IJSAR, 5(5), 2018; 06-14

International Journal of Sciences & Applied Research

www.ijsar.in

Using SARIMA Approach to Modeling and Forecasting Monthly Rainfall in Bangladesh

Md. DeluarJahan Moloy, MashfiqulHuq Chowdhury, Md. Binyamin, Somaresh Kumar Mondal*

Department of Statistics, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh.

Corresponding author: *Somaresh Kumar Mondal, Department of Statistics, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh.

Abstract

It is difficult task to predict the trend of precipitation meteorology and environmental sciences. The present study considered the monthly rainfall from 1960 to 2016 obtained from Bangladesh Meteorological Department (BMD).To examine Rainfall in Bangladesh and find a suitable model for forecasting is the main intention of this study with the help of SARIMA Approach.Original data plot shows that it is stationary and then tests Auto-Correlation Function (ACF), Partial Auto-Correlation Function (PACF). After taking first difference original data was transformed to stationary. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test with p-value of 0.1and Augmented Dickey- Fuller (ADF) test with p-value of 0.01 proved the stationarity of Rainfall data. On the basis ofleast Akaike Information Criterion (AIC)value we suggest SARIMA (0,0,0) $(2,1,1)_{12}$ model is the best to predict monthly rainfall data. After diagnostic checking on SARIMA (0,0,0) $(2,1,1)_{12}$ model, the Auto Regressive parameter was found to be statistically insignificant and SARIMA (0,0,0) $(2,1,1)_{12}$ model that best fit and was used to forecasting 120 months (January2017- December 2022) seasonal Rainfall in Bangladesh.

Keywords: Rainfall, ADF, SARIMA, Seasonality, Meteorology

Introduction

Climate change such as Temperature, Humidity and Rainfall are the main reason for many countries in the world one of the biggest threats for environmental to food production, forest biodiversity and livelihoods [1]. Specially, Climate change is more affected in developing country like Bangladesh [2] [3].Rainfall is natural climatic phenomena whose prediction is challenging and demanding. Worldwide numerous attempts were made to predict its behavioral pattern using various techniques [4].

Autoregressive Integrated Moving Average Model (ARIMA) is a significant statistical model which is proposed by Box and Jenkins (1970s) [5].ARIMA is a widely applicable for non-stationary time series to clearly identifiable trends [5]. Some periodical time series, ARIMA model is mainly used which namely non-seasonal ARIMA(p, d, q) model.The periodicity of periodical time series is normally due to seasonal changes (including monthly, quarterly and degree of weeks change),that why seasonal ARIMA(P,D,Q) model to be applied[6]. If the time period is equals to 12, SARIMA(p, d, q)(P, D, Q)12 model is particularly applicable [7]. In recent time,some researchers who were modeled rainfall drought on the basic of SARIMA model [8].For example, A. C. Akpanta*et al*.(2015) fit a SARIMA (0,0,0) $(1,1,1)_{12}$ model to forecast monthly rainfall in Nigeria which SARIMA mode was followed by the three conventional steps of identification, estimation and diagnostic checking procedures established by Box and Jenkins [8].

In this present study, to mainly used a seasonal ARIMA model and then forecasting in next 60 months.Specifically, in a SARIMA model, identified the parameters P, D, Q and p, d, q and also built up ACF and PACF plot, KPSS and ADF test, and selected the best SARIMA model also used AIC value.Other related methodologies were also used in the study.

Methodology

Seasonal Autoregressive Integrated Moving Average (SARIMA)

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is

ARIMA $(p,d,q)\times(P,D,Q)_S$ … … … … (1)

with p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

Without differencing operations, the model could be written more formally as

$$\Phi(B^S)\varphi(B)(x_t-\mu)=\Theta(B^S)\theta(B)w_t$$

The non-seasonal components are:

AR: $\varphi(B)=1-\varphi_1 B-\dots\dots-\varphi_p B^P$

MA: $\theta(B)=1+\theta_1 B+\dots\dots+\theta_q B^q$

The seasonal components are:

Seasonal AR: $\Phi(B^S)=1-\Phi_1 B^S-\dots\dots-\Phi_S B^{PS}$

Seasonal MA: $\Theta(B^S)=1-\Theta_1 B^S-\dots\dots\Theta_P B^{PS}$

Note that on the left side of equation (1) the seasonal and non-seasonal AR components multiply each other, and on the right side of equation (1) the seasonal and non-seasonal MA components multiply each other.

