IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT)
e-ISSN: 2319-2402,p- ISSN: 2319-2399.Volume 10, Issue 6 Ver. II (Jun. 2016), PP 22-29
www.iosrjournals.org

Modelling Monthly Precipitation in Faridpur Region of Bangladesh Using ARIMA

Md. Matiur Rahman Molla[1], Md. Shohel Rana[2] Dr. M. Sazzad Hossain[3], S.M. Nuruzzman[4]

Abstract: *ARIMA modelling has become a key tool of analysing climatic variables. To forecast rainfall of a particular region adequately time series ARIMA model may help the farmer as well as water management authority indispensably. For our study purpose, we take data from Bangladesh Meteorological Department (BMD) of Faridpur region for the period January1983 to December 2014. From our analysis* ***ARIMA(0,0,0)(2,0,2)12*** *model was found that's why this developed model were used for forecasting monthly Precipitation for next three years to take proper decision of water development management authority. The Akike information criteria, MSE, MAD, and MAPE were applied to test the accuracy and applicability of the developed ARIMA model at different stages*

Keywords: *Time series analysis, monthly Precipitation forecasting, Box-Jenkins ARIMA methodology and accuracy.*

I. Introduction

Water resources are talks of the world topics now which mostly depend on rainfall. Faridpur is a district of Bangladesh which is famous for cultivating Jute, Onion, Garlic, wheat and so on. The production of these cash crops and food grains mostly depend on rainfall. Amid invention of modern agricultural technology there is no perfect alternative like rainfall water for cultivating of crops. For predicting such kind of phenomena, various techniques including numerical and machine learning process have been adopted based on historical time series. For analysing time series data, the most statistical process is Box-Jenkins methodology which was introduced by Box-and Jenkins (1976). This approach is well known by naming Auto Regressive Integrated Moving Average (ARIMA). And it has been widely applied for the prediction of various meteorological variables worldwide. If the underlying selected variables exhibit a seasonal behaviour then this ARIMA model should be expanded to include this component and then called seasonal ARIMA, SARIMA (Sawsan,et al,2013) . Like other region of Bangladesh Fridpur region facing shortage of water for a decades.

Although, Rainfall forecast of Dhaka, Sylhet and northern region of Bangladesh has been done by Mahsin et al. (2012) and Sourav et.al (2015) but forecasting of Faridpur region hasn't been done yet. Inspite of producing vast quantity of Jute of the country, the region's specific upcoming rainfall pattern is not investigated as a statistical manner. Sowing of Jute mostly depends on rainfall in April- MAY month For finding an alternative source of water rain water would be the best source since this region also facing contamination of arsenicosis in water and layer of the water going down to down. For accomplishing current study, data of rainfall of Jessore station for the period 1983-2014 were used to explore the rainfall behaviour applying ARIMA technique.

II. Basic Terminologies

The following key words are used throughout the research approach

Stationary:

Stationarity means that there is no growth or decline in the data. The data must be horizontal along the axis. A time series is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time is computed**.**

Suppose y_t be a stochastic time series then,

$$E\left(y_t\right)=\mu$$

$$\operatorname{var}\left(y_t\right)=E\left(y_t-\mu\right)^2=\sigma^2$$

Box-Jenkin's methodology and ARIMA modelling

The general ARIMA model proposed by Box and Jenkins (1970) is written as ARIMA (p, d, q) but when the characteristic of the data is seasonal behaviour then it said to be SARIMA. And the seasonal ARIMA model is written as very formal notation like this

$ARIMA(p,d,q) \times (P,D,Q)_s$

$$\begin{pmatrix}\text{Non seasonal}\\ \text{part of the}\\ \text{model}\end{pmatrix}\begin{pmatrix}\text{Seasonal}\\ \text{part}\\ \text{of the model}\end{pmatrix}$$

AR: p = order of the autoregressive part
I: d = degree of differencing involved
MA: q = order of the moving average part
s= number periods per season

The basis of the Box-Jenkins modelling in time series analysis is summarized the following figure and consist of three phases: identification, estimation and testing, and application.

