Research Journal of Engineering Sciences E- ISSN 2278 – 9472
Vol. **5(3),** 17-31, March **(2016)** *Res. J. Engineering Sci.*

Rainfall Forecasting in Northeastern part of Bangladesh Using Time Series ARIMA Model

Md. Abu Zafor[1], Amit Chakraborty[1], Sheikh Md. Muniruzzaman[2] and Smriti Rani Mojumdar[1]
[1]Department of Civil Engineering, Leading University, Sylhet
[2]Department of English, Leading University, Sylhet
cezafor@gmail.com

Available online at: www.isca.in, www.isca.me
Received 29th February 2016, revised 3rd March 2016, accepted 25th March 2016

Abstract

To improve water resources management, time series analysis is an important tool. Bangladesh is a densely populated country and now facing shortage of drinking water. Rainwater harvesting is one of the major techniques to overcome this problem. For this purpose, it is very much important to forecast future rainfall events on a monthly basis. Box-Jenkins methodology has been used in this study to build Autoregressive Integrated Moving Average (ARIMA) models for monthly rainfall data from eight rainfall stations in the northeastern part of Bangladesh: (Sunamganj ,Lalakhal, Kanaighat, Sherupur, Chattak, Gobindoganj, Sheola, Zakiganj) for the period 2001-2012. Eight ARIMA models were developed forthe above mentioned stationsas follow: $(2,0,2)x(1,0,1)_{12}$, $(4,0,1)(1,0,0)_{12}$, $(1,0,1)(0,0,2)_{12}$, $(1,0,0)(2,0,0)_{12}$, $(1,0,0)(2,0,0)_{12}$, $(1,0,0)(2,0,0)_{12}$, $(1,0,0)(2,0,0)_{12}$, $(2,0,0)(2,0,0)_{12}$ respectively. The performance of the resulting successful ARIMA models were evaluated using the data year (2011). These models were used to forecast the monthly rainfall data for the up-coming years (2013 to 2020). The results supported previous work that had been carried out on the same area recommending the use of water harvesting in both drinking and agricultural practices.

Keywords: Time series analysis, Monthly Rainfall forecasting, Box-Jenkins (ARIMA) methodology

Introduction

Winstanley[1] reported that monsoon rains from Africa to India decreased by over five hundredth from 1957 to 1970 and expected that the long run monsoon seasonal rain, averaged over five to ten years is probably going to decrease to a minimum around 2030.

Laban[2] uses time series supported ARIMA and Spectral Analysis of areal annual rain of two same regions in East Africa and counseled ARMA(3,1) because the best appropriate region indice of relative wetness/dryness and dominant quasi-periodic fluctuation around 2.2-2.8 years,3-3.7 years,5-6 years and 10-13 years.

Kuo and Sun[3] used an intervention model for average10 days stream flow forecast and synthesis that was investigated by to influence the extraordinary phenomena caused by typhoons and alternative serious abnormalities of the weather of the Tanshui river basin in Taiwan.

Chiew et al[4] conducted a comparison of six rainfall-runoff modeling approach to reproduce daily, monthly and annual flows in eight unfettered catchments. They accomplished that time-series approach will offer adequate estimates of monthly and annual yields within the water resources of the catchments.
Langu[5] used statistic analysis to observe changes in rainfall and runoff patterns to go looking for important changes within the parts of variety y of rainfall statistic.

Box and Jenkins in early 1970's, pioneered in evolving methodologies for statistic modeling within the univariate case often referred to as Univariate Box-Jenkins (UBJ) ARIMA modeling [6].

Methodology

Study Area Description and Datasets: Sylhet district is the north eastern part of Bangladesh. It is sited on the bank of the river Surma. The total area of the district is 3452.07 sq. km. (1332.00 sq. miles). The district lies between 24º36' and $25^0$11' north latitudes and between 91º38' and $92^0$30' east longitudes.

Data Acquisition: The time series data of rainfall for the Sylhet region was collected from Bangladesh Water Development Board (BWDB). Observed monthly rainfall of eight monitoring weather stations for the period 2001 to 2012 was collected to perform this study.

ARIMA Model Methodology: In 1976, Box and Jenkins, give a methodology (Figure-2) in time series analysis to find the best fit of time series to past values in order to make future forecasts[6]. The methodology consists of four steps: i. Model identification. ii. Parameters estimation. iii. Diagnostic checking for the recognized model suitability for modeling and iv.Utilize the model for forecasting.

Table-1
Name of the Rainfall Station and Geographic Location

Name of Rainfall Station	Location of Station	Latitude	Longitude
R-127 Snamganj	BWDB office campus, Sulaghar, Snamganj	25.07	91.41
R-116 Lalakhal	Lalakhal tea garden, Lalakhal, Jaintia,Sylhet	25.11	92.16
R-228Kanaighat	Kanaighat, Sylhet	25.01	92.26
R-109 Gobindganj	Gobindganj, Syhet	24.93	91.68
R-107 Chhatak	Chhatak,Sunamganj	25.03	91.67
R-119 Sherpur	Sherpur, Syhet	24.63	91.68
R-125Sheola	BWDB office campus, Sheola, Sylhet	24.82	92.16
R-130 Zakiganj	BWDB office campus, Zakiganj,Sylhet	24.88	92.37

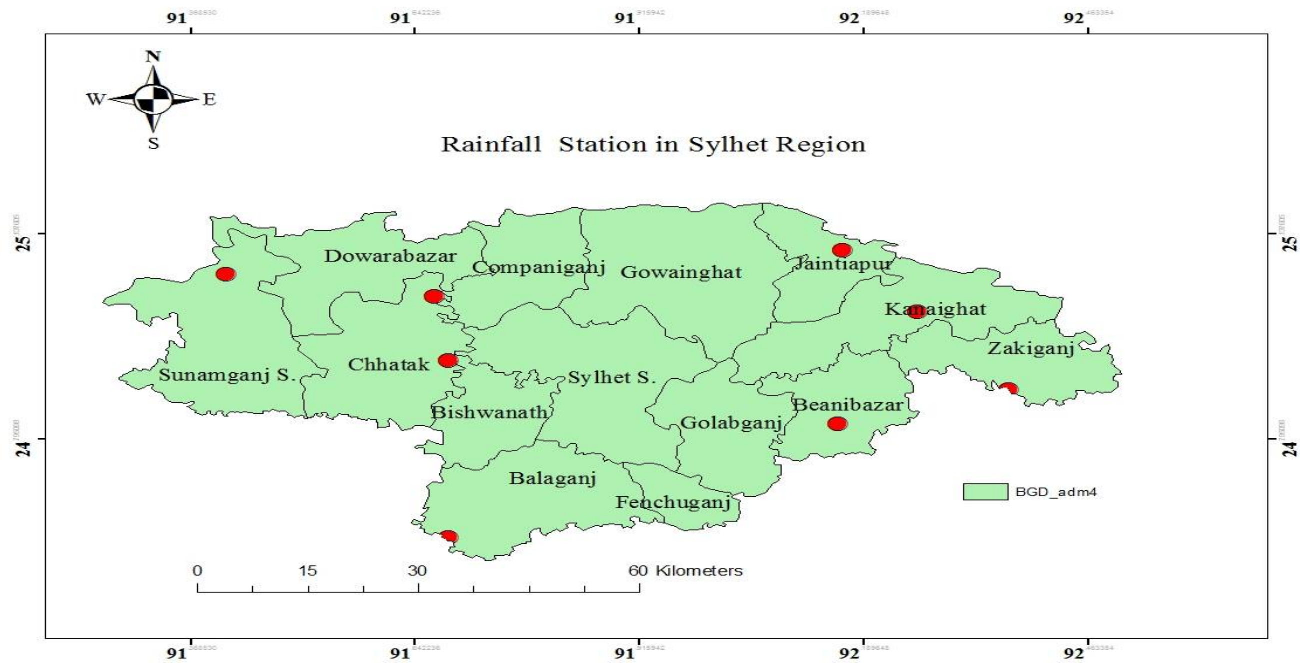

Figure-1
Study Area of Northeastern Part of Bangladesh

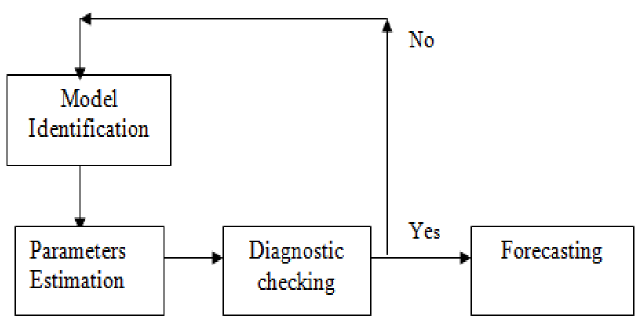

Figure-2
Box-Jenkins Methodology

ARIMA Process: The first conditions imply that the series Yt following (2) is stationary. In practice Yt might be non stationary, but with stationary first difference

$$Y_t - Y_{t-1} = (1-B)Y_t \quad (1)$$

If (1-B) Yt is no stationary, then need to take the second difference,

$$Y_t - 2Y_{t-1} + Y_{t-2} = (1-B)[(1-B)Y_t] = (1-B)^2 Y_t \quad (2)$$

In general, it may need to take the d^{th} difference (1-B) dY_t (although rarely is d larger than 2). Substituting (1-B) dY_t for Y_t in (7) yields the ARIMA (p,d,q) model (Bell, 1984) [8]:

$$\left(1-\varphi_1 B - \cdots - \varphi_p B^p\right)(1-B)^d Y_t = (1-\theta_1 B - \cdots - \theta_q B^q)a_t \quad (3)$$

Or $\varphi(B)(1-B)^D Y_t = \theta(B)a_t$

Where; d is the order of differencing.