ARIMA (0,0,0) $(2,1,1)_{12}$

The model includes that no non-seasonal AR(0) term, a seasonal AR(2) term, which indicates $(1-\Theta_1 B^{12}-\Theta_2 B^{12})$, no non-seasonal differencing($d=0$), seasonal differencing $(D=1)$ that is $(1-B^{12})$, no seasonal moving average i.e. MA(0) terms, seasonal moving average MA(1) that means $(1+\Theta_1 B^{12})$ and seasonal period is S = 12.

The model is $(1-\Theta_1 B^{12}-\Theta_2 B^{12})(1-B^{12})(x_t-\mu)=(1+\Theta_1 B^{12})w_t$

When we multiply the two polynomials on the right side, we get

$$(x_t-\mu)=(1+\theta_1 B+\Theta_1 B^{12}+\theta_1\Theta_1 B^{13})w_t$$
$$=w_t+\theta_1 w_{t-1}+\Theta_1 w_{t-12}+\theta_1\Theta_1 w_{t-13}$$

Thus the model has MA terms at lags 1, 12, and 13.

Results and Discussions

Identification of a Seasonal ARIMA Model

The data on the frequency of monthly rainfall in Bangladesh January, 1960 to December, 2015 and was obtained from the Bangladesh Meteorological Department (BMD). R Statistical Software [9] has been used to perform analysis. A plot of monthly average rainfall is plotted in Figure 1 (Top Panel) which is only conducted by raw data, to evaluate its stability. It is clearly delimitated that time series are stationary [10].

Unit Root test

ADF (Augmented Dickey-Fuller) and KPSS tests are used to identify the order of non-seasonal integration of the series. The ADF test checks the null hypothesis of unit root against the alternative of stationarity for the data generating process. The KPSS test checks the null hypothesis of stationarity against the alternative of a unit root for the data generating process.The ADF and KPSS tests value are in Table 1. At the 5% significant level, the ADF test rejects the null hypothesis of unit root and KPSS test does not rejects the null hypothesis of stationarity. Therefore conclusively the time series does not required non seasonal differencing [11].

The ACF and PACF plots of the original data, as shown in Figure 1, show that the temperature data is not stationary. To make the data stationary, seasonal first difference, $D = 1$, of the original data have been used. The ACF and PACF plots in Fig. 2 shows that the differenced and de-seasonalized Rainfall data are almost stable which support the assumption that the series is stationary in both the mean and the variance after having 1st order non seasonal difference. Therefore, an ARIMA (p, 0, q) (P, 1, Q) 12 model could be identified for the differenced and de-seasonalized Rainfall data [12].