The equation for the simplest ARIMA(p,d,q) model is as follows :

$$Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots\cdots\cdots + \emptyset_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \ldots\ldots\ldots - \theta_q e_{t-q}$$

The above equation can be expressed as backshift notation

$\left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) Y = C + (1 - \theta_1 B \text{-} \theta_2 B^2 - \cdots \theta_q B^q) e_t$

Where C= constant term. $\emptyset_i = i^{th}$ autoregressive parameter, $\theta_j = j^{th}$ moving average parameter., e_t =error term at time t and $B^k = k^{th}$order backshift operator.

When a time series exhibit a potentials seasonal indexed by s, using a multiplied seasonal ARIMA (p, d,q)(P,D,Q)s model is advantageous. The seasonal time series is transformed into a stationary time series with periodic trend components. A multiplied seasonal ARIMA model can be expressed as (Lee and Ko, 2011):

$\left(1 - \varphi_1 B - \varphi_2 B^2 - \cdots \varphi_p B^p\right)(1\text{-}\emptyset_1 B^s - \cdots \emptyset_p B^{ps})(1 - B^{\,s})^D Y_t = \left(1 - \theta_1 B - \cdots \theta_q B^q\right)\left(1 - \phi_1 B - \cdots - \right.$ $\phi QBQSat$or,

$$\varphi(B)\emptyset(B^s)(1 - B^s)^D Y_t = \theta(B)\phi(B^s) a_t$$

Where D is the order of the seasonal differencing, $\phi(B^S)$ and $\phi(B^S)$are the seasonal AR (p) and MA(q) operators respectively, which are defined as:

$\emptyset(B^s)$=$1 - \emptyset_1 B^s - \cdots \emptyset_p B^{ps})$

$\phi(B^S)$= $\left(1 - \phi_1 B - \cdots - \phi_Q B^{QS}\right)$

Where $\emptyset_1, \emptyset_2, \cdots, \emptyset_p$ are the seasonal AR (P) parameters and $\phi_1, \phi_2, \cdots, \phi_q$ are the seasonal MA (q) parameters.

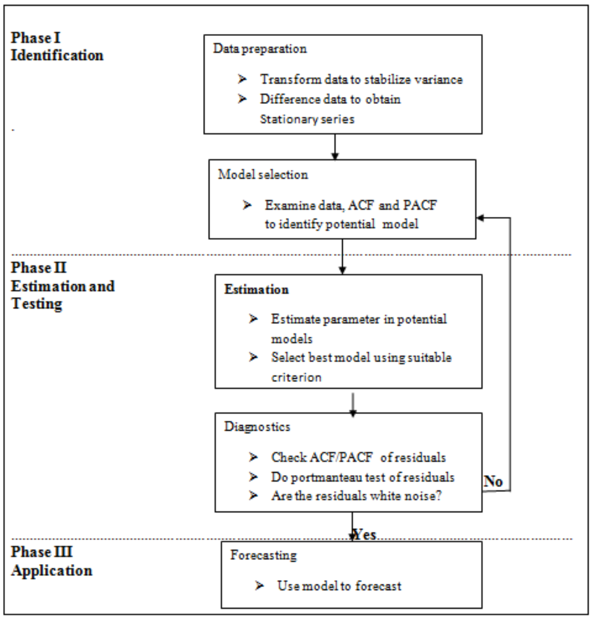


Figure1: Schematic representation of the **Box-Jenkins methodology** for time series modelling

III. Result and Discussion

(i) **Model Identification**:

As the data is a monthly rainfall, **Fig-2** shows that there is a seasonal cycle of the series and the series is nonstationary. The time series plot shows that there is a vivid indication of seasonality with periodicity of one year (twelve month) in the data set. And the both function of ACF and PACF of the original data; illustrated by **Fig-3**, show that the rainfall data is not stationary. In order to fit an ARIMA model stationary data in both variance and mean are needed. Stationary could be attained in the variance by log transformation and differencing of the original data to attain stationary in the mean. Since data series contain zero values straight forward **log transformation not possible**. As shown by Fig.5: Time series plot, Fig.6: the ACF and PACF for the seasonally adjusted series are almost stable which support the assumption that the series is stationary. Therefore, an $ARIMA(0,0,0) \times (P,0,Q)$ model could be identified. for the de-seasonalised rainfall data. After ARIMA model was identified above, the p, q and P and Q parameters need to be identified for our model.

