



# Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method

Ishtiak Mahmud<sup>1†</sup>, Sheikh Hefzul Bari<sup>2</sup>, M. Tauhid Ur Rahman<sup>3</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh

<sup>2</sup>Department of Civil Engineering, Leading University, Sylhet-3100, Bangladesh

<sup>3</sup>Department of Civil Engineering, Military Institute of Science and Technology, Dhaka-1216, Bangladesh

## ABSTRACT

Rainfall is one of the most important phenomena of the natural system. In Bangladesh, agriculture largely depends on the intensity and variability of rainfall. Therefore, an early indication of possible rainfall can help to solve several problems related to agriculture, climate change and natural hazards like flood and drought. Rainfall forecasting could play a significant role in the planning and management of water resource systems also. In this study, univariate Seasonal Autoregressive Integrated Moving Average (SARIMA) model was used to forecast monthly rainfall for twelve months lead-time for thirty rainfall stations of Bangladesh. The best SARIMA model was chosen based on the RMSE and normalized BIC criteria. A validation check for each station was performed on residual series. Residuals were found white noise at almost all stations. Besides, lack of fit test and normalized BIC confirms all the models were fitted satisfactorily. The predicted results from the selected models were compared with the observed data to determine prediction precision. We found that selected models predicted monthly rainfall with a reasonable accuracy. Therefore, year-long rainfall can be forecasted using these models.

**Keywords:** ARIMA model, Bangladesh, Forecast, Monthly rainfall, Time series

## 1. Introduction

Prediction of rainfall is tough due to its non-linear pattern and a large variation in intensity. Till today, numerous techniques have been used to predict rainfall. Among them, Autoregressive Integrated Moving Average (ARIMA) modeling, introduced by Box and Jenkins [1] is an effective method. The Box-Jenkins Seasonal ARIMA (SARIMA) model has several advantages over other models, particularly over exponential smoothing and neural network, due to its forecasting capability and richer information on time-related changes [2]. ARIMA model consider the serial correlation which is the most important characteristic of time series data. ARIMA model also provides a systematic option to identify a better model. Another advantage of ARIMA model is that the model uses less parameter to describe a time series.

Rahman et al. [3] used a comparative study of ANFIS and ARIMA model for the weather forecast in Dhaka city and found that ARIMA model performs better than ANFIS. Johnson and Montgomery [4] consider Box and Jenkins ARIMA method as possibly the most

precise method for forecasting of time series data. According to Dizon [5], the Box-Jenkins methodology is particularly suited for the development of a model that exhibit strong seasonal behavior. Momani [6] successfully used ARIMA model for predicting rainfall trend of Jordan.

In Bangladesh, ARIMA modeling has been applied to few of the available rainfall stations. Mohsin et al. [7] used ARIMA model to forecast the rainfall of Dhaka city. Rainfall forecast for Sylhet station has been done by Bari et al. [8]. They have concluded that ARIMA model was adequate in predicting monthly rainfall for these stations. However, rainfall forecast for all over Bangladesh has not been done yet. In this paper, Box-Jenkins approach has been used to build seasonal ARIMA model of monthly rainfall data of various stations in Bangladesh. Development of a reliable forecasting model of rainfall series will be helpful to mitigate natural hazards like flood and drought. A precise year-long prediction will also help in agricultural water use planning. The forecasted data series will also reveal likely future rainfall trend.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2017 Korean Society of Environmental Engineers

Received May 20, 2016 Accepted November 24, 2016

† Corresponding author  
Email: ishtiak006@gmail.com  
Tel: +88-01791327282

## 2. Materials and Methods

Bangladesh is a low-lying, riverine country with a largely wet jungle coastline. It extends from 20°34' to 26°38' north latitude and 88°01' to 92°41' east longitude [9]. The country has a sub-tropical humid climate characterized by extensive seasonal variations in rainfall, moderately warm temperature and high humidity [9]. Rainfall in Bangladesh varies from 1,400 mm in the west to more than 4,300 mm in the east. About 75% rainfall occurs during monsoon in Bangladesh [10]. A more detail discussion about the climate of Bangladesh is given by Rashid [9].