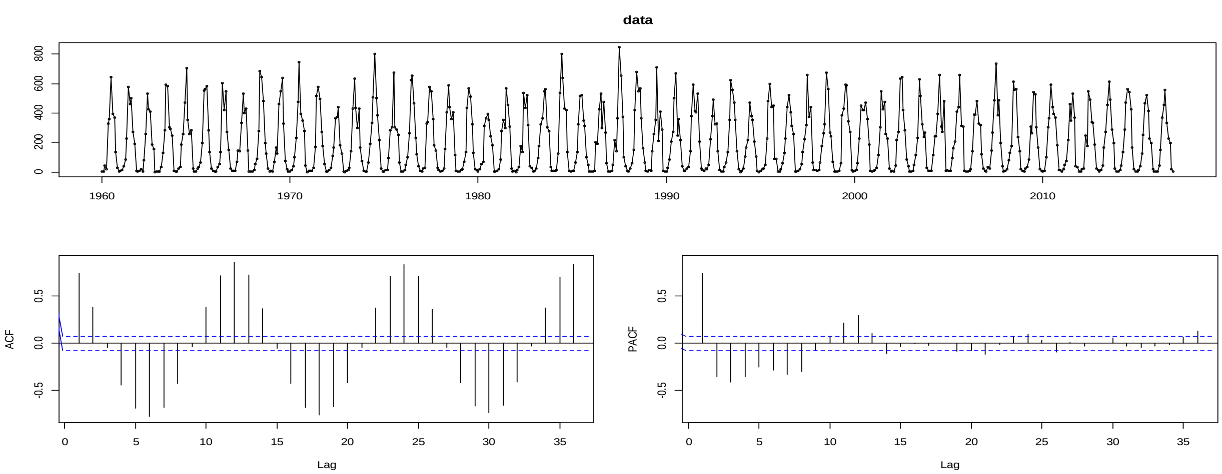

Figure 1: Observed Time Series Plot (Top Panel), ACF Plot (Centre Panel), PACF Plot (Bottom Panel).

Table 1: Summary results for ADF and KPSS tests.

Test	t-statistics	p- value
ADF	-19.182	0.01
KPSS	0.017029	0.1

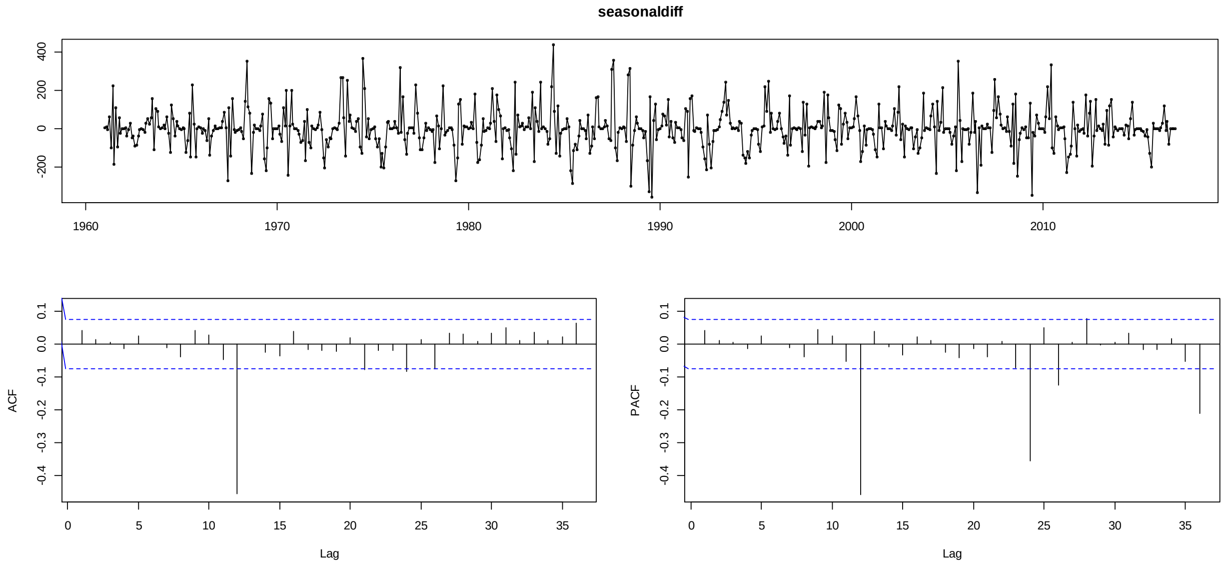

Figure 2: Plot of ACF and PACF Plot of the first seasonal difference data.

Table 2: AIC and BIC values of the fitted models.