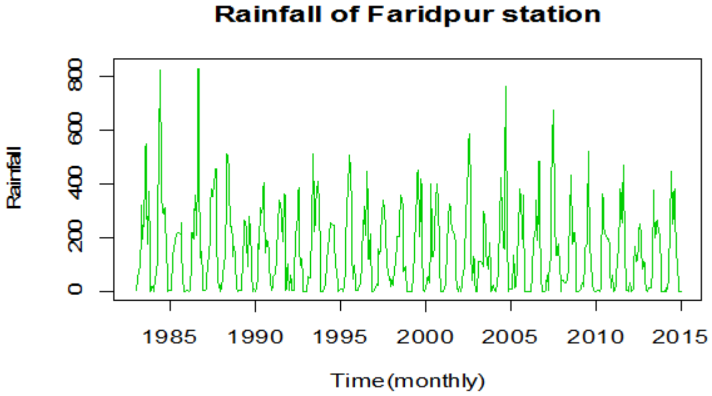


Figure: 2: Time series plot of original monthly rainfall data for Faridpur station (1983-2014)

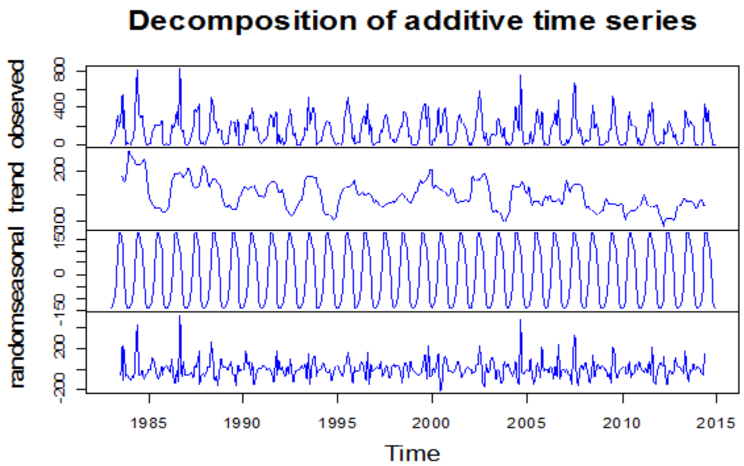
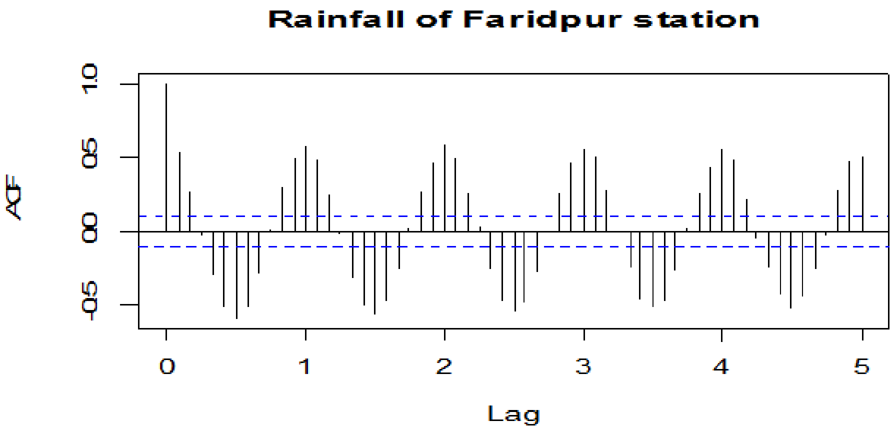