### 2.1. Data Collection and Quality Control

Rainfall data for 35 stations were collected from Bangladesh Agricultural Research Council (BARC). BARC collect weather data from Bangladesh Meteorological Department (BMD).

Quality control of data is an essential step before the calculation of indices because erroneous outliers can mislead the index calculation [11]. The key purpose of the quality control procedure was to identify errors in data processing, such as errors in manual keying [12]. As a first step, missing values were screened. Stations, containing more than 2% missing value were excluded from the modeling. Other checks such as identification of consecutive month with the same amount of rainfall along the year, the precipitation

value below 0 mm, winter rainfall more than average rainfall and monsoon rainfall below the threshold limit were also carried out [10]. After checking these, thirty stations having long-term data (more than 20 y) up to the year 2013 and passing the above tests were used in the present study. Locations of the stations are shown in Fig. 1.

The homogeneity of the data series was identified using Standard Normal Homogeneity test [13], Von Neumann Ratio test [14], Buishand Range test [15] and Pettitt test [16]. The rainfall data sets of all the stations were found homogenous except for Shrimangal station. There was a breakpoint for Shrimangal at 1961. Therefore, rainfall data from 1961 to 2013 was used for this station.

### 2.2. The ARIMA Model

Seasonal ARIMA model, proposed by Box and Jenkins [1] was used for model building and forecasting monthly rainfall. Seasonal ARIMA model can be labeled as ARIMA (p, d, q) \* (P, D, Q)<sub>s</sub> where (p, d, q) is the nonseasonal part and (P, D, Q)<sub>s</sub> is the seasonal part of the model which could be written as:

$$\phi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D z_t = \theta_q(B)\Theta_q(B^S)a_t \quad (1)$$

Where,  $p$  = non-seasonal Auto Regressive (AR) order,  $d$  = non-seasonal differencing,  $q$  = non-seasonal Moving Average (MA) order,  $P$  = seasonal AR order,  $D$  = seasonal differencing,  $Q$  = seasonal MA order, and  $S$  = length of the season.

Building an ARIMA model consists of four systematic stages (identification, estimation, diagnostic check and application or forecast). The identification phase involves the improvement of the stationarity and normality of the data. At this stage, the general form of the model is estimated. Model parameters are calculated using the method of maximum likelihood at estimation stage. The diagnostic checks are performed to reveal the possible inadequacies and to select the best model. Finally, the forecasting of the rainfall time series is done.

## 3. Results and Discussion

### 3.1. Model Identification

The identification stage involves checking the stationarity and normality of time series data. Initially, the data series were analyzed to check if the data are stationary and if there are any seasonality exists. The temporal correlation structure of the monthly time series was identified using Autocorrelation (ACF) and Partial Autocorrelation (PACF) Function [17]. Inspecting the time series plot, ACF, and PACF a clear seasonality of twelve months cycle was detected. Therefore, seasonal differencing of an appropriate order was done to remove seasonality and to achieve stationarity of data. A first order seasonal differencing was adequate to achieve stationarity for all the stations.

For example, time series plot of Rajshahi station is shown with ACF and PACF plots in Fig. 2. The sine-wave formation of ACF indicates that the data is seasonal. After a first order seasonal differencing ACF and PACF plot (Fig. 3) shows that the series is stationary. The ACF and PACF of the differenced series show