When a time series exhibits potential seasonality 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 non-periodic trend components. A multiplied seasonal ARIMA model can be expressed as (Lee and Ko, 2011)[9]:

$$\left(1-\varphi_1 B - \cdots - \varphi_p B^p\right)\left(1-\Phi_1 B^s - \cdots - \Phi_p B^{ps}\right)(1-B^s)^D Y_t = \left(1-\theta_1 B - \cdots - \theta_q B^q\right)\left(1-\Theta_1 B - \cdots - \Theta_Q B^{Qs}\right)a_t \quad (4)$$

Or $\varphi(B)\Phi(B^S)(1-B^s)^D Y_t = \theta(B)\Theta(B^s)a_t$

Where D is the order of seasonal differencing, Φ(Bs) and Θ(Bs) are the seasonal AR(p) and MA(q) operators respectively, which are defined as:

$$\Phi(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_p B^{ps} \quad (5)$$

$$\Theta(B^s) = \left(1 - \Theta_1 B - \cdots - \Theta_Q B^{Qs}\right) \quad (6)$$

Where; Φ1,…, Φp are the seasonal AR(p) parameters and Θ1,…, Θp are the seasonal MA (q) parameters.

To illustrate forecasting with ARIMA models, we shall use (2.9) written as:

$$Y_{t+1} = \Phi_1 Y_{n+l-1} + \cdots + \Phi_{p+d} Y_{n+l-p-d} + a_{n+l} - \theta_1 a_{n+l-1} - \cdots - \theta_q a_{n+l-q} \quad (7)$$

for t = n + l.

Assuming for now that the data set is long enough so that we may effectively assume it extends into the infinite past.

Seasonal ARIMA Model

A series which is governed by a mixed seasonal process can be expressed as

$$\varphi(L^s)\Delta_s^D \Delta^d y_t = \Theta(B^s)e_t \quad (8)$$

Where

$$\varphi(L)y_t = C + \theta(L)e_t \quad (9)$$

$$\Theta(L^s) = 1 - \Theta_1 L^2 - \Theta_2 L^{2s} - \cdots - \Theta_Q L^{Qs} \quad (10)$$

And Δ_s^D and Δ_d are the difference operators to induce stationary. We can denote the general mixed seasonal model as ARIMA (P, D,Q)s.

Where, P = Order of the seasonal autoregressive process. D = Number of seasonal differencing. Q = Order of the seasonal moving average. S = Number of the seasonality.

Model Selection Criteria: This is very difficult question to select the best algorithm .Real data do not follow any model. The general advice is that: firstly we have to choose what measure of forecast error is most appropriate for the particular situation in hand. We may use root mean squared error (RMSE), mean absolute error (MAE), one step error, 12-step error etc. The mathematical formula of RMSE and MAE are as follows:

$$RMSE = \sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(Y_t - \overline{Y}_t)^2\right]} \quad (3.19)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|Y_t - \overline{Y}_t| \quad (11)$$

There is some other statistic for model selection. Which are like Akike information Criteria (AIC), Bayesian Information Criterion (BIC) (Schwarz, 1978) and Schwartz's Criterion (SC), which is closely related to AIC[10]. Then the model that gives smallest value of these criteria. Also there are some contradictions among econometricians about which criteria perform better. However, in this thesis we use all of these model selection criteria.

Results and Discussion

Sulaghar (Sunamganj) station was selected as a illustration of calculations for time series ARIMA model. The first step according to the time series methodology is to cheek whether the time series (monthly Rainfall) is stationary and has seasonality.

In Figure-3, the monthly rainfall data shows that there is a strong seasonality present in data set and it is not stationary. The plots of ACF and PACF of the original data (Figure-4) also shows that the rainfall data is not stationary; where both ACF and PACF (Figure-4) have stationary time series has a constant mean and has no trend over time.

According to ARIMA methodology seasonal trend could be removed by having seasonal differencing (D) through subtracting the current observation from the previous to observation, as described before that rainfall in Sulaghar (Sunamganj) it almost extend for every twelve months. In general the seasonality in a time series is a usual pattern of changes that repeats over time periods (S).

If the differenced transformation is apply only once to a series, that means data has been "first differenced" (D=0). If a trend is present in the data, then non-seasonal (regular) differencing (d) is required. The monthly rainfall data, of Sulaghar (Sunamganj) station, required to have a first seasonal difference of the actual data in order to have stationary series. Then, the ACF and PACF for the differenced series should be tested to ensure the stationary. From all of the above, an ARIMA Model of (p, 0, q) x $(P, 0, Q)_{12}$ could be identified. In the Box-Jenkins methodology, the estimated model will be depending on the ACF and PACF (Fig. 5). After ARIMA model was recognized, the p, q, P and Q parameters need to be identified for Sulaghar (Sunamganj) monthly rainfall time series.

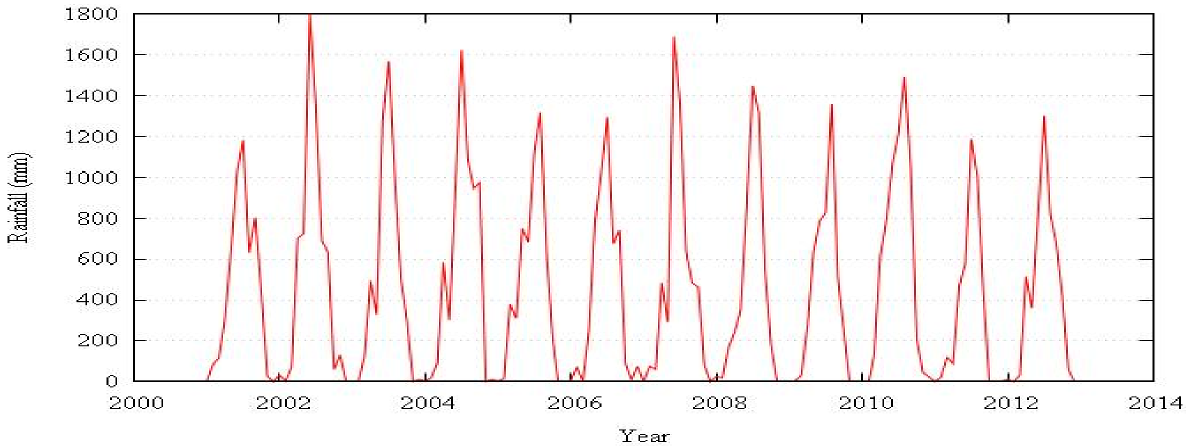

Figure-3
Monthly Rainfall Data of Sulaghar (Sunamganj) Station (2001-2011)

Table
Summary Statistics, using the observations 2001:01 – 2012:12

Mean	Median	Minimum	Maximum
451.222	299.000	0.00000	1800.00
Std. Dev.	C.V.	Skewness	Ex. kurtosis
479.226	1.06206	0.872856	-0.303962
5% Perc.	95% Perc.	IQ range	Missing obs.
0.00000	1357.00	753.500	0

