift operator.

Seasonal ARIMA (P, D, Q) parameters may also identified for specific time series data. These are seasonal autoregressive (P), seasonal differencing (D) and seasonal moving average (Q). The general expression of the seasonal ARIMA model (p, d, q) (P, D, Q)$_s$ is given by:

$$Ø_{AR}(B)\, Ø_{SAR}(B^s)\, (1-B)^d\, (1-B^s)^D .y_t = θ_{MA}(B)\, θ_{SMA}(B^s) .e_t \ \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Where

S = no. of periods in season

$Ø_{AR}$ = non-seasonal autoregressive parameter

$θ_{AM}$ = non-seasonal moving average parameter and

$θ_{SAM}$ = seasonal moving average parameter

The first step in developing Box- Jenkins model is to check the stationary of the time series and if there is any signification seasonality that need to be modeled. The Autocorrelation Function (ACF) and the partial Autocorrelation Function (PACF) are the most important elements of time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag K. The PACF plot helps to determine how many autoregressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of series, trend either in the mean level or in the variance of series. Stationary of data can also be identified by a number of tests. Ljung-Box is one of the most important tools and has the following expression:

$$Q^* = n(n+2)\sum_{k=1}^{n} \frac{r_k^2}{n-k} \cdots]\ (4)$$

It has a distribution closer to chi-square distribution with (h-m) degree of freedom then dose the Q statistic. The data are not white noise if the value of Q or Q* lies in the extreme 5% of the right hand tail of the X^2 distribution,[5].

The best model among the studied models could be selected according to specific criteria. Akaike's Information Criterion (AIC) is a criterion normally used for this purpose where the best one has the minimum AIC value, which has the following expression:

$$AIC(min) = n\ln σ^2 + 2m \ \ldots.(5)$$

Where: n: no. of observations, σ:standard deviation and m is the no. of model parameters.

The second step is the model parameter estimation in which least square and maximum Likelihood techniques can be used. The third step is the diagnostic check in which the residuals from the fitted model should be examined against adequacy which is done be correlation analysis through the residual ACF plots. If the residuals are correlated, then the model should be refined as in step one. Otherwise the autocorrelations are white noise and the model is adequate to represent our time series,[6].The calibrated model can be used for forecasting purposes.

Results and Discussions:

Monthly rainfall data of Baghdad International airport station were used to study the behavior of time series of Baghdad area. Figure 1- shows the plot of the monthly rainfall values.

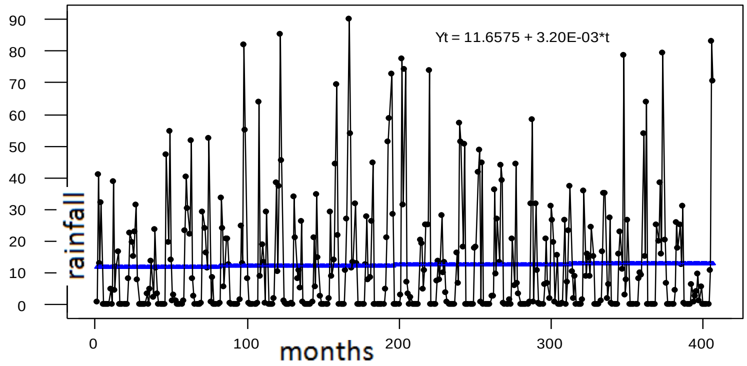

Figure 1- Ttime series and trend of Baghdad rainfall for the period 1980-2012.

