International Journal of Mathematics and Computer Research
ISSN: 2320-7167
Volume 09 Issue 02 February 2021, Page no. - 2187-2192
Index Copernicus ICV: 57.55, Impact Factor: 7.184
DOI: **10.47191/ijmcr/v9i2.02**

Forecasting of Monthly Rainfall in Cumilla of Bangladesh, Using SARIMA Model

Md. Shahidul Hoque[1], Muhammad Alamgir Islam[2]
[1]Assistant Professor, Department of Statistics, University of Chittagong.
[2]Assistant Professor, Department of Statistics, University of Chittagong.

ARTICLE INFO
Published Online:
15 February 2021

ABSTRACT
In this paper, seasonal autoregressive integrated moving average (SARIMA) model is developed to predict monthly rainfall in Cumilla using data for the period 1971 to 2017.The stationary condition of the data series are observed by ACF and PACF plots and then checked using the statistic such as Augmented Dickey-Fuller Test (ADF). The ADF test confirms that the monthly rainfall is stationary because the p-value of 0.01 is less than 0.05.The model for which the values of the criteria are smallest is considered as the best model. We found that the SARIMA (1,0,0)(2,1,0)[12]has been fitted to the data based on the Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AIC_C) and Schwarz Bayesian Information Criterion (BIC). Using Root Mean Square Error (RMSE), Absolute Mean Error (AME) and Mean Absolute Percentage Error (MAPE) to measure forecast accuracy. Then forecast of the data have been made using selected type of SARIMA model for the next five years.

Corresponding Author:
Md. Shahidul Hoque

KEYWORDS: Time Series, Forecasting, SARIMA model, ADF test, AIC & BIC, Rainfall data.

1. INTRODUCTION

Bangladesh is an agriculture-based country where better parts of its total people are directly or indirectly connected the wide range of agricultural activities. Bangladesh has a subtropical monsoon climate characterized by wide seasonal variations in rainfall, high temperatures and humidity. 'There are three distinct seasons in Bangladesh: a hot, humid summer from March to June; a cool, rainy monsoon season from June to October; and a cool, dry winter from October to March.Heavy rainfall is characteristic of Bangladesh. With the exception of the relatively dry western region of Rajshahi, where the annual rainfall is about 1600 mm, most parts of the country receive at least 2000 mm of rainfall per year. Because of its location just south of the foothills of the Himalayas, where monsoon winds turn west and northwest, the regions in northeastern Bangladesh receives the greatest average precipitation, sometimes over 4000 mm per year'[14]. Rainfall is the total amount of separate water drops whichfall to the earth from the clouds, having been formed bythe condensation of water vapor in the atmosphere.Condensation of this water vapor is brought about by the riseof air considerable height above the earth's surface, during agiven time; rainfall is measure by an instrument call raingauge [12]. Weather in Cumilla is influenced by Tropical Wet & Dry climate. There are more than two months with less than 60 mm rainfall. All average monthly temperatures are greater than 18°C (64°F).Rainfall during dry seasons are less than 100 mm. Happens mostly around the tropics, also close to the equator. The hottest month is March**,** when max temperature is about 33℃ fourth week is the warmest usually. But be aware of Thunderstorm and Rain. The coldest month is December**.** In this month temperature could be even 10℃ at night! And be prepared for Fog and Rain. The average temperature in Cumilla is 25.5 °C. The annual rainfall is 2295 mm [5]. Rainfall is the most important natural catalyst that determines the agricultural production in Bangladesh. The productiveness of farmland as well as the firmness of land resources depends on heavily upon the rainfall. The variability of rainfall and the extreme high or low precipitations are very important for the agricultural production as well as the economic growth of the country. It is well known that the rainfall is changing globally; and the regional scales due to global warming. Bangladesh is considered to be one of the most vulnerable country for change of climate rapidly and shortage or heavy rainfall. This is due to its unique geographic location, dominance of floods and rising of sea level. Due to such importance of rainfall in Bangladesh we have selected a best SARIMA model for predicting monthly rainfall in Cumilla. In this study we have selected Cumilla district because it is one of the most agriculture area in Bangladesh.