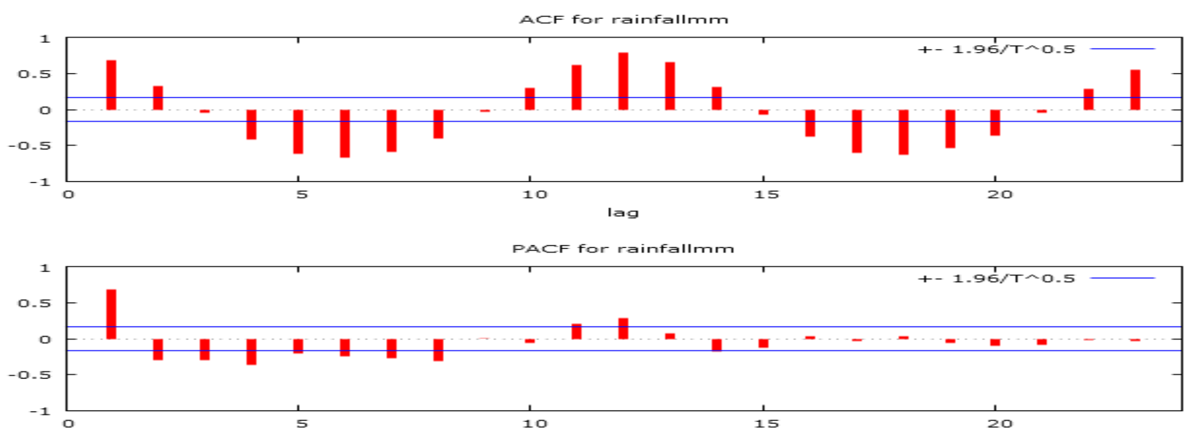

Figure-4
ACF and PACF for Original Monthly Rainfall Data of Sulaghar (Sunamganj)

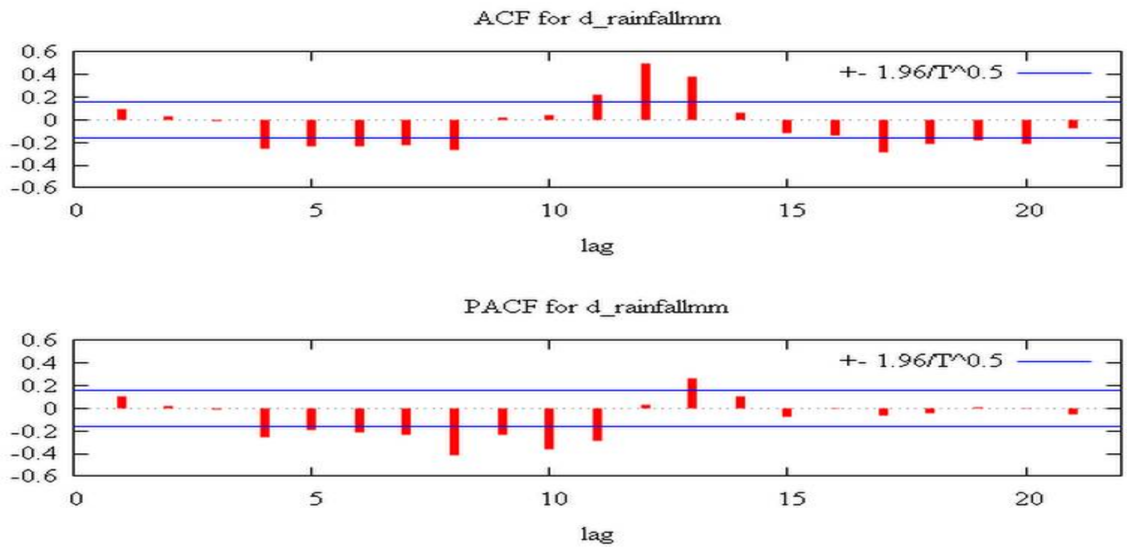

Figure-5
ACF and PACF after Taking First Difference in Rainfall Data of Sulaghar (Sunamganj)

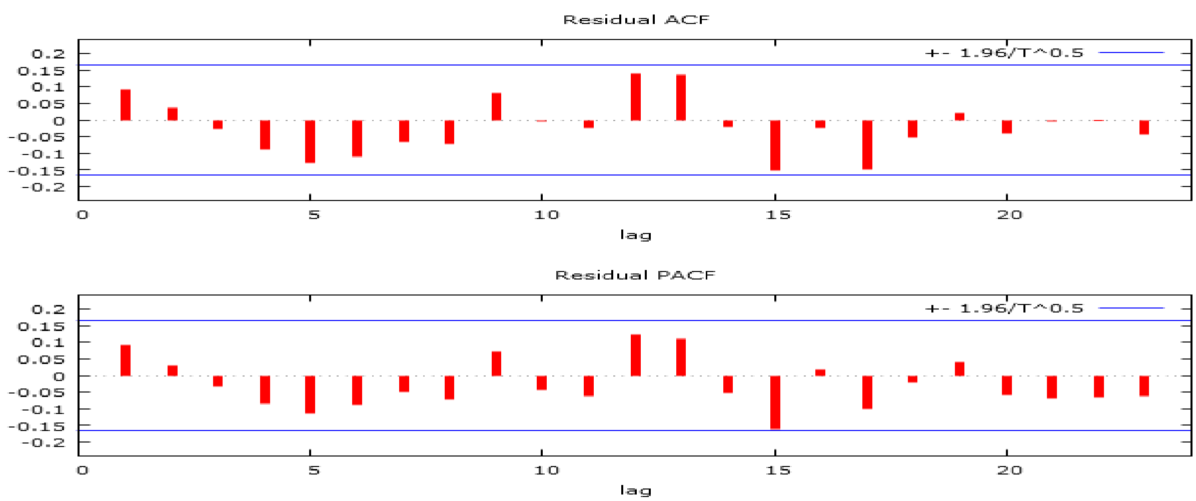

Figure-6
ACF and PACF Residuals of ARIMA Model (2,0,2)x(1,0,1)$_{12}$ of Sulaghar (Sunamganj)

For Sulaghar (Sunamganj), the ARIMA Model should be identified and tested. Then apply AIC to select the best ARIMA model. After selecting the most appropriate model it was found that ARIMA model (2,0,2)x(1,0,1)$_{12}$among several models that passed all statistic tests required in the Box-Jenkins methodology.

The model parameters are estimated using SAS10.1 software. Table 1 shows the estimated Parameters for the successful Sulaghar (Sunamganj) ARIMA model (2,0,2)x(1,0,1)$_{12}$

The residuals from the fitted model are examine for the sufficiency. This is done by testing the residual ACF and PACF plots that shows all the autocorrelation and partial autocorrelations of the residuals at different lags are within the 95% confidence limits.

The first calculation is the normal probability plot of the residuals (top-left of Figure-7) which is good as required for an adequate model and most of the residuals are on the straight line. The second measure for adequacy of model is the histogram of the residuals (bottom-left of Figure-7) which shows good normality of the residuals. The third measure is the plot of residuals against fitted values (top-right of Figure-7).

In this plot the data does not follow any symmetric pattern with the run order value. It shows almost random behavior of residuals with the increasing run order which indicates that the model is a good fit. Almost all of the residuals are within acceptable limits which indicate the adequacy of the recommended model.

The performance of the Sulaghar (Sunamganj) ARIMA model (2,0,2)x(1,0,1)$_{12}$ evaluated by forecasting the data for the year 2011. Both the forecasted and real weekly rainfall depth of the Sulaghar (Sunamganj) station for the year 2011 were fitted on the same plot to indicate the model adequacy, performance and comparison purposes (Figure-8). His similarity and matching between the forecasted and real rainfall depth were good. The above comparison increases confidence with the ARIMA(2,0,2)x(1,0,1)$_{12}$ to represent the rainfall data at Sulaghar (Sunamganj) station and can be used for forecasting the future rainfall data. Figure-9 shows the forecasting rainfall depth for the years 2013-2020 using ARIMA (2, 0, 2) x (1, 0, 1)$_{12}$

The same procedures of the Box-Jenkins methodology were followed for the other three stations (Lalakhal, Kanaighat, Sherupur, Chattak, Gobindoganj, Sheola, Zakiganj) to forecast future rainfall.

For all three stations, it was found that some ARIMA models passed all the statistical tests required in the Box-Jenkins methodology without significant residuals for ACF and PACF plots. These models are the following.