As appeared from this figure, there are significant fluctuations in these values and very slight increasing trend can be noticed. It is necessary to de-trend these data before pursuing the analysis. Plots of ACF and PACF of Baghdad time series figures (2 and 3) show non-stationary and clear seasonality variation in the present data set,

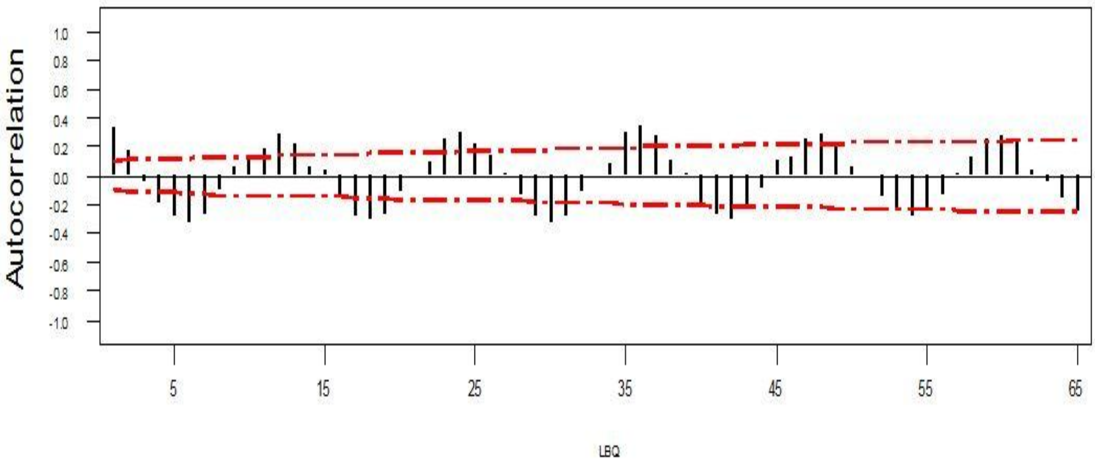

Figure 2- Autocorrelation function of (ACF) Baghdad rainfall series.

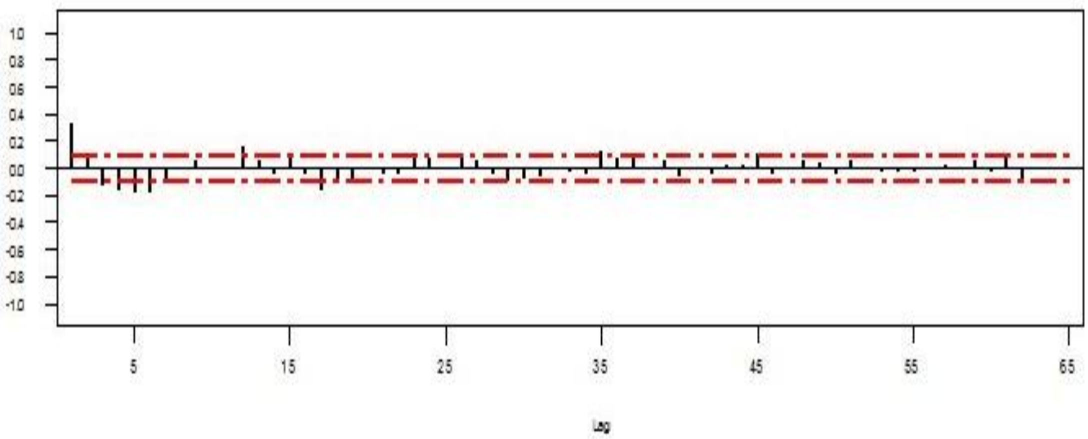

Figure 3- Partial autocorrelation function (PACF) of Baghdad rainfall series.

Therefore, the original series need to be differenced to make the series stationary. Figures (4 and 5) shows the differenced series of order one which transformed into stationary series and that all the lags are within the confidence limits.