Models	**AIC**
SARIMA(0,0,0)(0,1,1)$_{12}$	7713.535
SARIMA(0,0,0)(1,0,1)$_{12}$	7878.218
SARIMA(0,0,0)(1,1,0)$_{12}$	7983.916
SARIMA(0,0,1)(0,1,1)$_{12}$	7714.971
SARIMA(0,0,1)(1,0,1)$_{12}$	7878.917
SARIMA(0,0,1)(1,1,0)$_{12}$	7984.987
SARIMA(0,0,0)(2,1,1)$_{12}$	**7712.506**
SARIMA(0,0,0)(0,1,2)$_{12}$	7715.468
SARIMA(0,0,0)(2,1,0)$_{12}$	7884.006
SARIMA(0,0,1)(2,1,1)$_{12}$	7714.032
SARIMA(0,0,1)(0,1,2)$_{12}$	7716.917
SARIMA(0,0,1)(2,1,0)$_{12}$	7885.105
SARIMA(1,0,2)(2,1,1)$_{12}$	7717.999
SARIMA(1,0,2)(1,1,1)$_{12}$	7720.997
SARIMA(1,0,2)(1,1,0)$_{12}$	7988.785
SARIMA(1,0,1)(2,1,1)$_{12}$	7716.035
SARIMA(1,0,1)(1,1,2)$_{12}$	7718.961
SARIMA(1,0,1)(1,1,0)$_{12}$	7986.744

SARIMA (0,0,0) (2,1,1)$_{12}$ model shows least AIC values than the other model. Now this study checked the diagnosis checking of the SARIMA (0,0,0) (2,1,1)$_{12}$ model.

SARIMA (0,0,0) (2,1,1)$_{12}$, includes a non-seasonal AR (autoregressive) and a seasonal MA (moving average). To test the significance of the parameters, the coefficient of their estimated value and corresponding *p* values are given in the following table.

Table 3. The significance test of the parameter.

Parameter	Estimate	Std. Error	P-Values	Decision
sar1	-0.0167	0.0408	0.01	highly significant
sar2	-0.0889	0.0403	0.00	highly significant
sma1	-0.9831	0.0421	0.00	highly significant

Note: The above table shows that all the parameters are significant.

The diagnostics for the model SARIMA(1, 0, 1)(0, 1, 1)12 is displayed in Fig. 4. In order to check the residuals are white noise or not, Ljung-Box test has been conducted to check the normality test of the residuals. Here, the null and alternative hypotheses are:

H_0 : The residuals are white noise,

H_a : The residuals are not white noise.

The p value of Ljung-box test is found 6776.6 with 30 degrees of freedom, leading to a p-value < 2.2e-16 which indicates that the residuals are white noise [12].