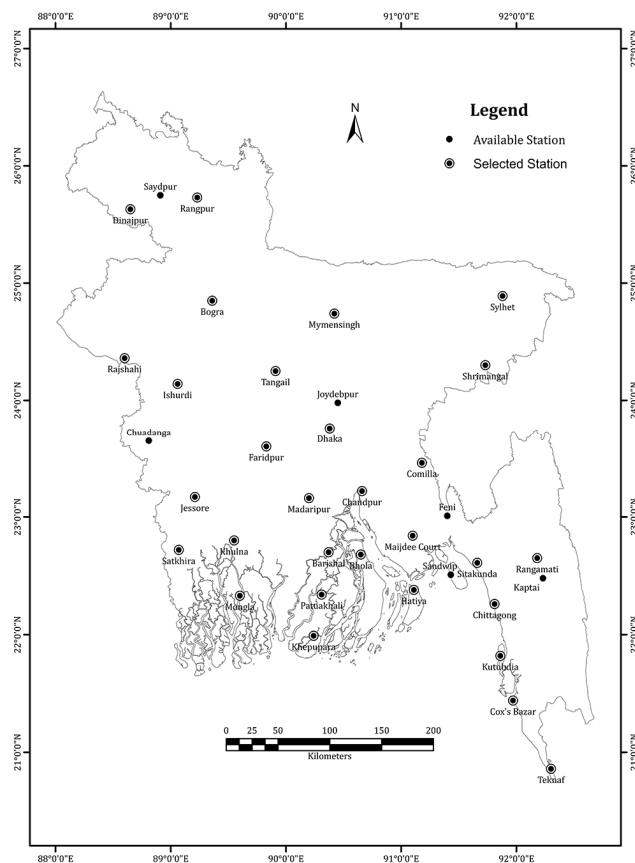


Fig. 1. Study area.

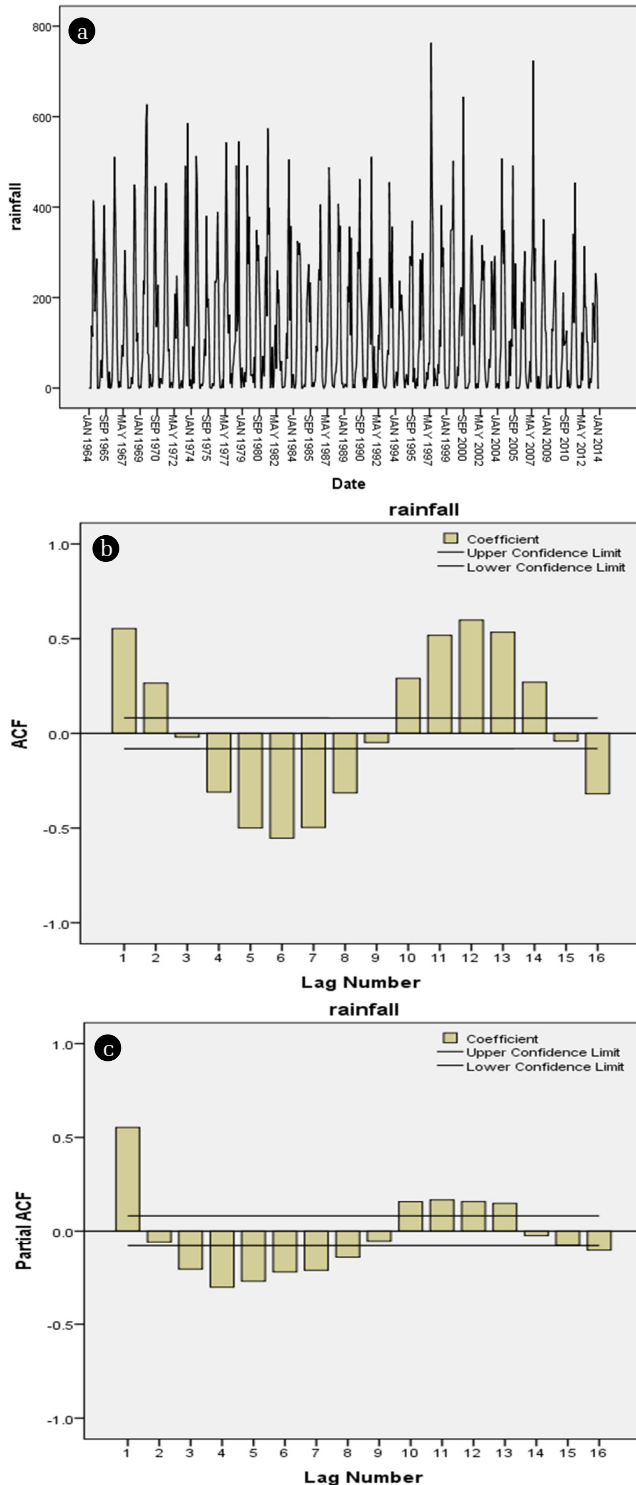


Fig. 2. Time series (a), ACF (b) and PACF (c) plot of rainfall for Rajshahi station.