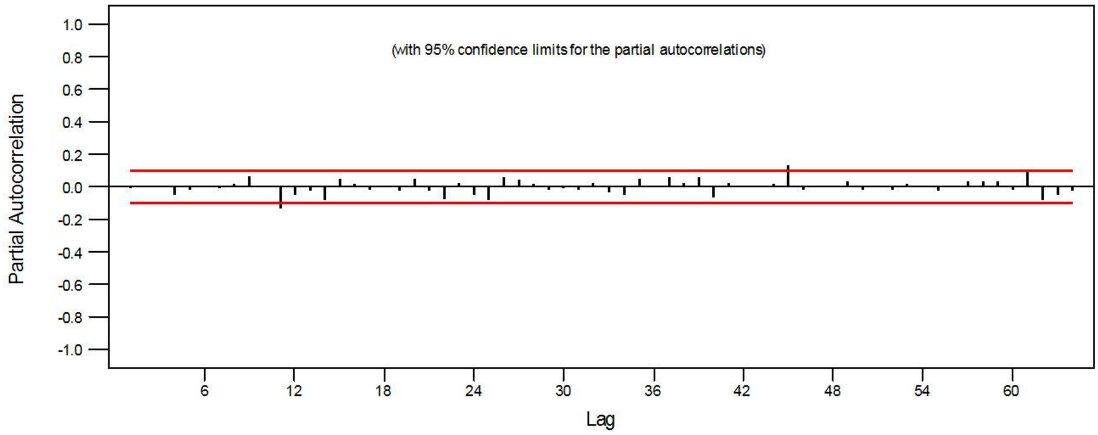

Figure 4- Autocorrelation function of Baghdad rainfall series after differencing of order one.

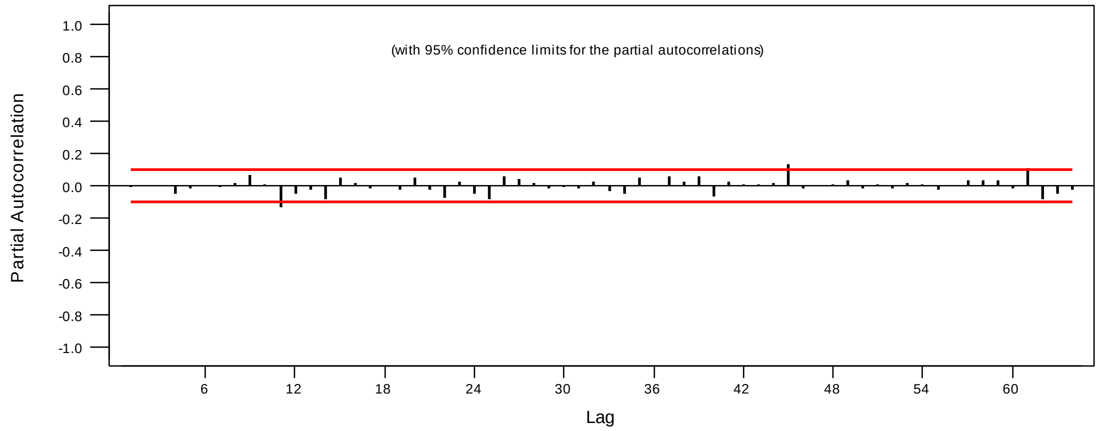

Figure 5- Partial autocorrelation function of Baghdad rainfall series after differencing of order one.

However, the series still need to remove the seasonality effect. Figure 6- shows the de-trended and de-seasonalized series which is now constant around the mean and variance. Months