2. REVIEW OF EMPIRICAL LITERATURE

A study by Ishtiak Mahmud and et.al (2017), monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method. They found that selected models predicted monthly rainfall with a reasonable accuracy [6]. J. C. Ramesh Reddy and et.al (2017), studyforecasting of monthly mean rainfall in Coastal Andhra.They found that the ARIMA (1,0,0)(2,0,0)[12] has been fitted to the data and the significance test has been made by using lowest AIC and BIC values[7]. J. C. Paul and et.al (2013), studying the Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh. ARIMA (2, 1, 2) is found as the best model for forecasting the SPL data series and forecasts of the data have been made using selected type of ARIMA model[8]. Md. Mahsin and et.al (2012) in their study modeling rainfall in Dhaka division of Bangladesh using time series analysis found that the ARIMA (0, 0, 1) (0, 1, 1)12 is the best model for forecasting the rainfall data [10]. Samuel Olorunfemi Adams and et.al (2019) in their study found that the SARIMA (1, 0, 1) (1, 1, 1)[12] model best fit the data and used to make forecast[13]. This study therefore attempts to identify the best SARIMA model that fits the dataand forecast the monthly rainfall of Cumilla using the selected SARIMA model for the next five years.

3. PURPOSE OF THE STUDY

The aim of this study is to use time series analysis to forecasting the monthly rainfall in Cumilla of Bangladesh. The specific objectives of this study are as follows:

1. To check whether the selected time series data is stationary or not.
2. Investigate the best SARIMA model using some selection criteria.
3. To fit and forecast the monthly rainfall using the selected SARIMA model.
4. Finally, forecast the monthly rainfall for the next five years.

4. DATA AND METHODOLOGY

In this study, we have done analysis like forecasting ofmonthly total rainfall of Cumilla for succeeding years. For thestudy, we have taken data of monthly total rainfall in mm unit from Bangladesh Meteorological Department (BMD), Dhaka, Bangladesh. The data is having the informationof monthly total rainfall from year 1971 to 2017. In this research work we have used Seasonal Autoregressive Integrated Moving Average (SARIMA) model for forecasting the monthly rainfall in Cumilla city. Augmented Dickey-Fuller test has been used to check the stationary of the data.

4.1. Statistical Software Package

In this study we have taken the help of Rprogramming that is now one of most demanded software inthe field of data science and statistics. In R software majorly we need packages for forecasting model. In this study we have used some popular packages for predicting the time series model. These packages are 'ggplot2', 'forecast' and 't-series'.

4.2. Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

In time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step can be applied one or more times to eliminate the non-stationarity of the mean function i.e., the trend .When the seasonality shows in a time series, the seasonal-differencing[4] could be applied to eliminate the seasonal component. The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past [3]. The I indicates that the data values have been replaced with the difference between their values and the previous values. The purpose of each of these features is to make the model fit the data as well as possible.

Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, *p* is the order of the autoregressive model, *d* is the degree of differencing and *q* is the order of the moving-average model. Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an algebraic statement that describes how a time series is statistically related to its own past. The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model given as;

$$SARIMA \underbrace{(p,d,q)}_{non-seasonal} \underbrace{(P,D,Q)_m}_{seasonal}$$

Where m refers to the number of periods in each season, and the uppercase P, D, Q refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model [11].

The SARIMA modeling approach is concerned with finding a parsimonious seasonal ARIMA model that describes the underlying generating processed of the observed time series. Box and Jenkins [2] established a three step modeling procedure: identification, estimation and diagnostic checking steps. The identification step is to tentatively choose one or more ARIMA/SARIMA model(s) using the estimated ACF and PACF plots. The ACF plot of the AR (Auto Regressive)/ SAR (Seasonal Auto Regressive) process shows an exponential decay while its PACF plot truncates at lag p seasonal lag p and diminishes to zero afterwards. The ACF plot of the MA process truncates to zero after lag q while its PACF decays exponentially to zero. The two processes: AR (p)/SAR (P) and MA (q)/SMA (Q), could be combined to form the ARMA (p, q)/SARMA (P, Q) process which has ACF and PACF that decays exponentially to zero. The maximum likelihood estimation method could be used to estimate the parameters of the identified model(s) in the identification stage. The last diagnostic checking stage involves assessing the adequacy of the identified and fitted models through possible statistically significant test on the residuals to verify its consistency with the white noise process e.g. the Ljung-Box test [9]. Finally, the best fitting model would be selected among other satisfactory, competing models e.g. the information criteria statistics on the basis of the AIC or BIC [1] rule of thumb (Models with the lowest information criterion is the best) and forecast is made with the model of best fit.Three goodness of fit statistics that are most commonly used for the model selection namely; Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc) and Schwarz Bayesian Information Criterion (BIC). The AIC, AICc and BIC are determined based on a likelihood function. The AIC, AICc and BIC are calculated using the formulas below:

$$\mathrm{AIC} = \frac{2k}{n} + \ln(SSE)$$

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k(k+1)}{(n-k-1)}$$

$$\mathrm{BIC} = \ln(SSE) + \frac{k}{n}\ln(n)$$

Where; n is the total number of observations, SSE is the sum of the squared errors and k=p+q+P+Q+d+m.

Accuracy Measures of the Forecast Values

The accuracy measures used in this paper are Root Mean Square Error (RMSE), Absolute Mean Error (AME) and Mean Absolute Percentage Error (MAPE).

Root Mean Square Error (RMSE): The square root of the sum of square of the deviation of the predicted values from the observed value dividing by their number of observation is known as the root mean square error. The root mean square error is defined as

$$\mathbf{RMSE} = \sqrt{\frac{1}{T}\sum_{I=1}^{T}\left(Z_{obs} - Z_{pred}\right)^2}$$

Where, T is the number of periods. This criterion is used for the comparison of the models in three periods.

Absolute Mean Error (AME): The mean of the absolute deviation of predicted and observed values is called absolute mean error and is defined as

$$\mathbf{AME} = \sum_{I=1}^{T}\frac{\left|Z_{obs} - Z_{pred}\right|}{T}$$

Mean Absolute Percentage Error (MAPE): The mean of the sum of absolute deviation of predicted and observed value dividing by the observed value is called mean absolute error. For comparison we have multiplied by 100, which is called mean absolute percent error and which is defined as

$$\mathbf{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|Z_{obs} - Z_{pred}\right|}{Z_{obs}} \times 100$$

Where, the parameters bear the usual meaning

5. RESULTS AND DISCUSSION

Table-1: Summary details

Cumilla	Min	1st quartile	Median	Mean	3rd quartile	Max
	0.0	11.0	124.0	171.0	277.8	892.0

From the above table seen that the minimum and maximum rainfall is 0.0 and 892.0, first and third quartiles are 11.0 and 277.8. Average rainfall of Cumila is 171.0 and median rainfall is 124.0. Depending on the followingsummary, we cannot give any decision about the rainfall of Cumilla.

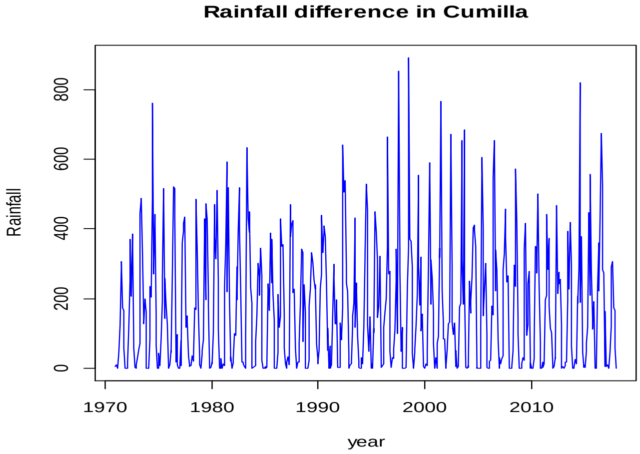
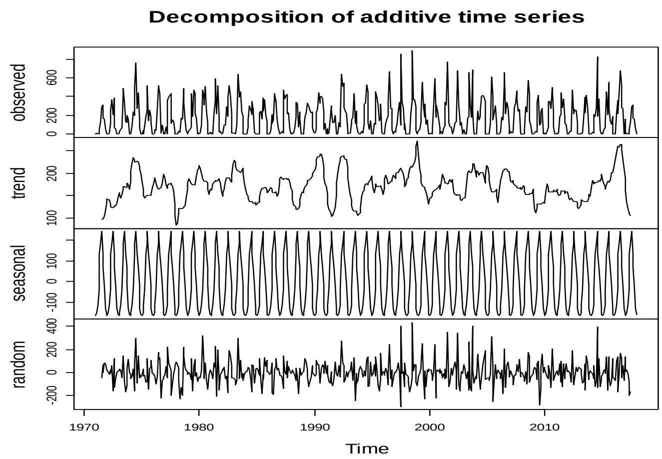