Figure: 3 Decomposition rainfall series of Faridpur regional station



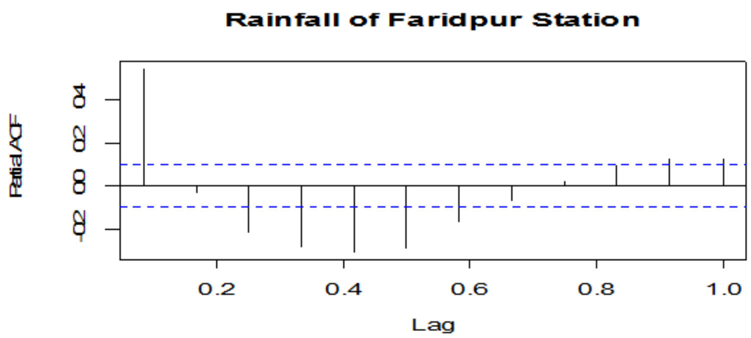


Figure: 4 Time series plot of ACF (top) and PACF (bottom) of original rainfall data.

From the above decomposition & time series plot, ACF and PACF it is observed that the original rainfall data is non-stationary and seasonal behaviour

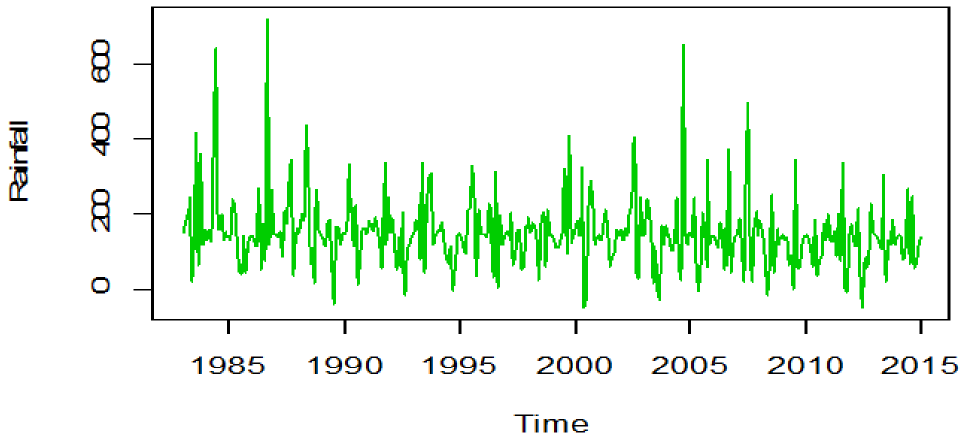


Figure 5: Time series Plot: after removing seasonal effect for Faridpur station.

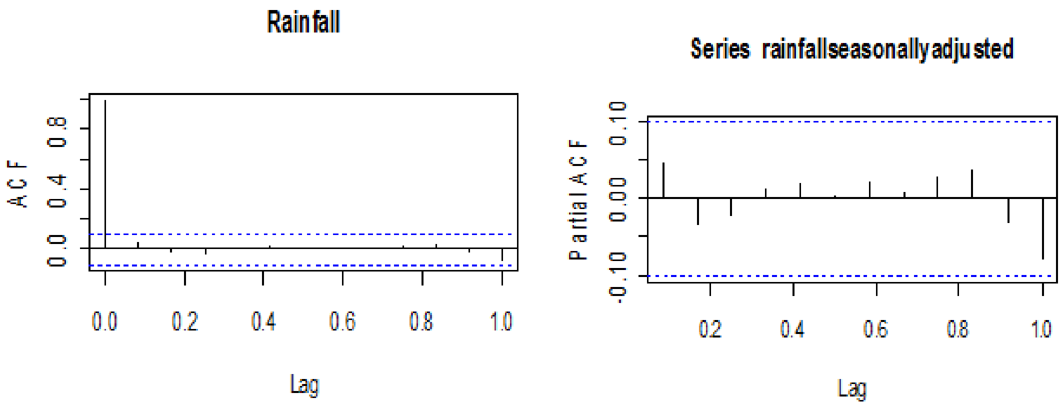


Figure 6: ACF and PACF curve after removing seasonal effect of original monthly rainfall data

Now, it looks like stationary.