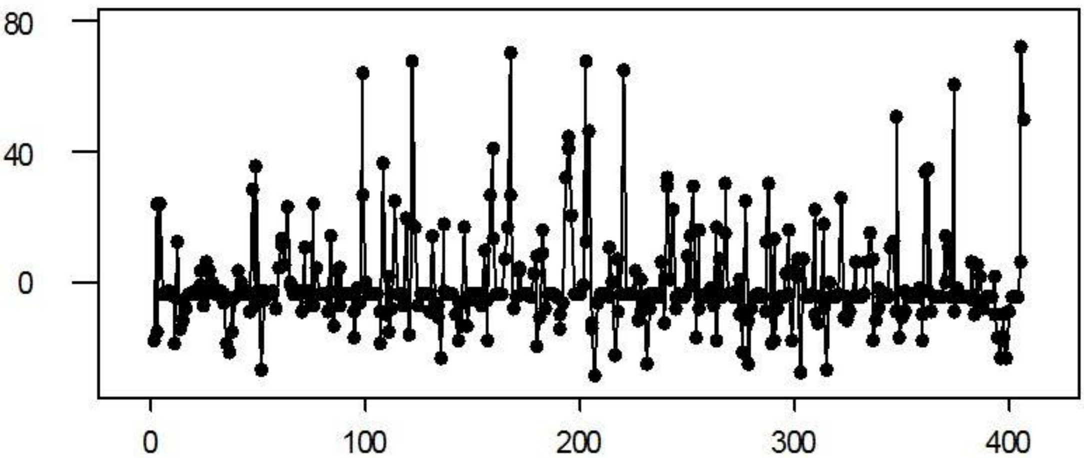

Figure 6- De-trended and de-seasonalised series of Baghdad rainfall.

Several SARIMA models have been tested and checked for the adequacy. Due to the above reasons, SARIMA model of different parameters can be identified. It should be mentioned that if the best ARIMA model identified, this doesn't mean that this model is the only model can be considered in which other ARIMA models with values of AR and MA less than the same parameters of the considered ARIMA models [7]. This shows the need for a specific criterion to select the most realiable model. AIC was selected to test the best model after estimating its parameters by using the maximum likelihood technique. Table 1- shows the values of mean square errors and the AIC values of some tested ARIMA models with different parameters orders.

Table.1- Mean square and AIC values of selected models

SARIMA	Mean Square	AIC
(4,1,0)x(1,1,0)	406	2379
(2,1,0)x(0,1,0)	599	2526
(0,1,1)x(0,1,0)	487	2442
(3,1,0)x(0,1,1)	286	2238
(2,1,3)x(0,1,1)	**231**	**2158**
(4,1,2)x(0,1,1)	240	2175

According to the above table-1 , it was found that ARIMA model(2,1,3)x(0,1,1)$_{12}$ is the best model with a minimum value of the mean square error and then after the AIC value. It is very important to check the residual of the fitted model for adequacy purposes. This can be done by testing the ACF and PACF of these residuals at deferent lags. Figures (7and 8) show the plots of ACF and PACF of the residuals of the present data.

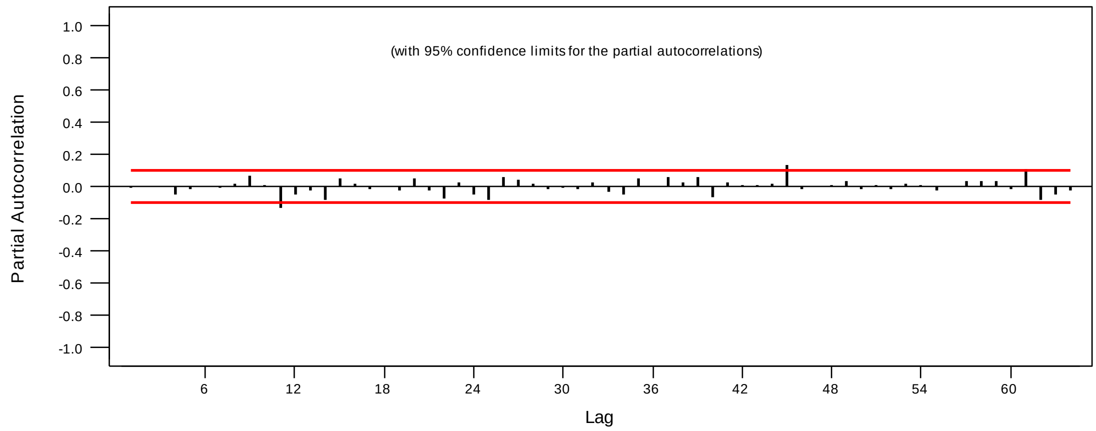

Figure 7- Autocorrelation function (ACF) of the residuals of Baghdad rainfall series.

