[Islam et. al., Vol.6 (Iss.4): April, 2018]
(Received: Jan 03, 2018 - Accepted: Apr 16, 2018)
ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.29121/granthaalayah.v6.i4.2018.1475

INTERNATIONAL JOURNAL OF RESEARCH – GRANTHAALAYAH
A knowledge Repository

Science

MODELING AND FORECASTING HUMIDITY IN BANGLADESH: BOX-JENKINS APPROACH

Mashfiqul Huq Chowdhury [1], Somaresh Kumar Mondal [1], Jobaidul Islam [*2]
[1] Department of Statistics, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh
[*2] Department of Physics, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh

Abstract

Humidity (atmospheric moisture) is an important atmospheric component and has significant influence on plant growth and development. The rate of growth and the form that a plant attains is controlled by humidity. The present study is an attempt to analyze the seasonal humidity's of Bangladesh by employing appropriate statistical techniques. The main objective of this study is to examine humidity over time in Bangladesh and find a suitable model for forecasting. This study utilizes humidity data from Bangladesh Meteorological Department (BMD), recorded at 6 divisional meteorological stations for the period of 1976 to 2015. This study found that annual average humidity of Bangladesh is 78.88%. Initially data set is checked for whether it is stationary or not through Augmented Dickey Fuller test. Data was found non-stationary but it was transformed to stationary after taking first difference. Then seasonal ARIMA model was built using Box and Jenkins approach. After examining of all diagnostic procedures, ARIMA (2,0,1)(2,1,1)12 model has been identified as an appropriate model for forecasting 60 months (2016-2020) seasonal humidity.

Keywords: Box-Jenkins; Humidity; Ljung-Box Test; Sarima; Shapiro Test.

Cite This Article: Mashfiqul Huq Chowdhury, Somaresh Kumar Mondal, and Jobaidul Islam. (2018). "MODELING AND FORECASTING HUMIDITY IN BANGLADESH: BOX-JENKINS APPROACH." *International Journal of Research - Granthaalayah*, 6(4), 50-60. https://doi.org/10.29121/granthaalayah.v6.i4.2018.1475.

1. Introduction

Humidity (atmospheric moisture) is an important atmospheric component [1]. Humidity level has significant influence on plant growth and development. The rate of growth, the composition and the form that a plant attains are controlled by humidity. The influence of humidity upon plants is apparent by examining the different patterns of natural vegetation that have evolved in different areas of the world and the species that grow and mature in different seasons. Certain vegetation patterns are closely linked with atmospheric moisture levels [2].

 ISSN- 2350-0530(O), ISSN- 2394-3629(P)
(Received: Jan 03, 2018 - Accepted: Apr 16, 2018) DOI: 10.5281/zenodo.1241452

The relative humidity of indoor environments over the range of normal indoor temperatures of 19ºC to 20ºC has both direct and indirect effects on health and comfort. Both very low and high relative humidity may cause some physical discomfort. Extremely, low (below 20%) relative humidity may cause eye irritation and moderate to high levels of humidity have been shown to reduce the severity of asthma. Epidemiological studies suggest that relative humidity and humidification equipment can indirectly affect the incidence of allergies and infectious respiratory diseases. These indirect effects may partially account for the suspected relationship between respiratory infections and nose or throat irritation and relative humidity [3].

A warmer and humid climate leads to an increase of some air pollutants and more frequent extreme weather events. It increases the rates and ranges of transmission of infectious diseases through unclean water and contaminated food, and by affecting vector organisms (such as mosquitoes). Changes in temperature, humidity and other atmospheric elements affects agricultural product in many regions, specially least developed or developing countries like Bangladesh and thus jeopardizing child health and growth and the overall health and functional capacity of adults [4].

Bangladesh is mainly an agricultural country. The agricultural activities of Bangladesh are largely depends on climate. But due to unnatural behavior of atmosphere, the cultivation is often hampered. For the development of agriculture sector and agricultural production, it is very important to extensively study the humidity condition of Bangladesh [5].