(ii) Model Estimation

Since the order P, p and Q, q necessary to adequate model for a given problem is not known to us. It is required to determine the model that the best fits the data based on observing the ACF and the PACF of the difference data and on the basis of Alike information Criteria, BIC, BICc and log likelihood Among the combinations of order, the best order is chosen Model (0,0,0)(2,0,2)[12]

(iii) Model Diagnostic Checking

Once the models have been fitted to the data, a number of diagnostic checks were initialized. If the model fits well, the residuals should be uncorrelated with constant variance. Moreover, in developing model this is often assumed that the errors are normally distributed. Hence, we expect the residuals to be more or less normally distributed. Standard checks for ARIMA is to compute the ACF and the PACF of the residuals.

Further diagnostics checking can be done by looking at the residuals in various ways (Figure: 08).If the residuals are normally distributed, they should all more or less lie on a straight upward sloping line (Bowerman, and O'Connell, 1993). After accomplishes the above series checking the ARIMA (0, 0, 0) (2, 0, 2)12 was found to give significant results.

Figure: 9 shows the graphical plot for the actual rainfall series versus the predicted rainfall series and from the visual inspection of the plot it is sufficient evident that the chosen model is well enough since the predicted series is close to the observed series that's why this model could be used to forecast the next three years rainfall in Faridpur region, Bangladesh.

Figure 11: Comparison plot of original value and fitted value Using ARIMA (0,0,0) (2,0,2) [12] From January 1983 to December 2014 for Faridpur station. where Green colour indicates Original observations and Blue indicates Fitted observations