no significant spikes indicating simplest Seasonal ARIMA process. However, the model could be a combination of both AR and MA process. The ACF and PACF plot direct a possible SARIMA model

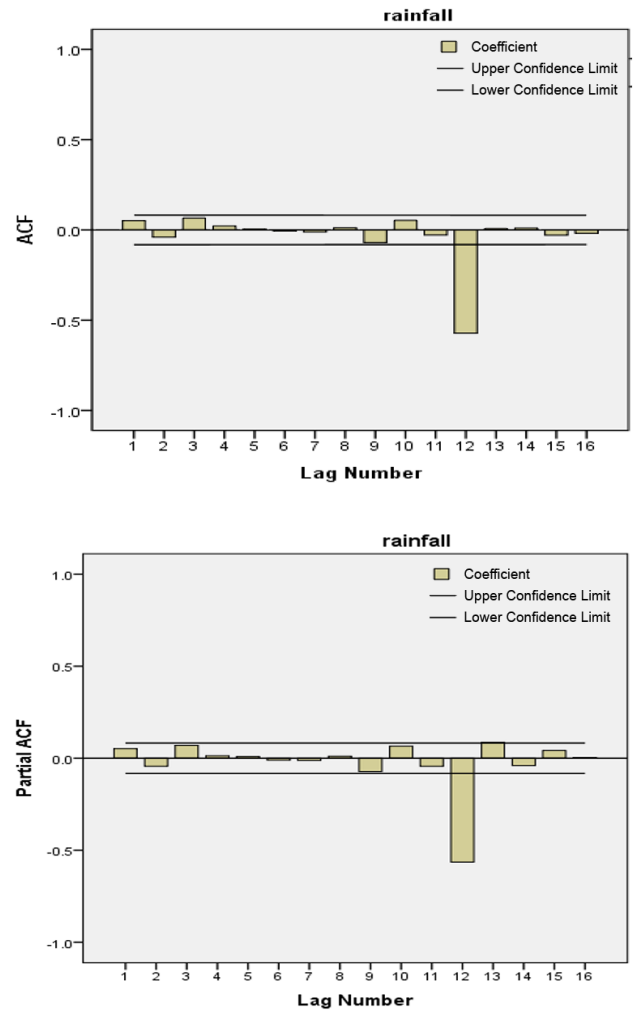


Fig. 3. ACF and PACF plot of transformed series (Rajshahi).

with  $P = 0-2$  and  $Q = 0-2$ . All the possible combinations using several  $P$  and  $Q$  ranging from zero to two were examined to determine the best ARIMA model from the nominee models. The model that gives the best combination of minimum RMSE, maximum R-squared, and least Normalized BIC was selected as the best fit model.

### 3.2. Parameter Estimation

Primary estimation of the parameters was done from AR and MA at the identification stage. This preliminary evaluation was then used to compute the final parameter by the procedure described by Box and Jenkins (1976) [1]. The value of the parameters, standard error (SE), t-ratio, and p-values for each station are shown in Table 1. The standard error calculated for the relevant parameter is small compared to the parameter values. Therefore, the parameters are statistically significant [2].