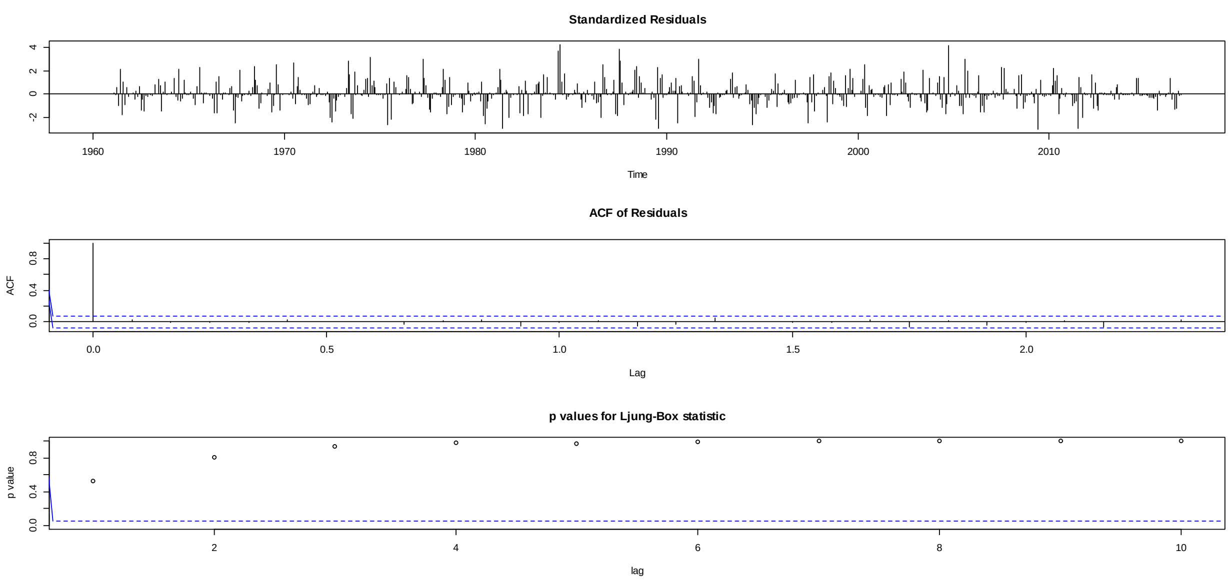

Figure 3: Diagostic Plots for the Fitted ARIMA (0,0,0)(2,1,1)$_{12}$.

The graph of the (Q-Q) plot for the residual data look fairly linear, the normality assumptions of the residuals hold, as shown in Fig. 4[10].

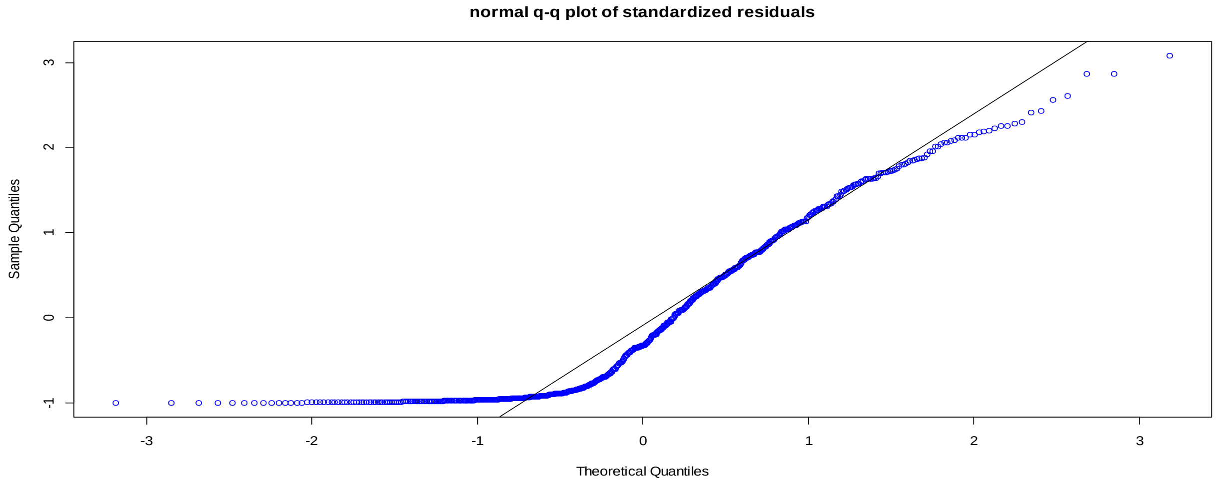

Figure 4: Q-Q Plot of standardized residuals of SARIMA (0,0,0)(2,l,1)$_{12}$ model.

Figure 5 shows a very close agreement between the fitted model and the actual data.