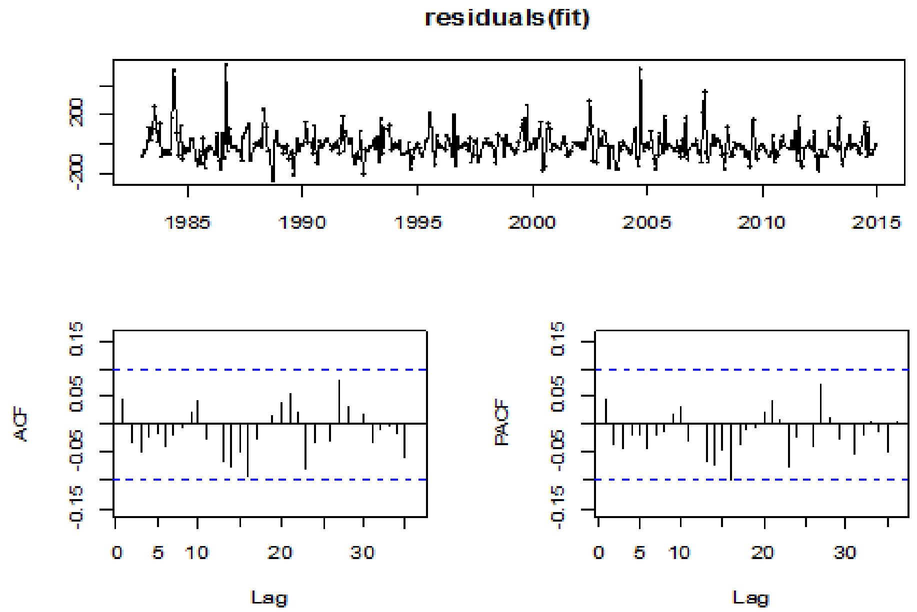


Figure 7 Plot: ACF, PACF and Fitted curve for residuals

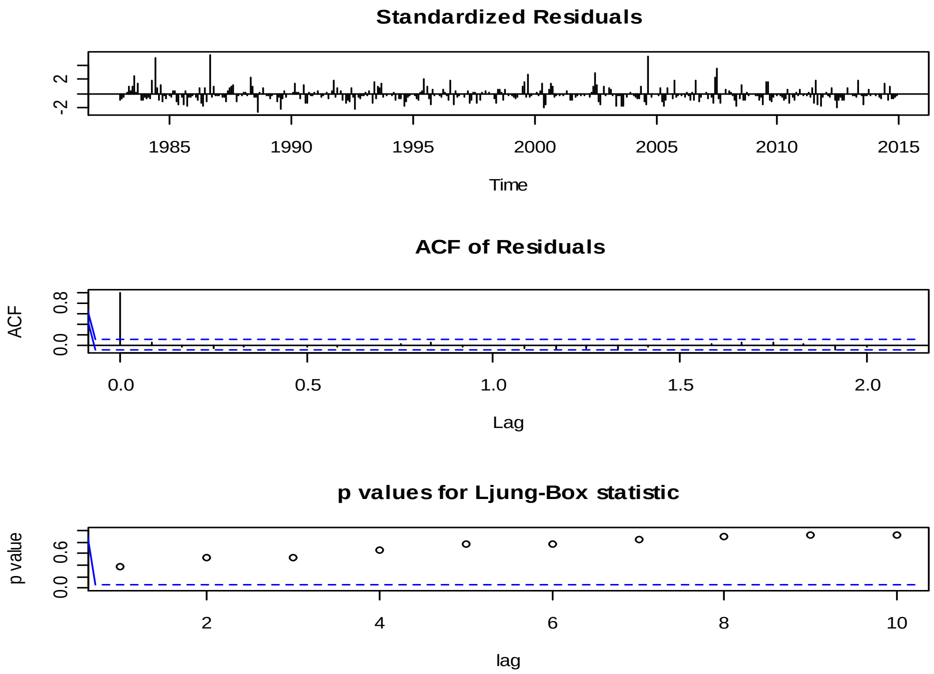


Figure 8: Plot of standardized residuals, ACF of residuals & Ljung-Box statistic

Box –Ljung test
Data: residuals
X-squared =4.7519, df=12, p-value =0.9658
Null hypothesis is rejected for residual of the fitted model. So, the residuals follow white-noise and hence resudual stationary

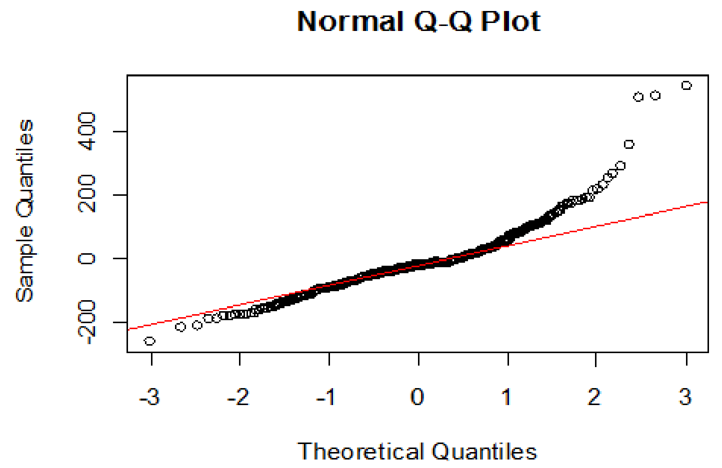


Figure 9 Plot: Normal Q-Q plot for Normality checking for Residuals

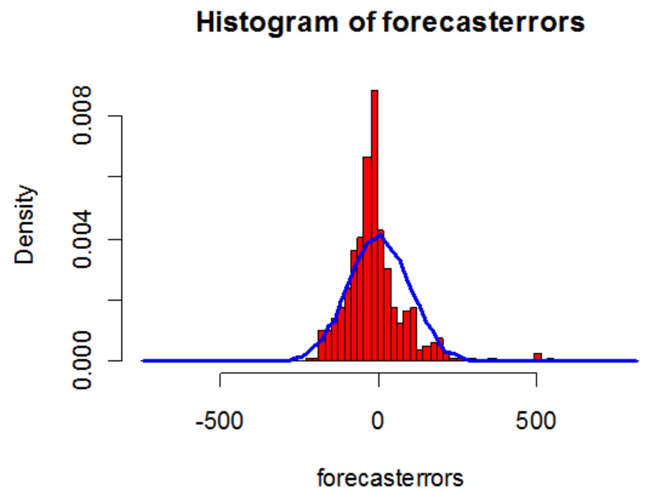


Figure10: Histogram of forecast errors

Comment: Observing above Q-Q plot and histogram of forecast error, we may decide that the model is well fitted.