Table-2
Parameter Estimation for a ARIMA (2,0,2)x(1,0,1)$_{12}$ Model

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	1.58494	0.07369	21.5096	3.9E-45
AR1,2	1	2	-0.7238	0.08529	-8.4865	3.6E-14
AR2,12	2	12	0.99248	0.00463	214.541	8E-171
MA1,1	1	1	0.57112	0.1434	3.98261	0.00011
MA1,2	1	2	-0.0143			
MA2,12	2	12	0.79395			
Intercept	1	0	449.966	282.11	1.595	0.11309

Table-3
Model Selection Parameter

DF	133
Sum of Squared Errors	433009.3166
Variance Estimate	3255.709147
Standard Deviation	57.05882182
Akaike's 'A' Information Criterion	1563.053173
Schwarz's Bayesian Criterion	1583.64467
RSquare	0.960378467
RSquareAdj	0.95859103
MAPE	26.56722653
MAE	50.86241458
-2LogLikelihood	1549.053173

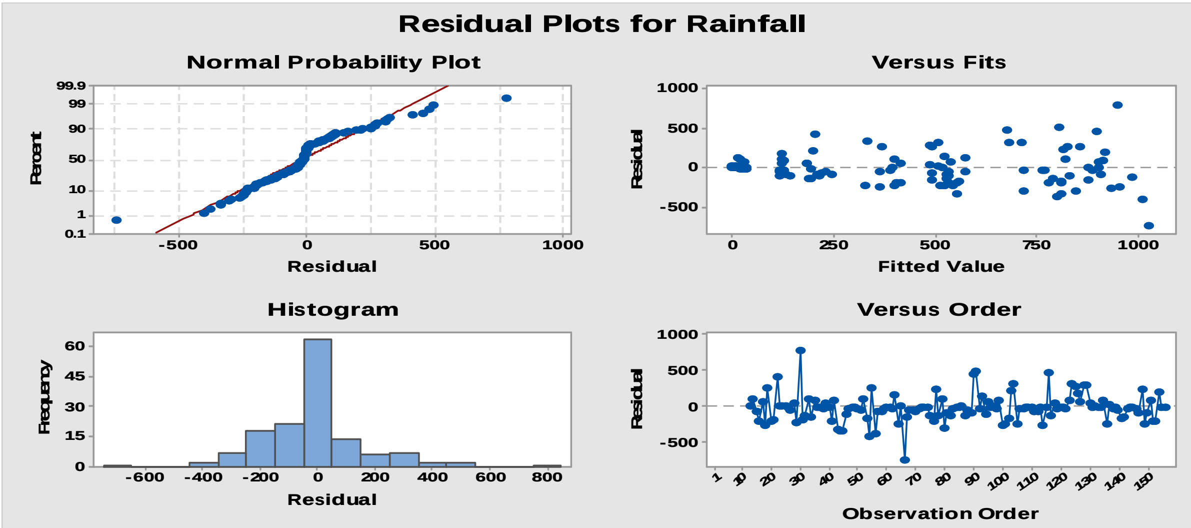

Figure-7
Residual Plots of Sulaghar (Sunamganj) ARIMA Model (2,0,2)x(1,0,1)$_{12}$

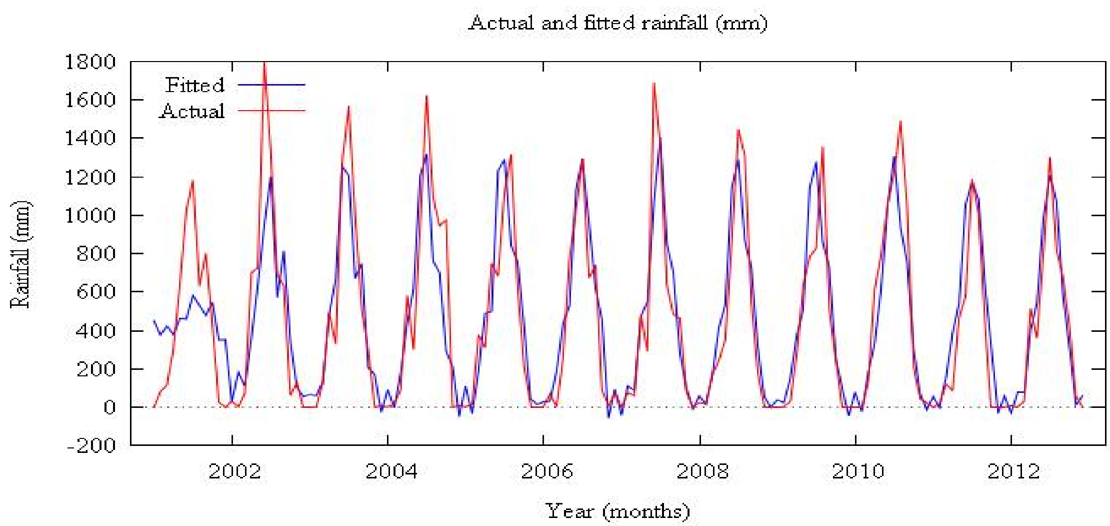

Figure-8
Actual vs. Fitted Values of the Rainfall Data for the Station Sulaghar (Sunamganj)

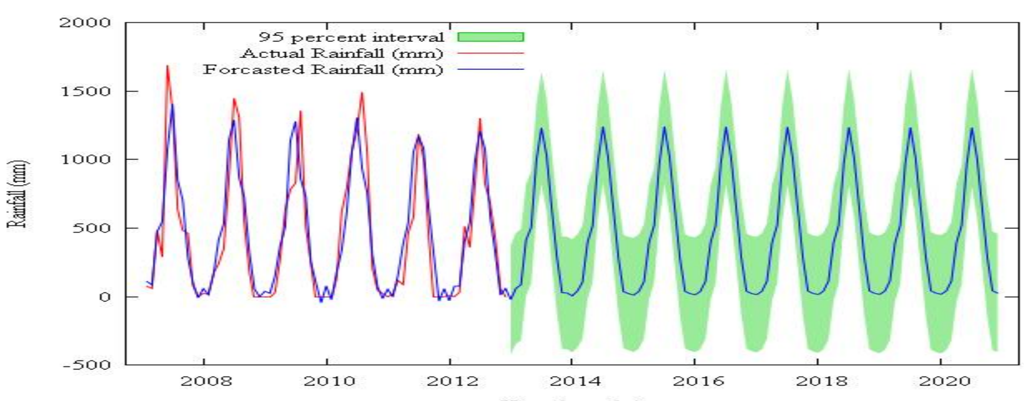

Figure-9
Actual and Forecasted Monthly Rainfall of Sulaghar (Sunamganj)

Table-4
Parameter Estimation for seven stations

Name of rainfall station	**ARIMA Model**	**Model Parameter**						
R-109 Gobindganj	(1,0,0)(2,0,0)12	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
		AR1,1	1	1	0.2984657	0.0015457	193.09819	9.58E-179
		AR1,2	1	2	0.1314892	0.0004493	292.67578	3.18E-205
		AR2,12	2	12	0.9999936	2.185E-06	457594.82	0
		MA1,1	1	1	-0.8094073	0.003818	-211.9991	1.1E-184
		MA1,2	1	2	0.1905893	0.0010197	186.91535	1.12E-176
		MA2,12	2	12	0.9944184	0.0015631	636.17191	9.51E-255
		Intercept	1	0	489.40086	157.81037	3.1011959	0.0023111
R-116 Lalakhal	(4,0,1)(1,0,0)12	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
		AR1,1	1	1	0.3866046	0.0776261	4.9803446	1.934E-06
		AR1,2	1	2	0.2136635	0.0766545	2.787358	0.0060926
		AR1,3	1	3	-0.0189977	0.0393815	-0.4824013	0.6303139
		AR1,4	1	4	-0.2121051	0.0791416	-2.6800698	0.0082915
		AR1,5	1	5	-0.353523	0.0778449	-4.5413783	1.239E-05
		MA1,1	1	1	-0.9999977	0.021685	-46.114625	1.111E-83
		MA2,12	2	12	0.1754065	0.1075021	1.6316568	0.1051174
		MA2,24	2	24	-0.1667476	0.095806	-1.740471	0.0840892
		Intercept	1	0	361.17484	17.646405	20.467332	6.308E-43
R-228Kanaighat.	(1,0,1)(0,0,2)12	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
		AR1,1	1	1	0.6088328	0.0709431	8.5819927	1.733E-14
		AR2,12	2	12	0.4834523	0.079123	6.110134	9.577E-09
		AR2,24	2	24	0.296149	0.0876962	3.3769889	0.0009524
		Intercept	1	0	290.10776	89.576213	3.2386696	0.0015045
R-119 Sherpur	(1,0,0)(2,0,0)12	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
		AR1,1	1	1	0.5622503	0.0715144	7.8620593	6.831E-13
		AR2,12	2	12	0.5531746	0.0822998	6.7214588	3.519E-10
		AR2,24	2	24	0.2648989	0.0862319	3.0719357	0.0025262
		Intercept	1	0	351.84718	98.258143	3.580845	0.000462