### 3.3. Diagnostic Check

The diagnostic check was done once the model parameters were

**Table 1.** Statistical Analysis of Parameter

Station	Model	Parameter		Estimate	SE	t	Significance (p value)
Barishal	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.862	0.032	26.646	0
Bhola	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.959	0.053	18.099	0
Bogra	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.969	0.013	72.797	0
Chandpur	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.97	0.027	35.299	0
Chittagong	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.967	0.027	36.259	0
Comilla	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.972	0.028	34.524	0
Cox's Bazar	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.919	0.021	43.526	0
		P	lag1	-0.573	0.368	-1.558	0.12
Dhaka	ARIMA(0,0,0)(1,1,2)	Q	lag1	0.485	0.387	1.254	0.21
			lag2	0.515	0.386	1.334	0.183
Dinajpur	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.929	0.02	46.887	0
Faridpur	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.961	0.03	31.633	0
		q	lag1	-0.158	0.04	-3.906	0
Hatiya	ARIMA(0,0,1)(0,1,1)	Q	lag1	0.939	0.022	42.319	0
Ishurdi	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.991	0.042	23.656	0
			lag1	0.979	0.037	26.42	0
Jessore	ARIMA(0,0,0)(0,1,2)	Q	lag2	-0.044	0.037	-1.191	0.234
		q	lag1	0.145	0.046	3.176	0.002
Khepupara	ARIMA(0,0,1)(0,1,1)	Q	lag1	0.921	0.028	33.039	0
Khulna	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.955	0.02	47.747	0
Kutubdia	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.929	0.043	21.66	0
Madaripur	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.926	0.034	27.165	0
Maijdeecourt	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.971	0.048	20.361	0
			lag1	0.96	0.062	15.548	0
Mongla	ARIMA(0,0,0)(0,1,2)	Q	lag2	-0.079	0.062	-1.272	0.205
Mymensingh	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.962	0.015	64.464	0
Patuakhali	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.939	0.029	32.123	0
		P	lag1	-0.201	0.041	-4.853	0
Rajshahi	ARIMA(0,0,0)(1,1,1)	Q	lag1	0.986	0.033	29.441	0
		P	lag1	-0.089	0.043	-2.079	0.038
Rangamati	ARIMA(0,0,0)(1,1,1)	Q	lag1	0.972	0.031	31.371	0
Rangpur	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.945	0.02	47.163	0
Satkhira	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.983	0.039	25.329	0
Shrimangal	ARIMA(0,0,0)(0,1,2)	Q	lag1	0.948	0.022	43.083	0
Sitakunda	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.9	0.036	24.749	0
			lag1	1.098	0.273	4.023	0
Sylhet	ARIMA(0,0,0)(0,1,2)	Q	lag2	-0.102	0.045	-2.243	0.025
Tangail	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.957	0.082	11.672	0
Teknaf	ARIMA(0,0,0)(0,1,1)	Q	lag1	0.917	0.03	30.931	0

calculated to verify the adequacy of the prediction. The residuals should be white noise for a useful forecasting model. Several tests carried out on residuals are described below:

### 3.3.1. ACF and PACF of residuals

The majority of the ACF and PACF values of residuals lie within the confidence limit which indicates no significant correlation among them. Illustration of the ACF and PACF of residuals for rainfall series of Rajshahi station are given in Fig. 4. The figure

shows that the residuals are white noise.

### 3.3.2. Histogram of residual

Histogram of residuals of rainfall series of Rajshahi station (Fig. 4(b)) shows that the residuals are normally distributed. Remaining series also exhibit similar result.