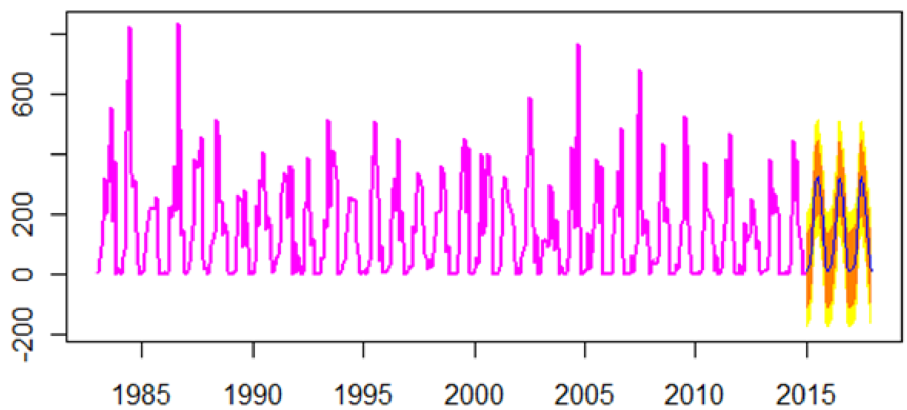


Figure 11 Forecasted plot for next three years of Faridpur Station Using ARIMA (0,0,0)(2,0,2)[12] model where, Red=Original, black= point estimate, yellow=80% CI, pink= 95%.

Inspection of the time plot of the standardized residuals in Figure 7 & 8 shows adequate supporting for the model. The ACF of the standardized residuals shows no apparent departure from the model assumptions. The Ljung-Box test for this model gives a chi-squared value of 4.7519with 12 degrees of freedom, leading to a *p*- value of 0.9658—a further indication that the model is quite well fitted.The normal Q-Q plot of the residuals shows nearly close to the theoretical counterpart and normality assumption is met by the histogram of the forecasts error.

Table 1: Fitted Coefficient for Rainfall Data of Faridpur station Using ARIMA (0, 0, 0) (2, 0, 2) [12] with non zero mean

	Sar1	Sar2	Sma1	Sma2	intercept
Coefficient	0.4174	0.5801	-0.4246	-0.4645	148.9329
S.E	0.3049	0.3046	0.3209	0.3054	35.4495

Estimated sigma^2 =9638, AICc = 4655.06, BIC = 4678.5

Table 2: Forecasted amount of rainfall (& CI) in Faridpur station, Bangladesh during the period July 2015- June 2017.

Time	Point Forecast	LCL at 95% CI	UCL at 95% CI
January-15	11.38210	-181.12520	203.8894
February-15	23.67352	-168.83378	216.1808
March-15	37.71771	-154.78959	230.2250
April-15	91.52520	-100.98210	284.0325
May-15	213.32691	20.81960	405.8342
June-15	302.93517	-110.42786	495.4422
July-15	323.77927	131.27196	516.2866
August-15	280.13232	87.62501	472.6396
September-15	239.25799	46.75068	413.7653
October-15	155.30773	-37.19957	347.8150
November-15	31.08011	-161.42719	223.5874
December-15	11.61880	-180.88850	204.1261
January-16	11.68474	-180.81288	204.1824
February-16	22.80280	-169.69483	215.3004
March-16	33.95305	-158.54458	226.4507
April-16	88.01403	-104.48359	280.5117
May-16	197.72106	5.223426	390.2187
June-16	314.62890	122.13126	507.1265
July-16	318.93251	126.43487	511.4301
August-16	286.99297	94.49534	479.4906
September-16	231.00304	38.50540	423.5007
October-16	144.38469	-48.11294	336.8823
November-16	32.13004	-160.36758	224.6277
December-16	11.59953	-180.89809	204.0972
January-17	11.85054	-181.87142	205.5725
February-17	23.62164	-170.10032	217.3436
March-17	36.42300	-157.29895	230.1450
April-17	90.20247	-103.51948	283.9244
May-17	206.53307	12.93111	400.3750
June-17	307.43295	113.71098	501.1549
July-17	321.32142	127.59946	515.0434

Our key task was to predict the monthly average rainfall for next three years. Depending on the developed model **Table: 2** illustrate the forecasted average monthly rainfall along with 80% and 95% confidence limit for three years.