Name of rainfall station	ARIMA Model	Model Parameter						
R-107 Chhatak	(1,0,0)(2,0,0)12	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
		AR1,1	1	1	0.5622503	0.0715144	7.8620593	6.831E-13
		AR2,12	2	12	0.5531746	0.0822998	6.7214588	3.519E-10
		AR2,24	2	24	0.2648989	0.0862319	3.0719357	0.0025262
		Intercept	1	0	351.84718	98.258143	3.580845	0.000462
R-125Sheola	(1,0,0)(2,0,0)12	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
		AR1,1	1	1	0.6206417	0.0691543	8.9747339	1.845E-15
		AR2,12	2	12	0.4031892	0.0745843	5.4058187	2.758E-07
		AR2,24	2	24	0.3959028	0.0793066	4.992055	1.772E-06
		Intercept	1	0	306.81009	90.182785	3.4020916	0.0008752
R-130 Zakiganj.	(2,0,0)(2,0,0)12	Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
		AR1,1	1	1	0.9273448	0.1173495	7.9024169	8.041E-13
		AR1,2	1	2	-0.3501857	0.0957686	-3.6565806	0.0003634
		AR2,12	2	12	0.3323488	0.0957762	3.4700545	0.0006962
		AR2,24	2	24	0.2436615	0.0924652	2.6351699	0.0093785
		Intercept	1	0	326.42385	56.741583	5.7528153	5.485E-08

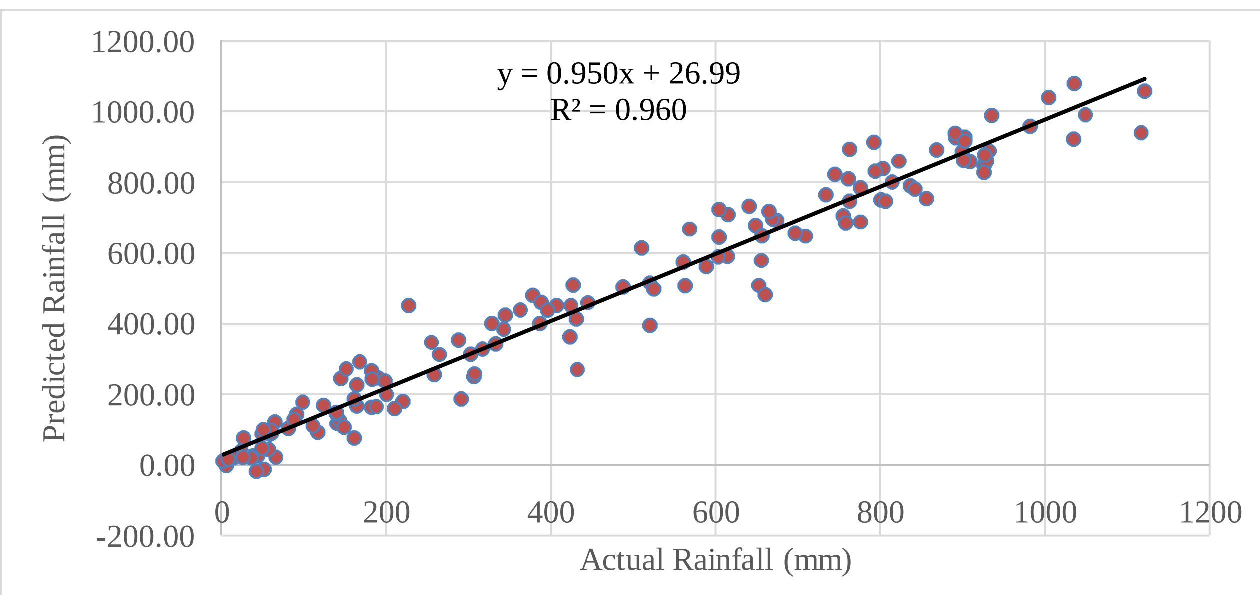

Figure-10
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Sulaghar (Sunamganj)