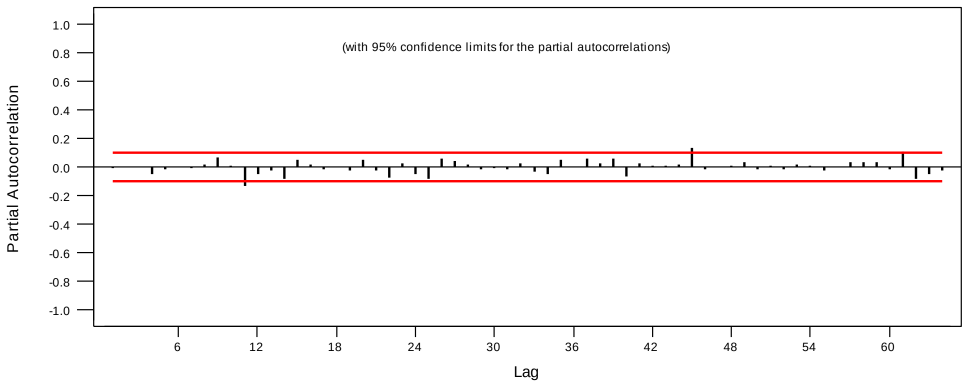

Figure 8- Partial autocorrelation function (PACF) of the residuals of Baghdad rainfall series.

Based on these results from the above figures, no significant correlation can be observed which , in turn exhibit white noise properties . Furthermore, the assumption of un-correlateness can also be tested by plotting the histogram of these residuals, figure 9.

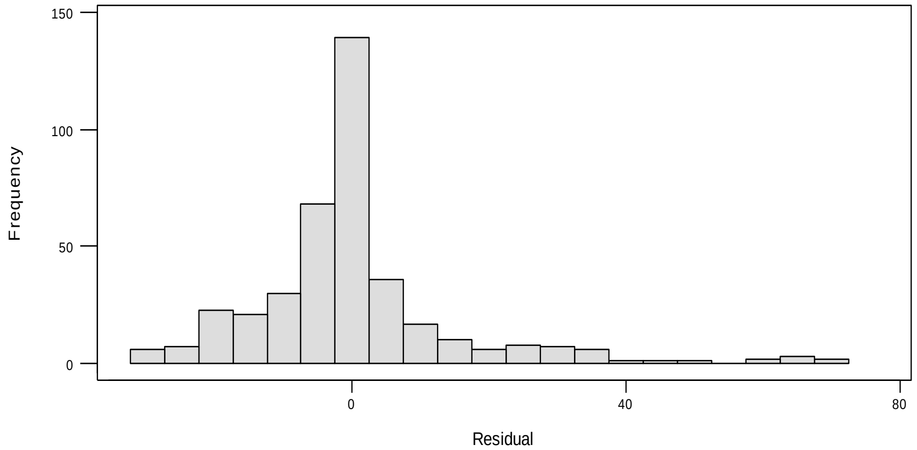

Figure 9- Histogram of the residuals distribution.

As appeared from the analysis, the error term of the fitted model is normal i.e. "bell shaped" in which the center is nearly at zero. Normal plot of the residuals, which is an another way to test the normality shows that all the points laid on and around the line and that theses residuals are normally distributed figure 10. The best and the most adequate ARIMA model can be used now for forecasting purposes. This model was used for forecasting the monthly rainfall values for the period 2013-2016, figure 11. The forecasted values show similar pattern of the original data series