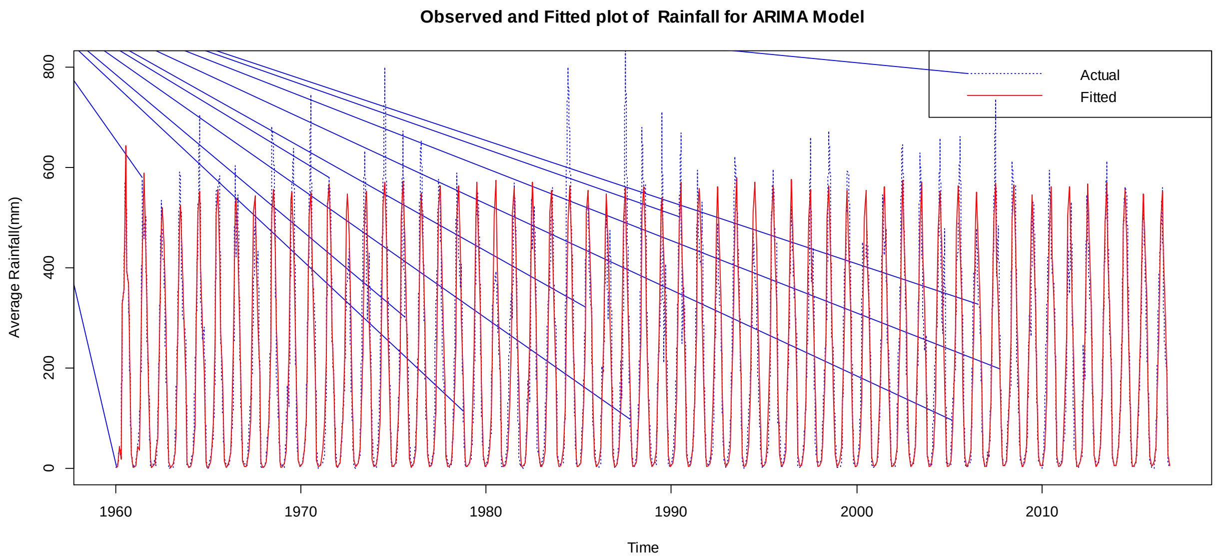

Figure 5: Comparison between observed and fitted plot.

Time Period	**Point Forecasts**	**95% Confidence Intervals**
January-2017	5.211771	(-137.1674590, 147.5910)
February-2017	14.05672	(-128.3225119, 156.4359)
March-2017	46.62335	(-95.7558852, 189.0026)
April-2017	144.4834	(2.1091681, 286.8676)
May-2017	273.6111	(131.2318413, 415.9903)
June-2017	436.6068	(344.2275196, 628.9860)
July-2017	555.069	(412.6897826, 697.4482)
August-2017	439.3532	(296.9739793, 581.7324)
September-2017	332.6363	(190.3075692, 475.0660)
October-2017	174.4233	(32.0441096, 316.8026)
November-2017	26.36636	(-116.0128738, 168.7456)