Figure-1: Time Series Plot of the Original data



Figur-2: Decomposition of additive time series

From the figure-1 indicates that, there exist seasonal variations in Cumilla for the original data.The figure-2 represents the original data, seasonal component, trend component and the remainder and this shows the periodic seasonal pattern extracted out from the original data and the trend. There is a bar at the right hand side of each graph to allow a relative comparisonof the magnitudes of each component. For this data the change in trend is less than the variation doing to the monthly variation. The stationary condition of the data series are observed by ACF and PACF plots shown in Figure- 3 and also checking stationary using the statistic such as Augmented Dickey-Fuller Test (ADF).

Hypothesis

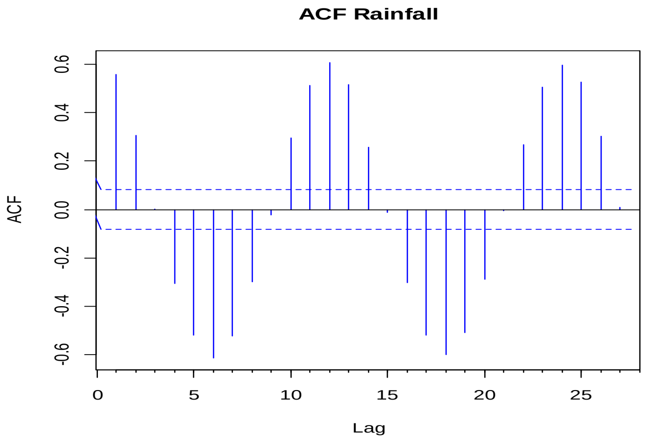
H_0: The rainfall data is unit root non stationary

H_A: The rainfall data is stationary.

Table-2: Augmented Dickey-Fuller test.

Test	Test Statistics	Lag Order	p-value
Dickey-Fuller	-12.292	8	0.01

Since p-value = 0.01 is less than level of significant = 0.05, the alternative hypothesis is accepted. Thus, strong evidence against the null hypothesis at 5% level of significance and we can observe the data is stationary presented in table-2.



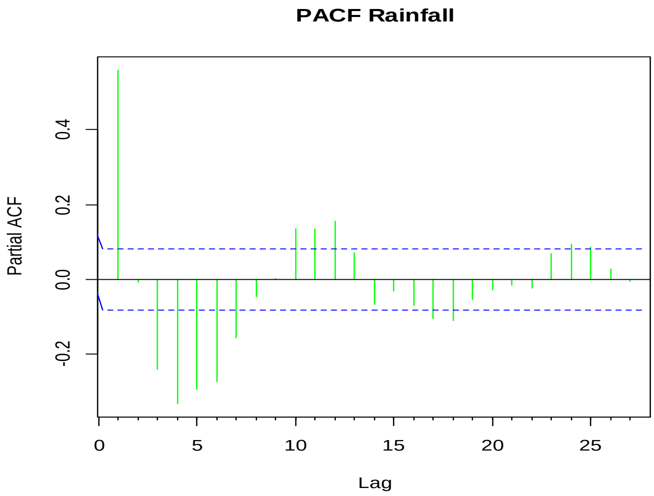


Figure-3: ACF and PACF plots respectively of the original data

5.1 SARIMA Model for Data

We are using auto arima method for finding the best forecasting model for data set. Automate ARIMA model for the data is SARIMA (1,0,0)(2,1,0)[12]. By applying auto arima, we acquired the best fitted model which has the smallest AIC, AICc and BIC. Comparing with AIC, AICc and BIC values, it is important that all models have the same orders of differencing. However, when comparing models using a test set, it does not matter how the forecasts were formed - the comparisons are always valid. The given model has been satisfied the residual tests. In practice, we would normally use the best model we could find, even if it did not satisfy all tests. From table-3, the forecast accuracy measures used SARIMA (1,0,0)(2,1,0)[12] model as the best model, because it has the least accuracy measures values.