### 3.3.3. Normal probability of residuals

The cumulative distribution of the residuals data usually appears

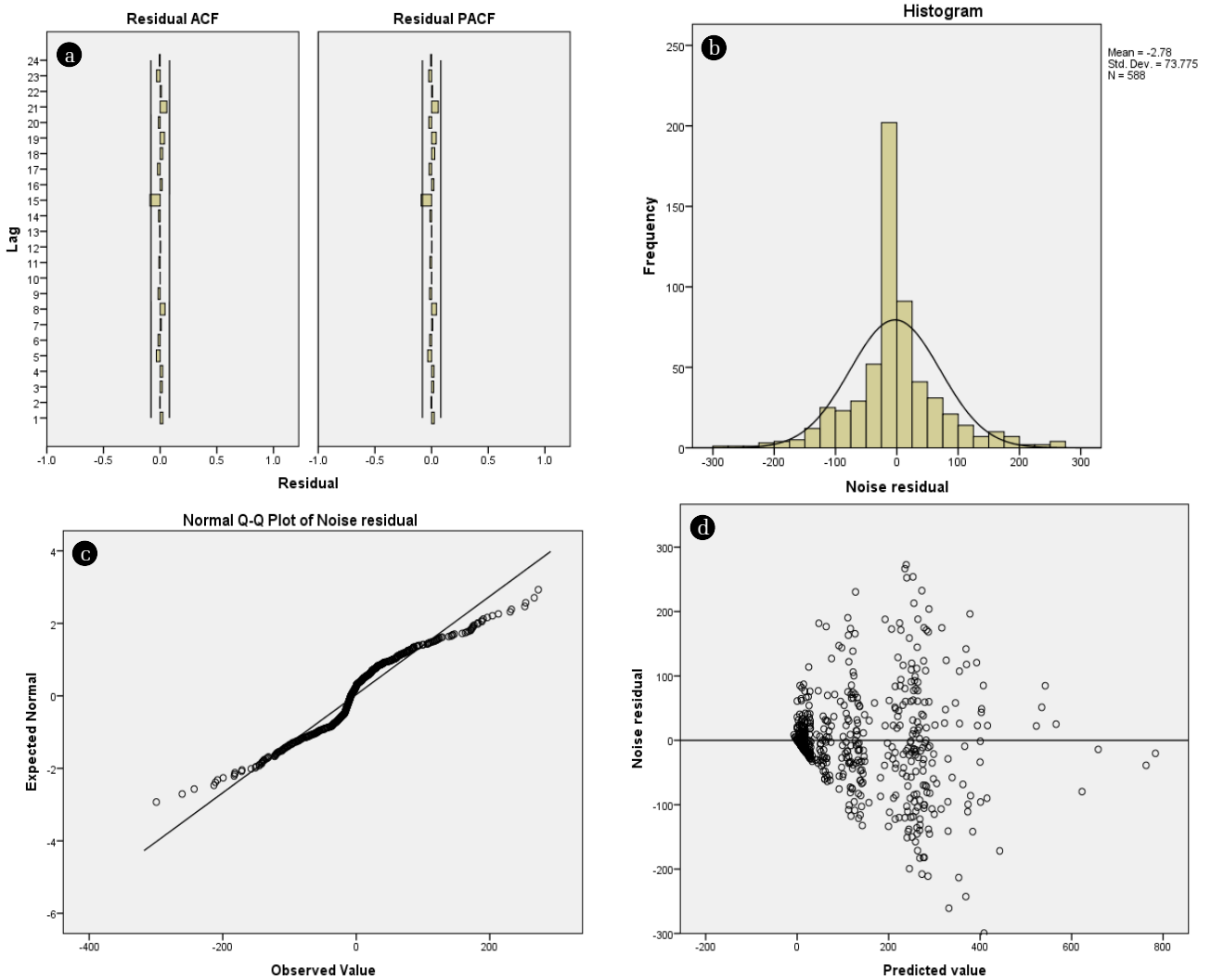


Fig. 4. Diagnostic check for best fitted model for rainfall series of Rajshahi.

as a straight line when plotted on normal probability paper [18]. The normal probability plot of residuals were found fairly linear for all stations, indicating the residuals are normally distributed. Fig. 4(c) represents the normal probability of residuals for Rajshahi station.

**3.3.4. Residuals versus prediction**

Residuals were plotted against predicted values. The plot (Fig. 4(d)) for Rajshahi station shows that residuals are evenly distributed around mean, which indicate that the model is well fitted. Similar results were found for other stations also.

**3.3.5. Lack of fit test**

The modified Ljung Box statistics was used to verify the null hypothesis that the model is correctly specified [19]. The test statistics is shown in Table 2. The associated P-value is greater than 0.05 for maximum stations, which hold the null hypothesis of being white noise. The test indicates the models are adequate.

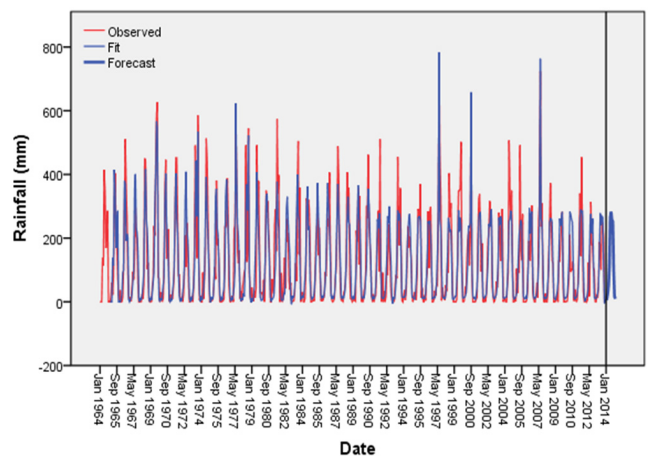


Fig. 5. Comparison of observed data with predicted data using best ARIMA model (Rajshahis).