December-2017	5.34465	(-137.0345802, 147.7239)
January-2018	5.24992	(-137.1160079, 147.6158)
February-2018	14.81917	(-127.5467626, 157.1851)
March-2018	45.95374	(-96.4121854, 188.3197)
April-2018	142.0557	(-0.3102160, 284.4216)
May-2018	264.5736	(122.2126670, 406.9445)
June-2018	436.9522	(344.5863106, 629.3182)
July-2018	551.576	(409.2100315, 693.9419)
August-2018	444.7938	(302.4328423, 587.1647)
September-2018	330.9245	(188.5585566, 473.2904)
October-2018	174.7043	(32.3383834, 317.0702)
November-2018	26.19314	(-116.1677899, 168.564)1
December-2018	5.200924	(-137.1650039, 147.5669)
January-2019	5.020458	(-137.6447936, 147.6857)
February-2019	13.61737	(-129.0473858, 156.2831)
March-2019	45.67566	(-96.9895960, 188.3409)
April-2019	142.796144	(0.1308926, 285.4614)
May-2019	273.4167	(130.7514211, 416.0819)
June-2019	484.2344	(341.5696068, 626.9001)
July-2019	552.0013	(409.3366297, 694.6671)
August-2019	435.3766	(292.7113903, 578.0419)
September-2019	321.4545	(178.7892178, 464.1197)
October-2019	176.6462	(33.9809255, 319.3114)
November-2019	25.27913	(-117.3861204, 167.9444)
December-2019	4.940651	(-137.7246007, 147.6059)
January-2020	5.020393	(-137.6696095, 147.7114)
February-2020	13.57017	(-129.1203418, 156.2607)
March-2020	45.7398	(-96.9507087, 188.4303)
April-2020	143	(0.3094483, 285.6905)
May-2020	274.0718	(131.3812762, 416.7623)
June-2020	484.2495	(341.5590084, 626.9400)
July-2020	552.3052	(409.6146605, 694.9957)
August-2020	435.05	(292.3595089, 577.7405)
September-2020	321.7691	(179.0786350, 464.4597)
October-2020	176.5338	(33.8982885, 319.2793)
November-2020	25.30942	(-117.3810885, 167.9999)
December-2020	4.957767	(-137.7327407, 147.6483)
January-2021	5.041281	(-137.6922559, 147.7748)
February-2021	13.67771	(-129.0558262, 156.4112)
March-2021	45.76344	(-96.9700976, 188.4970)
April-2021	142.9308	(0.1972215, 285.6643)
May-2021	273.2755	(130.5419512, 416.0090)
June-2021	484.4907	(341.7572036, 627.2243)
July-2021	552.2623	(409.5287209, 694.9958)
August-2021	435.8927	(293.1591916, 578.6263)
September-2021	322.6054	(179.8718685, 465.3389)
October-2021	176.4172	(33.6836609, 319.1507)

November-2021	25.39058	(-117.3429595, 168.1241)
December-2021	4.980609	(-137.7529276, 147.7141)
January-2022	5.040901	(-137.7080689, 147.7899)
February-2022	13.68015	(-129.0688163, 156.4291)
March-2022	45.75735	(-96.9916257, 188.5063)
April-2022	142.9138	(0.1648320, 285.6628)
May-2022	273.2306	(130.4815954, 415.9795)
June-2022	484.4854	(341.7364410, 627.2344)
July-2022	552.236	(409.4870531, 694.9850)
August-2022	435.9077	(293.1587158, 578.6567)
September-2022	322.5635	(179.8145139, 465.3125)
October-2022	176.4252	(33.6761905, 319.1741)
November-2022	25.38653	(-117.3624392, 168.1355)
December-2022	4.978707	(-137.7702634,147.7277)

The forecast values with the 95% confidence intervals are shown in figure 6, where the two lines indicate the forecast values for next 72 months and the green lines indicate the 95% confidence intervals for those forecasts.