IV. Conclusion

In this study, our main task was to develop a ARIMA model to predict the next three years rainfall of Faridpur region, Bangladesh. Ultimately, on the basis of nature of the data, it has been developed an ARIMA model. Comparing the observed and forecasted value with 95% confidence interval limit, the proposed model gives an adequate result. Model diagnostic assessment presented that ARIMA (0,0,0)(2,0,2)12 should have significant result. The predicted model is reliable as the RMSE values on test data are comparatively less. And this developed model obviously may help the decision makers to set up strategies and proper management of water resources.

Reference

[1]. Ali. S.M., (2013) Analysis of Baghdad rainfall using ARIMA method .Iraki Journal of Science, Vol. 54, No. 4, pp: 1136-1142

[2]. Anderson, T.W. (1994) The statistical Analysis of Time Series, Willey India pvt.Ltd, New Delhi, India.

[3]. AKAIKE, h. (1974) A new look at statistical model identification, *IEEE transaction and automatic control*, AC-19, 716-723.

[4]. AWAL, M.A. and M. A.B. Siddique (2011). Rice Production in Bangladesh Employing by ARIMA Model, *Bangladesh Agricultural Journal*.

[5]. Bari S. H.,Rahman M.T.Hossain M.M. and Ray S. (2015).Forecasting monthly precipitation in Sylhet city using ARIMA model. Civil And Environmental Research Vol. 7, No. 1, 2015, ISSN 2224-5790 (Paper) ISSN 2225-0514(online)

[6]. BROCKWELL, P., J. and R.A .DAVIS (1996) An introduction to time series and forecasting. New York: Springer-Verlag.

[7]. Box, G.E.P., G.M. Jenkins, and G.C. Reinsell (1994) Time series analysis: Forecasting and control, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall.

[8]. BOX, G.E.P and D.A. PIERCE (1970) Distribution of the residual autocorrelation in autoregressive-integrated moving average time series models, *Journal of theAmerican Statistical Association,* 65, 1509-1526.

[9]. Chattfield, C. and M. YAR (1988) Holt-Winters' forecasting: some practical issues, *The Statistician*, 37, 129-140.

[10]. Dougherty, C.,(2011), Introduction to Econometrics, Oxford University Press Inc., Newyork.

[11]. DICKEY, D.A. and W.A. FULLER (1979) Distribution of the estimators for autoregressive time series with a unit root, Journal of the *American Statistical Association*, 74, 427-431.

[12]. Faisal, F. (2012) Time Series ARIMA Forecasting of Natural Gas Consumption in Bangladesh's power sector *Elixir International Journal*, Elixir Prod. Mgt. 49(2012) 9985-9990.

[13]. Gujarati, D.N. (2014), *Basic Econometrics*, Tata McGraw-Hill Edition, New Delhi, India.

[14]. Hyndman, R.J., (2014) Forecasting principles and practice. robjhyndman.com/uwa

[15]. LEPOJEVIC, V. and M. A. PESIC (2011) Forecating Electricity Consumptionby Using Holts'-Winters and Seasonal Regression Model, University Servia Economic Organization, Vol.8 pp. 421-431.

[16]. Makridakis, S., and S.C. Wheelwright and R. J. Hyndman ( ) Forecastin Method and applications, Willey and Sons (Asia).

[17]. Mhsin M., Akhter Y. And Begum M. (2012) Modelling rainfall in Dhaka division of Bangladesh using time seres analysis. Journal of Mathematical Moeliing and Application, Vol. 1, No.5, 67-73, ISBN: 2178-2423

[18]. PEGELS, C.C.(1969) Exponential forecasting: some new variations, *Management Science*, 12, No. 5, 311-315.

[19]. Zafor M.A, Chakraborty A, Muniruzzaman S.M. and Mojumder S.R. (2016). Rainfall Forecasting in northern part of Bangladesh. *Research Journal of Engineering Science*, Vol. 5(3), 17-31