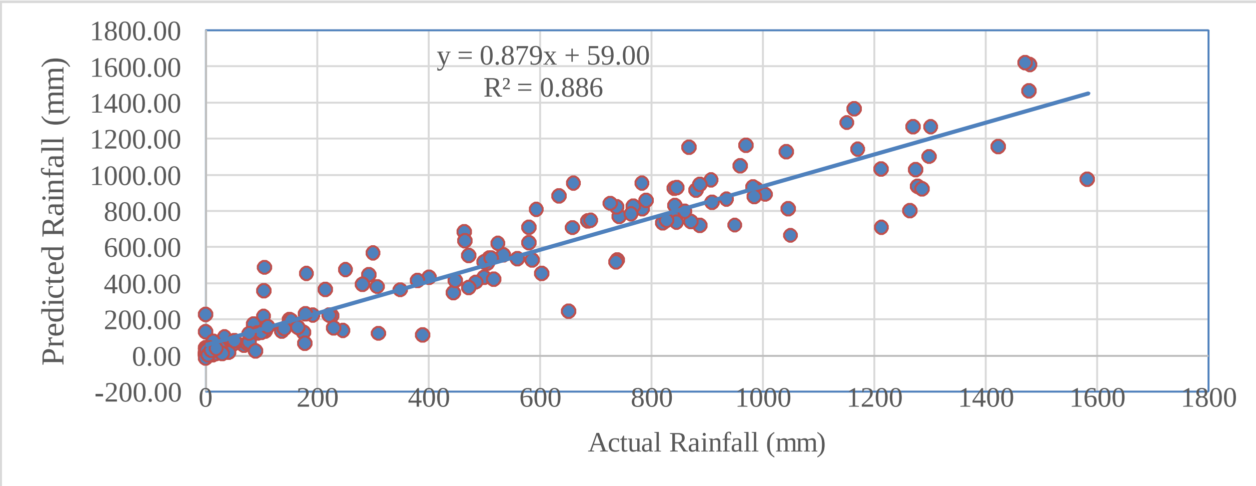

Figure-11
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Gobindganj

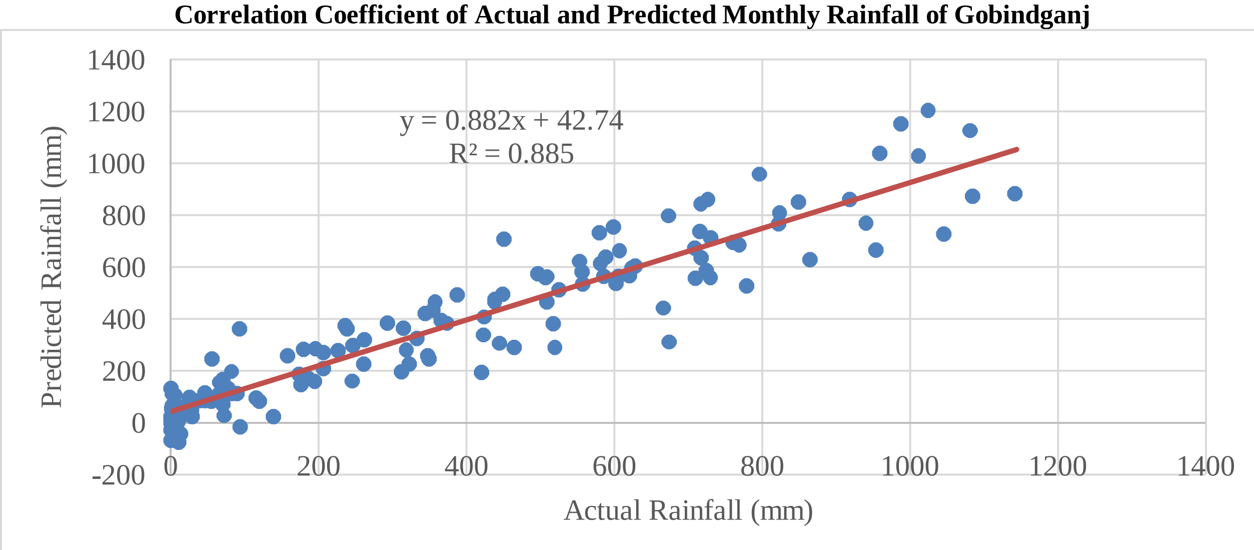

Figure-12
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Lalakhal

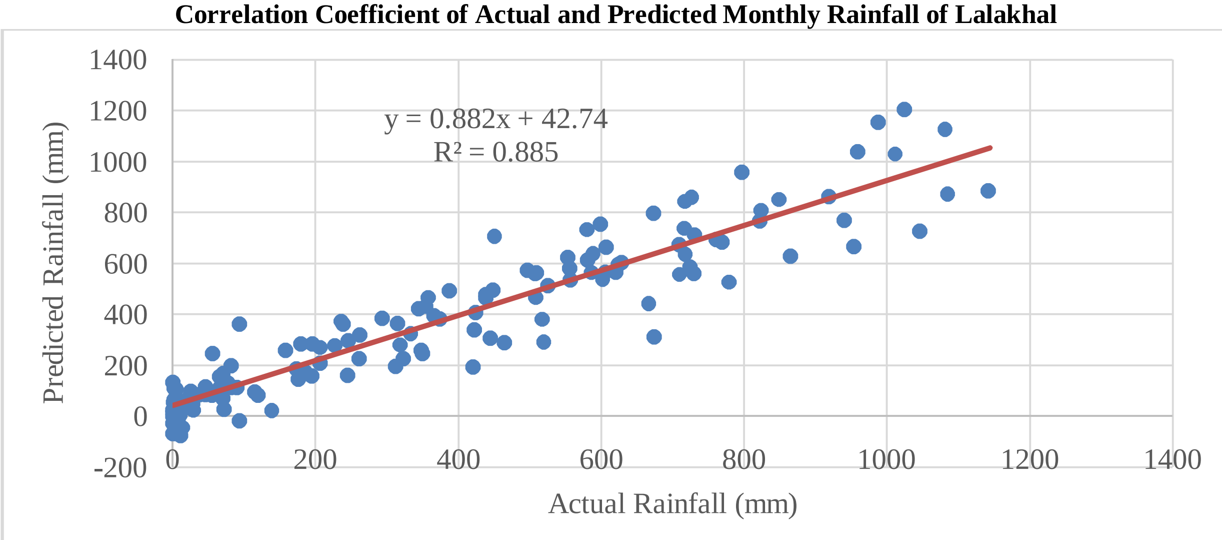

Figure-13
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Kanaighat

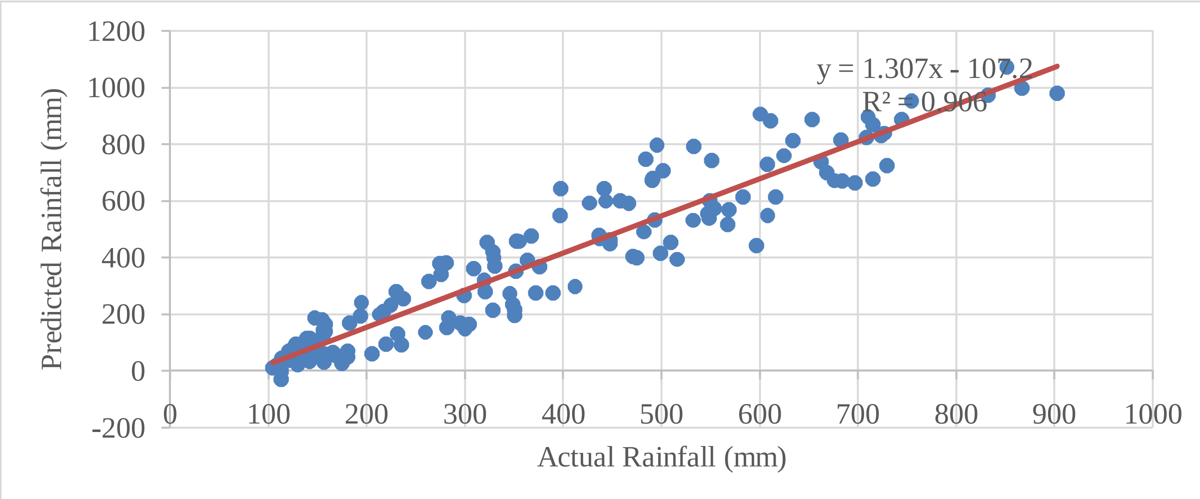

Figure-14
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Sherpur

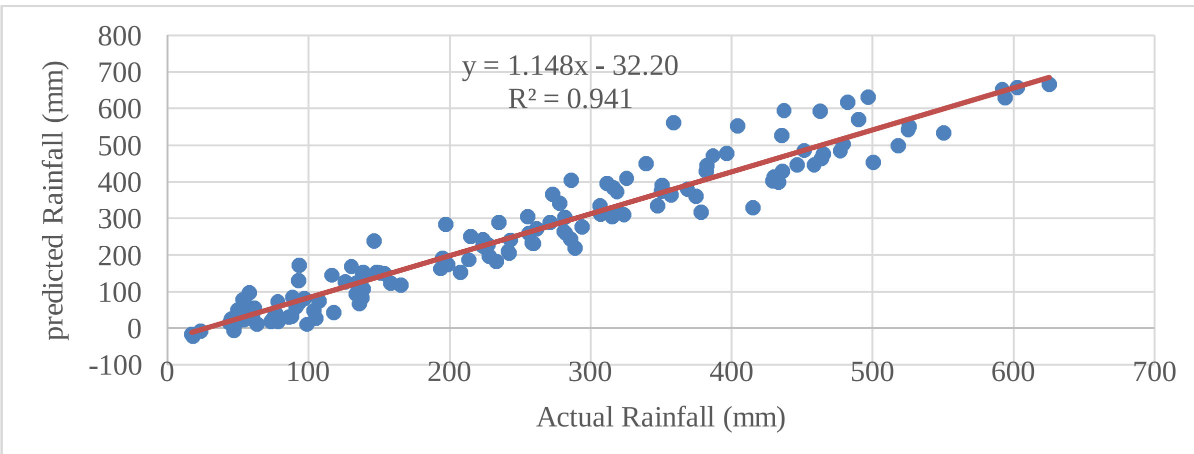

Figure-15
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Chhatak

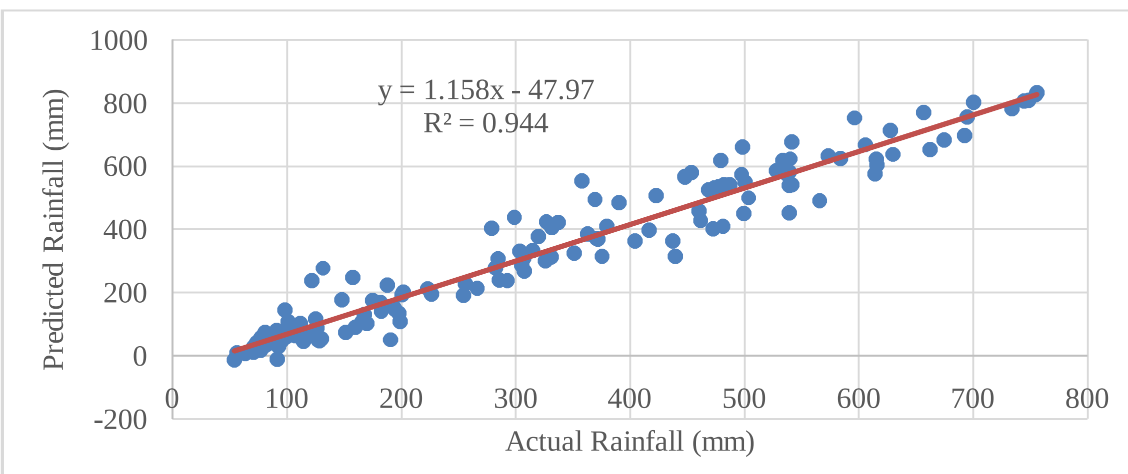

Figure-16
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Sheola

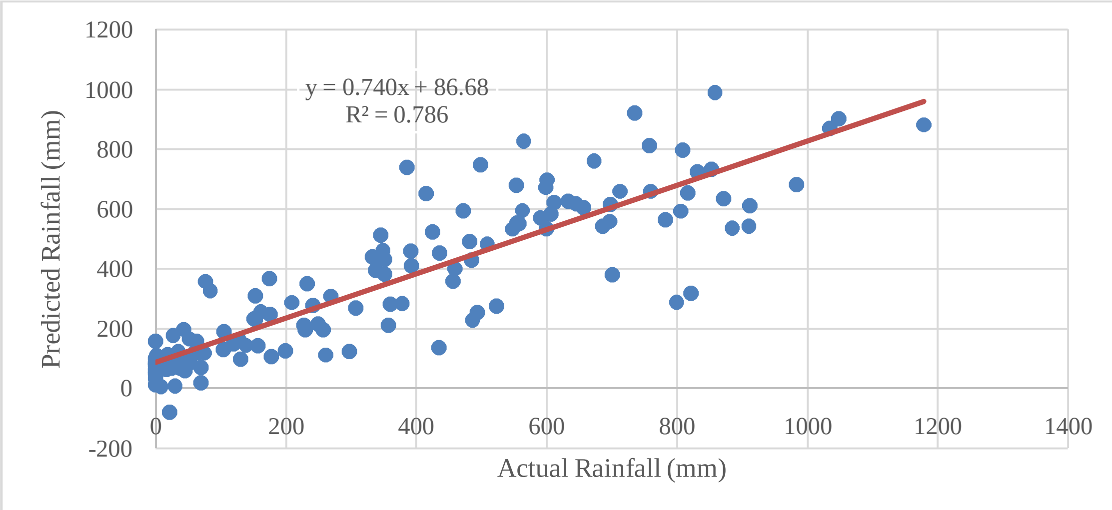

Figure-17
Correlation Coefficient of Actual and Predicted Monthly Rainfall of Zakiganj