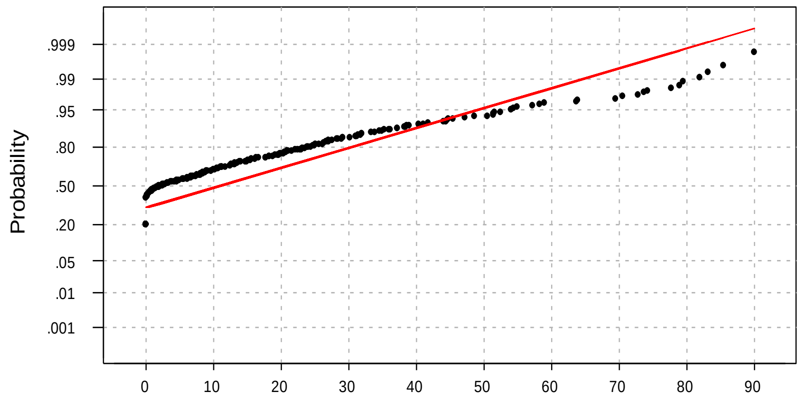

Figure 10- Normality diagnostic plot of the residuals.

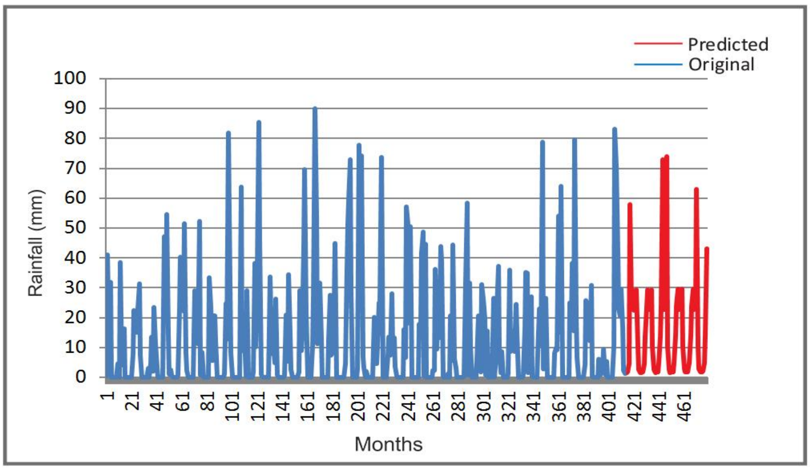

Figure 11- Original Baghdad series and the forecasted values for the period 2013-2016.

Cnoclusions:

Rainfall pattern of Baghdad area shows significant changes over time especially in the last years. Seasonality has significant effects on the general pattern of the rainfall series. SARIMA(2,1,3)x(0,1,1) is found to be good and reliable for forecasting purpose in which slight increasing trend can also be noticed .

References:

1. Suhartono, **2011**, Time series forecasting by using seasonal Autoregressive Integrated Moving Average: Subset, multiplicative or additive model, *J. of Mathematics and Statistics*, 7(1), pp: 20-27.
2. Iraqi Meteo. Organ. **2012**. *Records of meteorological data for the period* 1980-2012
3. Bisgaard, M. and Kulachi, M.**2011**,*Time series analysis and forecasting* by examples. Wiley and Sons Inc..
4. Box,G.E .and Jenkins,G.M.,**1976**,*The time series analysis, forecasting and control*, 3rd edition, Prentice-Hall
5. Abdul-Aziz, A.R., Anoky,M., Kwarn,H., Munyukazi,L. and Nuameh,N.,**2013**, Modeling and forecasting rainfall pattern in Ghana as a seasonal arima process. The case of Ashanti region. International *J. of Humanities and Social Sceinces*.
6. Suhartono, **2011**, Time series forecasting by using seasonal Autoregressive Integrated Moving Average: Subset, multiplicative or additive model, *J. of Mathematics and Statistics,* 7(1), pp: 20-27
7. Mahsin, M. ,Akhter,Y., Begum,M.,**2012**, Modeling rainfall in Dhaka division of Bangladesh using time series analysis, *J. of Mathematical modeling and Applications*, 1(5), pp: 67-73.