Table-3: SARIMA (1,0,0)(2,1,0)[12]

Coefficients	AR(1)	SAR(1)	SAR(2)
Estimate	0.0776	-0.6989	-0.3855
SE	0.0427	0.0398	0.0401
sigma^2 estimated as 15390		log likelihood=-3446.5	
AIC=6901		AICc=6901.07	BIC=6918.25
RMSE = 122.3969		AME= 80.11452	MAPE = 0.7702246

5.2. Forecasting Of Monthly Rainfall

We evaluate forecasts with the selected model SARIMA (1,0,0) (2,1,0) [12] for the next five years are shown in the table-4.

Table-4: Forecasted amount of rainfall with confidence interval in Cumilla, Bangladesh during the period January 2018 to December 2022

Year	Forecasted Rainfall (millimeter)	Confidence Interval (95%) of Forecasted Rainfall (mm)	
		LCL	UCL
Jan 2018	2.1955234	-245.344218	240.9532
Feb 2018	23.3173924	-220.563175	267.1980
Mar 2018	100.1434367	-143.741537	344.0284
Apr 2018	158.6502698	-85.234730	402.5353
May 2018	256.1246416	12.239642	500.0096
Jun 2018	253.4018557	9.516856	497.2869
Jul 2018	514.2225059	270.337506	758.1075
Aug 2018	364.0903073	120.205307	607.9753
Sep 2018	195.3055958	-48.579404	439.1906
Oct 2018	178.6569170	-65.228083	422.5419
Nov 2018	91.9635053	-151.921495	335.8485
Dec 2018	50.7850110	-193.099989	294.6700
Jan 2019	0.9064496	-253.727706	255.5406
Feb 2019	13.2384625	-241.459124	267.9360
Mar 2019	119.1883777	-135.509592	373.8863
Apr 2019	186.9110873	-67.786884	441.6091
May 2019	177.0997989	-77.598173	431.7978

Jun 2019	281.3631524	26.665181	536.0611
Jul 2019	506.6021466	251.904175	761.3001
Aug 2019	408.6013085	153.903337	663.2993
Sep 2019	221.4159467	-33.282025	476.1139
Oct 2019	211.4391324	-43.258839	466.1371
Nov 2019	47.4827815	-207.215190	302.1808
Dec 2019	77.7323652	-176.965606	332.4303
Jan 2020	1.5118787	-271.483769	274.5075
Feb 2020	15.1498175	-257.952429	288.2521
Mar 2020	67.2768472	-205.826042	340.3797
Apr 2020	106.0067657	-167.096127	379.1097
May 2020	151.3406827	-121.762210	424.4436
Jun 2020	211.5565320	-61.546361	484.6594
Jul 2020	425.8871609	152.784268	698.9901
Aug 2020	355.4854189	82.382526	628.5883
Sep 2020	194.1832893	-78.919604	467.2862
Oct 2020	183.6477910	-89.455102	456.7507
Nov 2020	65.0953554	-208.007538	338.1982
Dec 2020	39.3227505	-233.780143	312.4256
Jan 2021	0.1069345	-309.905508	309.6916
Feb 2021	17.6988240	-292.307816	327.7055
Mar 2021	96.2189058	-213.788988	406.2268
Apr 2021	151.6607225	-158.347179	461.6686
May 2021	199.8048396	-110.203062	509.8127
Jun 2021	249.5693357	-60.438566	559.5772
Jul 2021	485.2391879	175.231286	795.2471
Aug 2021	375.4533535	65.445452	685.4613
Sep 2021	203.1529862	-106.854915	513.1609
Oct 2021	190.4363023	-119.571599	500.4442
Nov 2021	69.9303615	-240.077540	379.9383
Dec 2021	55.7818229	-254.226079	365.7897
Jan 2022	0.7911526	-325.896019	327.4783
Feb 2022	15.1804949	-311.604655	341.9656
Mar 2022	95.9995116	-230.786229	422.7853
Apr 2022	150.9360251	-175.849719	477.7218
May 2022	175.8602606	-150.925483	502.6460
Jun 2022	249.9077191	-76.878025	576.6935
Jul 2022	474.8674113	148.081667	801.6532
Aug 2022	381.9705346	55.184791	708.7563
Sep 2022	207.3805568	-119.405187	534.1663
Oct 2022	196.4037309	-130.382013	523.1895
Nov 2022	59.7622239	-267.023520	386.5480
Dec 2022	59.0829451	-267.702799	385.8687