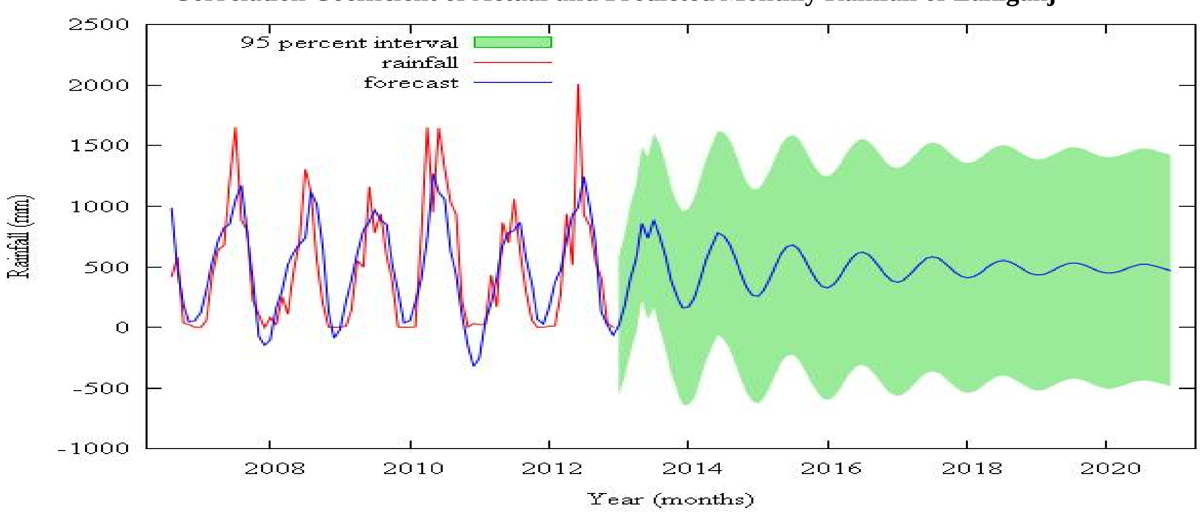

Figure-18
Actual and Forecasted Monthly Rainfall of Gobindganj

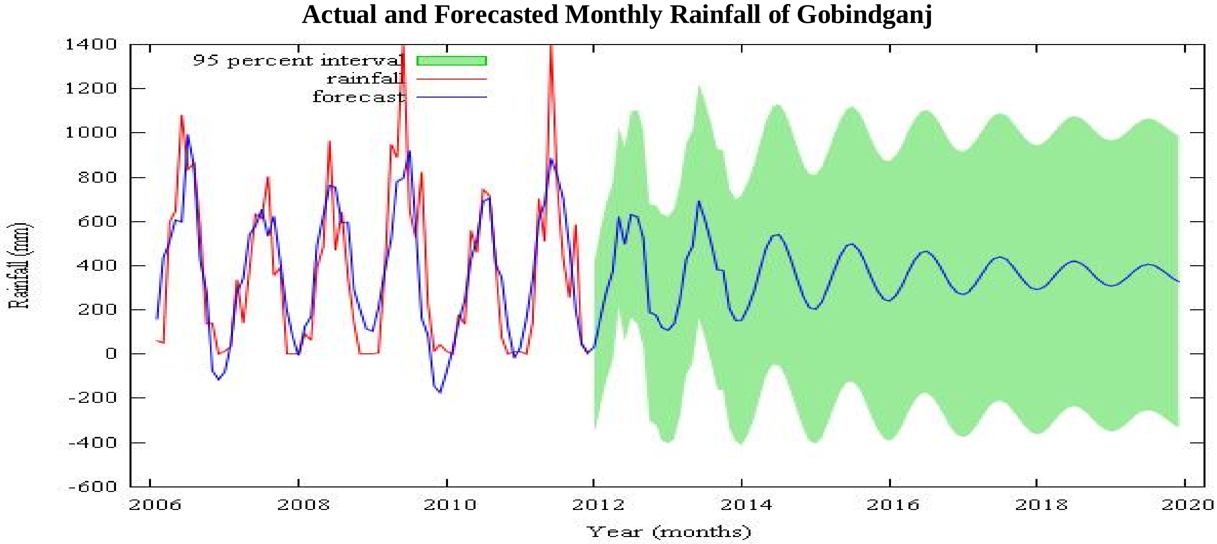

Figure-19
Actual and Forecasted Monthly Rainfall of Lalakhal

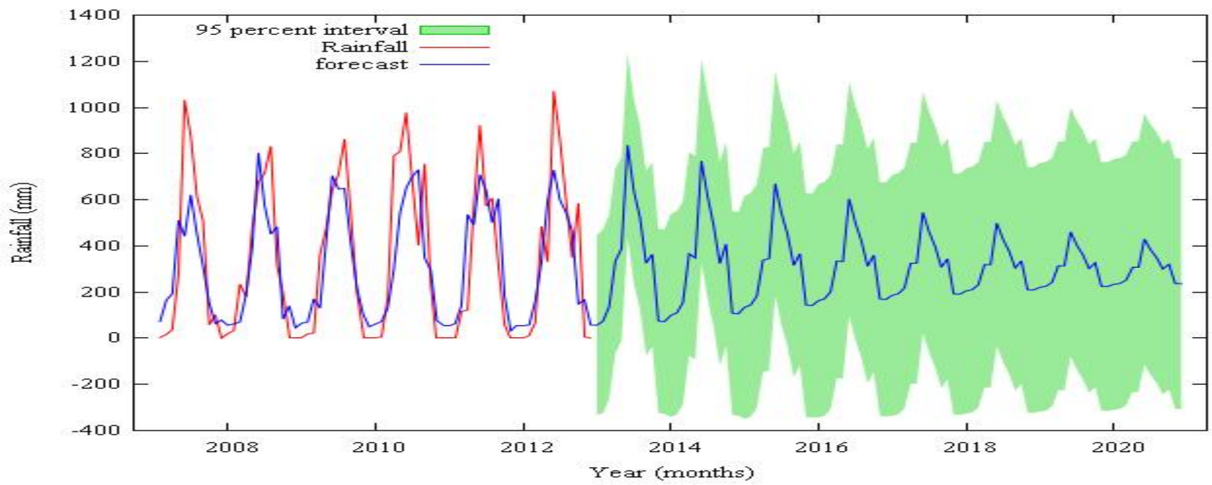

Figure-20
Actual and Forecasted Monthly Rainfall of Kanaighat

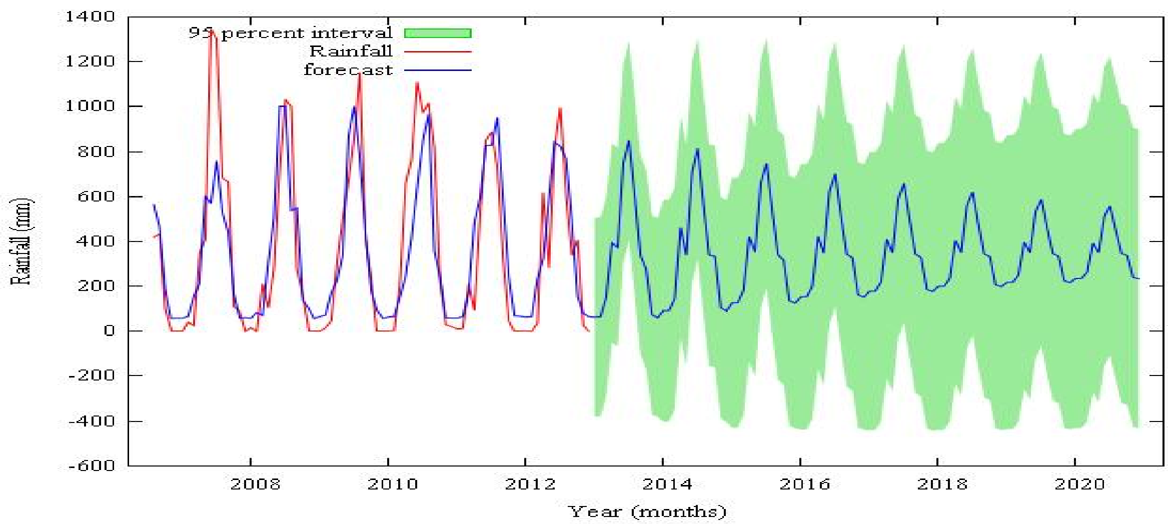

Figure-21
Actual and Forecasted Monthly Rainfall of Sherpur

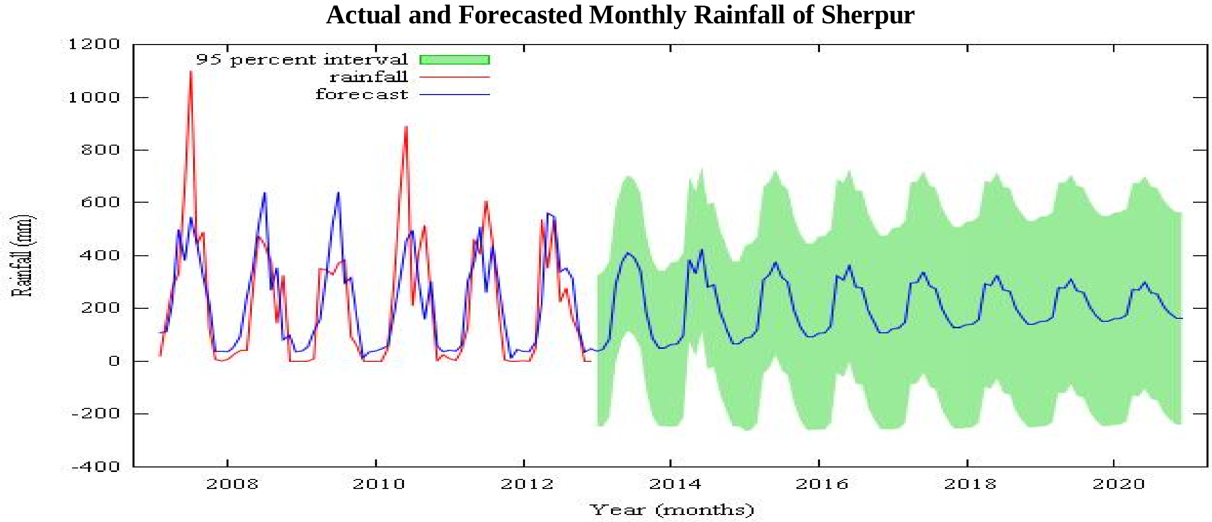

Figure-22
Actual and Forecasted Monthly Rainfall of Chhatak

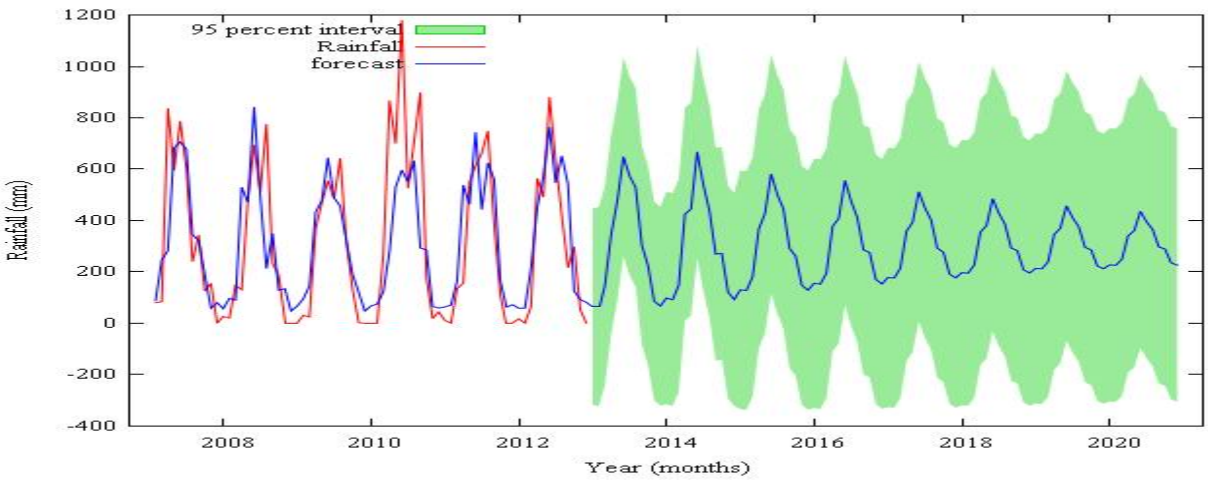

Figure-23
Actual and Forecasted Monthly Rainfall of Sheola

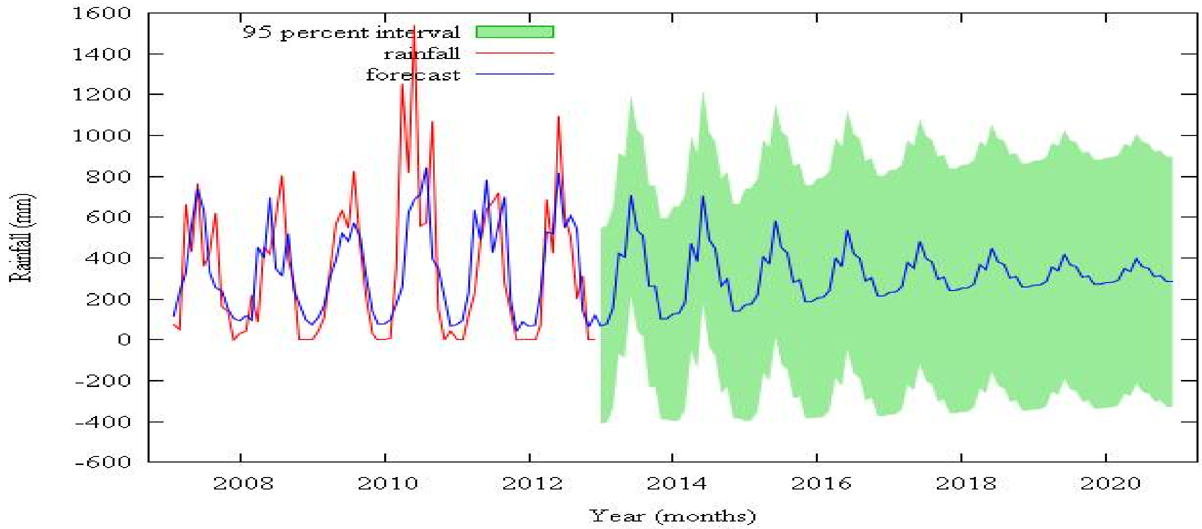

Figure-24
Actual and Forecasted Monthly Rainfall of Zakiganj

Conclusion

ARIMA model was applied to do monthly rainfall forecasting in this study. The models were developed and tested by a monthly rainfall record of eight rainfall stations (Snamganj, Lalakhal, Gobindoganj, Chattak, Sherupur, Sheola, Zakiganj, Kanaighat) ranging from 2001-2012 in the North Western part of Bangladesh. The prediction model is reliable as the RMSE values on test data are comparatively less. The rainfall forecasts are accurate enough comparing the fitted and actual values of rainfall data using the specific model. The predicted monthly rainfall data found from ARIMA $(2,0,2)x(1,0,1)_{12}$ with 95% confidence interval would enable decision makers to formulate strategies, give priorities and use water resources properly in Sylhet region. The probable excess rainwater could be stored and used when necessary.