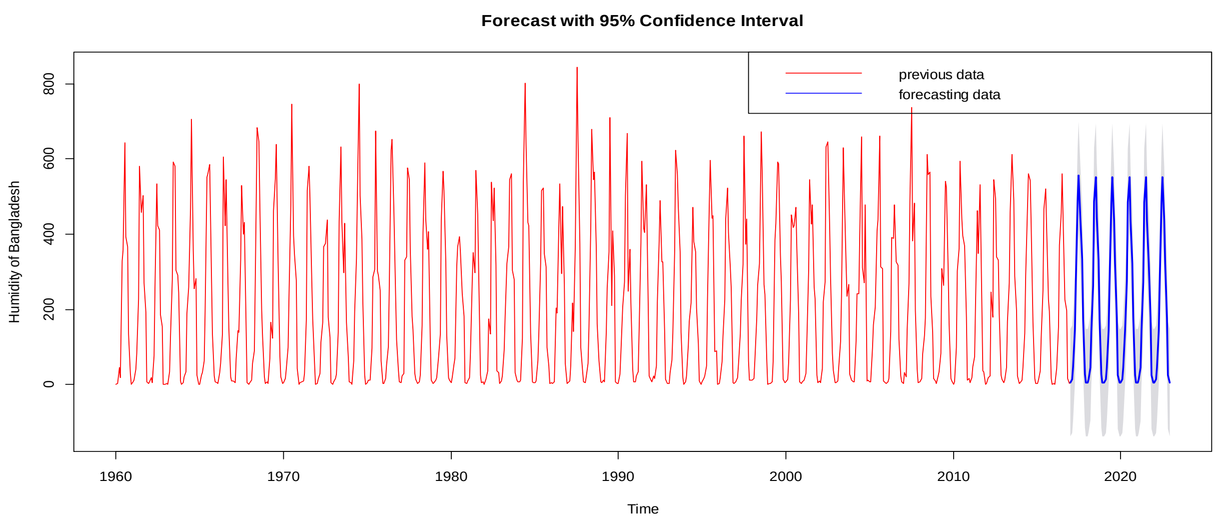

Figure 6: Forecasts with 95% confidence interval using ARIMA (0,0,0)(2,1,1)$_{12}$ model.

Conclusions

This paper has considered the seasonal autoregressive integrated moving average (SARIMA) modeling of Bangladesh monthly average Rainfall. In this study, an ARIMA model that incorporates the seasonality of time series was presented. Using the time series of monthly Rainfall precipitation in Bangladesh, we build a seasonal SARIMA (0, 0, 0) (2, 1, 1)$_{12}$. It was found that the model fitted the data well and the stochastic seasonal fluctuation was successfully modeled except some extreme values. The predictions based on this model indicate that the precipitation in the next three years will decrease.

References

[1] E.S. Chung, K. Park and K.S. Lee. 2011. The relative impacts of climate change and urbanization on the hydrological

response of a Korean urban watershed. Hydrological Processes, 25:544-560.

[2] D. Machiwal and M. K. Jha. 2006. Time Series Analysis of Hydrologic Data For Water Resources Planning and Management: A Review. J. Hydrol. Hydromech., 54(3): 237–257

[3] S.A. Shamsnia, N. Shahidi, A. Liaghat, A.Sarraf and S.F. Vahdat. 2011. Modeling ofWeather Parameters (Temperature, Rainfall AndHumidity) Using Stochastic Methods. Internat. Conference on Environment and Industrial Innovation, IPCBEE, Singapore, 282-285.

[4] M. Nirmala, S. M. Sundaram (2010), a seasonal arima model for forecasting monthly rainfall in tamilnadu, National Journal on Advances in Building Sciences and Mechanics, Vol. 1, No.2,

[5] Stoffer, D.S. and R.H. Dhumway, 2010. Time Series Analysis and its Application. 3rd Edn., Springer, New York, ISBN-10: 1441978658, pp: 596.

[6] He, S.Y., 2004. Applied Time Series Analysis. 1st Edn., Peking University Press, Beijing.

[7] Wang, Y., 2008. Applied Time Series Analysis. 1st Edn., China Renmin University Press, Beijing.

[8] A. C. Akpanta1, I. E. Okorie, N. N. Okoye,(2015), SARIMA Modelling of the Frequency of Monthly Rainfallin Umuahia, AbiaAtate of Nigeria, American Journal of Mathematics and Statistics 2015, 5(2): 82-87

[9] R Development Core Team R: A language and environment for statistical computing // R Foundation for Statistical Computing. - Vienna, Austria: [s.n.], 2014. - 3-900051-07-0.

[10] Hisham Mohamed Hassan Ali and Tariq Mahgoub Mohamed (2014).Rainfall Drought Simulating UsingStochastic SARIMA Mo dels for GadarefRegion, Sudan,MPRA Pap er No. 61259, p osted 11. February 2015 06:44 UTC.

[11] Yakubu Musa (2014), Modeling an Average Monthly Temperature of SokotoMetropolis Using Short Term Memory Models, *International Journal of Academic Research in Business and Social Sciences, Vol. 4, No. 7*

[12] Tanusree Deb Roy and Kishore Das K.(2012).Time Series Analysis OfDibrugarh Air Temperature, Journal Of Atmospheric And Earth Environment, VOLUME 1 NUMBER 1 (2012) 30-34.