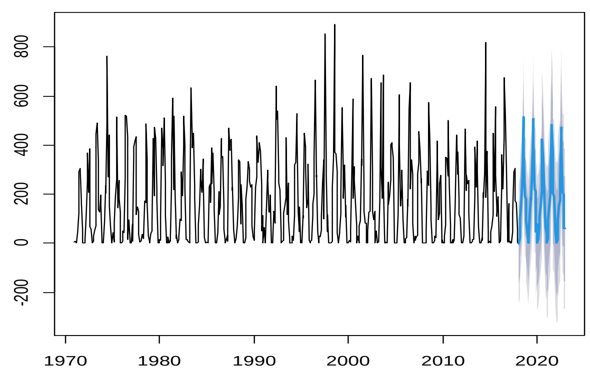


Figure-4: Forecast from SARIMA (1,0,0)(2,1,0)[12]

6. CONCLUSION

In this study, the data was analyzedwith the help of R programming that is now one of greatest necessitated software in the field of data science and statistics. The stationary condition of the data series are observed by ACF and PACF plots and then checked using the statistic such as Augmented Dickey-Fuller Test (ADF). The ADF test confirms that the monthly rainfall is stationary because the p-value of 0.01 is less than 0.05.We have applied auto arima to find and check the best model using R. Based on the auto arima the best fitted model has been found SARIMA (1,0,0)(2,0,0)[12] with minimum Akaike information criterion (AIC) of 6901,Corrected Akaike Information Criterion (AIC_C) of 6901.07 and Schwarz Bayesian Information Criterion (BIC) of 6918.25.The results are useful for predicting the expected rainfallin Cumilla in the next five years and also provide information that would be helpful for decision makers in conveyingplans to mitigate the problems of flood and drought in Cumilla. The rainfall forecast is also important for future strategies regarding agriculture, and to commodity merchants within the stock market, and hence the effect on the economy of country.

REFERENCES

1. Akaike H. (1978) A Bayesian analysis of the minimum AIC procedure.
2. Box G.E.P. and Jenkins G.M. (1976). Time series Analysis, Forecasting and Control. San Francisco: Holden-Day.
3. Box, George E. P. (2015). *Time Series Analysis: Forecasting and Control*. WILEY. ISBN 978-1-118-67502-1.
4. Hyndman, Rob J; Athanasopoulos, George. *8.9 Seasonal ARIMA models*. *Forecasting: principles and practice*. Texts. Retrieved 19 May 2015.
5. http://hikersbay.com/climate/bangladesh/comilla?lang=en.
6. Ishtiak Mahmud and et.al (2017), "Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method", *Environmental Engineering Research*; 22(2): 162-168.
7. J. C. Ramesh Reddy and et.al (2017), Study Forecasting of Monthly Mean Rainfall in Coastal Andhra, *International Journal of Statistics and Applications*, 7(4): 197-204.
8. J. C. Paul and et.al (2013), Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh: A Case Study on Square Pharmaceutical Ltd.; *Global Journal of Management and Business Research Finance* ,Volume 13 Issue 3 Version 1.0,15-25.
9. Ljung G. M. and Box G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. Biometrika. 65: 297.

10. Md. Mahsin and et.al (2012), “Modeling Rainfall in Dhaka Division of Bangladesh using Time Series Analysis”*Journal of Mathematical Modelling and Application* , Vol. 1, No.5, 67-73.
11. *“Notation for ARIMA models” Time Series Forecasting System*. SAS Institute. Retrieved 19 May 2015.
12. Samuel Olorunfemi Adams and et.al (2020), Modeling and Forecasting Seasonal Behavior of Rainfall in Abuja, Nigeria; A SARIMA Approach American Journal of Mathematics and Statistics, 10(1): 10-19.
13. Samuel Olorunfemi Adams and et.al (2019), “Seasonal Autoregressive Integrated Moving Average (SARIMA) Model for the Analysis of Frequency of Monthly Rainfall in Osun State, Nigeria”, *Physical Science International Journal, 22(4): 1-9.*
14. www.weatheronline.co.uk › reports › climate › Bangladesh.