**Table 2.** Performance Criteria of Selected SARIMA Model

Station	Model	Model Fit statistics			Ljung-Box Q(18)		
		R-squared	RMSE	Normalized BIC	Statistics	DF (Degree of freedom)	Sig.
Barishal	ARIMA(0,0,0)(0,1,1)	0.672	104.2	9.3	21.1	17	0.221
Bhola	ARIMA(0,0,0)(0,1,1)	0.708	108.9	9.4	18.0	17	0.386
Bogra	ARIMA(0,0,0)(0,1,1)	0.765	81.60	8.9	20.6	17	0.241
Chandpur	ARIMA(0,0,0)(0,1,1)	0.728	99.80	9.3	9.5	17	0.921
Chittagong	ARIMA(0,0,0)(0,1,1)	0.735	147.4	10.0	10.8	17	0.862
Comilla	ARIMA(0,0,0)(0,1,1)	0.680	104.2	9.3	12.9	17	0.742
Cox's Bazar	ARIMA(0,0,0)(0,1,1)	0.813	164.4	10.3	6.2	17	0.992
Dhaka	ARIMA(0,0,0)(1,1,2)	0.720	93.20	9.2	11.0	15	0.749
Dinajpur	ARIMA(0,0,0)(0,1,1)	0.800	85.90	9.0	27.6	17	0.049
Faridpur	ARIMA(0,0,0)(0,1,1)	0.704	84.60	8.9	12.5	17	0.763
Hatiya	ARIMA(0,0,1)(0,1,1)	0.832	120.6	9.7	20.2	16	0.209
Ishurdi	ARIMA(0,0,0)(0,1,1)	0.787	70.50	8.7	18.7	17	0.342
Jessore	ARIMA(0,0,0)(0,1,2)	0.730	78.20	8.8	11.7	16	0.76
Khepupara	ARIMA(0,0,1)(0,1,1)	0.762	123.2	9.7	18.9	16	0.271
Khulna	ARIMA(0,0,0)(0,1,1)	0.716	85.30	8.9	20.6	17	0.243
Kutubdia	ARIMA(0,0,0)(0,1,1)	0.792	146.2	10.1	3.4	17	0.96
Madaripur	ARIMA(0,0,0)(0,1,1)	0.707	91.30	9.1	8.3	17	0.96
Maijdeecourt	ARIMA(0,0,0)(0,1,1)	0.799	131.5	9.8	23.8	17	0.124
Mongla	ARIMA(0,0,0)(0,1,2)	0.771	82.30	8.9	10.4	16	0.844
Mymensingh	ARIMA(0,0,0)(0,1,1)	0.729	105.2	9.4	18.8	17	0.337
Patuakhali	ARIMA(0,0,0)(0,1,1)	0.720	124.5	9.7	15.1	17	0.588
Rajshahi	ARIMA(0,0,0)(1,1,1)	0.741	74.40	8.7	9.0	16	0.911
Rangamati	ARIMA(0,0,0)(1,1,1)	0.721	124.1	9.7	16.2	16	0.434
Rangpur	ARIMA(0,0,0)(0,1,1)	0.763	100.1	9.3	4.6	17	0.999
Satkhira	ARIMA(0,0,0)(0,1,1)	0.701	82.60	8.9	16.4	17	0.496
Shrimangal	ARIMA(0,0,0)(0,1,1)	0.720	103.8	9.3	14.9	17	0.597
Sitakunda	ARIMA(0,0,0)(0,1,1)	0.732	147.7	10.1	11.0	17	0.853
Sylhet	ARIMA(0,0,0)(0,1,2)	0.750	169.3	10.3	7.1	16	0.971
Tangail	ARIMA(0,0,0)(0,1,1)	0.761	82.40	9.04	17.5	17	0.414
Teknaf	ARIMA(0,0,0)(0,1,1)	0.868	161.6	10.3	17.0	17	0.448

### 3.4. Rainfall Forecasting

The forecast is done for 12 months lead time, using the best model calculated from historical data. The plot between observed and predicted values (Rajshahi station) in Fig. 5 indicates that the predicted values follow the observed data closely enough. The performance criteria of the selected seasonal ARIMA model are shown in Table 2. The high value of R-square indicates the performance of the model is fair enough. The appropriateness of the models is confirmed by investigating Normalized BIC. The models have least BIC value, upheld the significance of the model. In addition to statistical terms, the performance of the developed models was compared with the available literature in the country (e.g. Bari et al. 2015 and Mahsin et al. 2012) and found sound in rainfall forecasting comparing to these.