References

1. Winstanley D. (1973a). Recent rainfall trends in Africa, the middle East and India‖, *Nature*, 243, 464–465.

2. Winstanley D. (1973b). Rainfall Patterns and General atmospherically circulation', *Nature*, 245, 190–194.

3. Makridakis S., Wheelwright S.C. and Hyndman R.J. (1998). Forecasting Methods and Applications, 3rd Edition, John Wiley, New York.

4. Chiew F.H.S., Stewardson M.J. and McMahon T.A. (1993). Comparison of six rainfall-runoff modelling approaches. *Journal of Hydrology*, 147(1), 1-36.

5. Kuo J.T. and Y.H. Sun (1993). An intervention model for average 10 day stream flow forecast and synthesis. *J. Hydrol.*, 151, 35-56.

6. Box G.E.P., Jenkins G.M. and Reinsel G.C. (1994). Time series analysis: Forecasting and control, Pearson Education, Delhi.

7. Langu E.M. (1993). Detection of changes in rainfall and runoff patterns. *J. Hydrol.*, 147, 153-167

8. Bell W.R. (1984). An introduction to forecasting with time series models. *Insurance: Mathematics and Economics*, 3(4), 241-255.

9. Lee C.M. and Ko C.N. (2011). Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, 38(5), 5902-5911.

10. Schwarz G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.