### 4. Conclusions

Rainfall forecast becomes difficult in Bangladesh due to its non-linear pattern and the spatiotemporal variation in size. However, prediction of rainfall is necessary for flood management, rainwater harvesting, urban planning, water resource management, planning and optimal operation of the irrigation system. Considering the significance, rainfall forecasting using widely accepted ARIMA modeling technique is done in this paper. The highest R-squared value (0.868) was found for Teknaf station and the lowest value (0.672) was found for the Barishal station. Only two stations contain R-squared value below 0.70 which indicate the SARIMA models developed to forecast rainfall in the present study is reasonably precise. Hence, these SARIMA models can be used as a convenient tool for nationwide rainfall forecasting.

## References

1. Box GE, Jenkins GM. Time series analysis: Forecasting and control. Rev. ed. San Francisco: Holden-Day; 1976.
2. Mishra AK, Desai VR. Drought forecasting using stochastic models. *Stoch. Environ. Res. Assess.* 2005;19:326-339.
3. Rahman M, Islam AHMS, Nadvi SYM, Rahman RM. Comparative study of ANFIS and ARIMA model for weather forecasting in Dhaka. Informatics, Electronics & Vision (ICIEV), 2013 International Conference on, Dhaka; 2013. p. 1-6.
4. Johnson LA, Montgomery DC. Forecasting and time series analysis. New York: McGraw-Hill; 1976.
5. Dizon CQ. ARMA modeling of a stochastic process appropriate for the angat reservoir. *Philipp. Eng. J.* 2007;28:1-20.
6. Momani PENM. Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. *Am. J. Environ. Sci.* 2009;5:599-604.
7. Mahsin M, Akhter Y, Begum M. Modeling rainfall in Dhaka division of Bangladesh using time series analysis. *J. Math. Model. Appl.* 2012;1:67-73.
8. Bari SH, Rahman MT, Hussain MM, Ray S. Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil Environ. Res.* 2015;7:69-77.
9. Rashid HE. Geography of Bangladesh. 2nd ed. Dhaka: Univ. Press Ltd.; 1991. p. 545.
10. Shahid S. Rainfall variability and the trends of wet and dry periods in Bangladesh. *Int. J. Climatol.* 2010;30:2299-2313.
11. You Q, Kang S, Aguilar E, Yan Y. Changes in daily climate extremes in the eastern and central Tibetan Plateau during 1961-2005. *J. Geophys. Res-Atmos.* 2008;113:D07101.
12. Mortuza MR, Selmi S, Khudri MM, Ankur AK, Rahman MM. Evaluation of temporal and spatial trends in relative humidity and dew point temperature in Bangladesh. *Arab. J. Geosci.* 2014;7:5037-5050.
13. Alexandersson H. A homogeneity test applied to precipitation data. *J. Climatol.* 1986;6:661-675.
14. Von Neumann J. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* 1941;12:367-395.
15. Buishand TA. The analysis of homogeneity of long-term rainfall records in the Netherlands. KNMI Scientific Report WR 81-7, De Bilt, The Netherlands; 1981.
16. Pettitt AN. A non-parametric approach to the change-point problem. *J. Roy. Stat. Soc. C-App.* 1979;28:126-135.
17. Box G, Jenkins G, Reinsel G. Time series analysis: Forecasting & control [Internet]. 3rd ed. New Jersey: Prentice Hall; c1994 [cited 23 February 2016]. Available from: <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0130607746>.
18. Chow VT, Maidment DR, Mays LW. Applied hydrology [Internet]. New York: McGraw-Hill Book Company; c1988 [cited 23 February 2016]. Available from: <http://documentatiecentrum.watlab.be/imis.php?module=ref&refid=127685&basketaction=add>.
19. Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika* 1978;65:297-303.