Though some analysis on rainfall and temperature are done in different studies but study on humidity yet not been analyzed extensively in the country like Bangladesh. The present study is a modest attempt to analyze the humidity data of Bangladesh by employing appropriate statistical techniques. Since the production of agriculture, human health, ecosystem, biodiversity and many other important factors depends on humidity, it is of immense importance to observe the pattern and variations of humidity in the air of Bangladesh. In this study, the long range behavior of the average humidity of Bangladesh from 1976 to 2015 was studied. Then a suitable seasonal ARIMA model was built to forecast relative humidity in Bangladesh. It is expected that this report will be helpful for understanding the future climate of Bangladesh.

1.1. Data Processing

In regard to data processing this study extensively uses R programming language (version 3.3.2). Several packages of R programming language are used such as“ggplot2”, “tseries”, “timeSeries”, “forecast”, “gridExtra”“TTR” and “reshape2”.

2. Materials and Methods

2.1. Seasonal Autoregressive Integrated Moving Average (SARIMA)

Seasonal ARIMA (SARIMA) is used when the time series exhibits a seasonal variation. Natural phenomena such as temperature, rainfall etc. has strong components corresponding to seasons. Hence, the natural variability of many physical, biological and economic processes tends to match with seasonal fluctuations. Because of this, it is appropriate to introduce autoregressive and

 ISSN- 2350-0530(O), ISSN- 2394-3629(P)
(Received: Jan 03, 2018 - Accepted: Apr 16, 2018) DOI: 10.5281/zenodo.1241452

moving average polynomials that can be identified with seasonal lags [6, 7]. The ARIMA notation can be extended readily to handle seasonal aspects and the general shorthand notation is as follows:

$$\text{ARIMA}\,(p,d,q)(P,D,Q)_S$$

Where,

(p,d,q) Refers to the non-seasonal part (autoregressive parameter, difference parameter and moving average parameter) of the model

(P,D,Q) Refers to the seasonal part (autoregressive parameter, difference parameter and moving average parameter) of the model and s refers to the number of periods per season

2.2. Box-Jenkins Methodology

In econometrics, the Box-Jenkins methodology, named after the statisticians George Box and Gwilym Jenkins, applies autoregressive moving average ARMA or ARIMA models to find the best fit of a time series to past values of this time series, in order to make forecasting. The Box-Jenkins methodology consists of a four-step iterative procedure: tentative identification, estimation, diagnostic checking and forecasting [8].

The first step in developing a Box-Jenkins model is to determine if the time series is stationary and if there is any significant seasonality that needs to be modeled. Stationarity can be assessed from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay. An augmented Dickey Fuller test (ADF) is a test for a unit root in a time series sample. The augmented Dickey-Fuller (ADF) statistic, used in the test, is a negative number. At the model identification stage, the goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For many series, the period is known and a single seasonality term is sufficient. However, it may be helpful to apply a seasonal difference to the data and regenerate the autocorrelation and partial autocorrelation plots. This may help in the model identification of the non-seasonal component of the model. Box-Jenkins forecasting models are based on statistical concepts and principles and are able to model a wide spectrum of time series behavior. The series also needs to be at least weakly stationary [8].

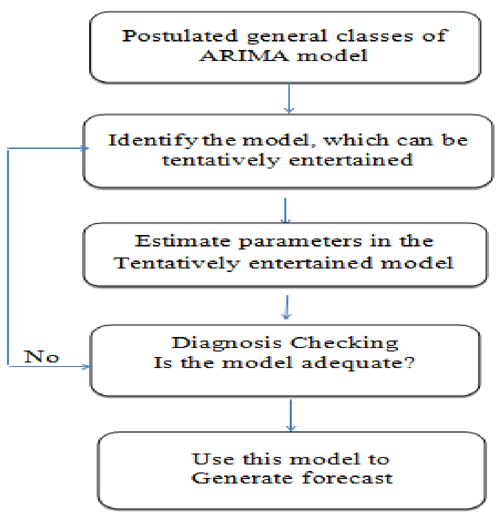

Figure 1: Box-Jenkins methodology for optimal model selection

3. Results and Discussions

[Islam et. al., Vol.6 (Iss.4): April, 2018] ISSN- 2350-0530(O), ISSN- 2394-3629(P)
(Received: Jan 03, 2018 - Accepted: Apr 16, 2018) DOI: 10.5281/zenodo.1241452

3.1. Statistical Analysis

3.1.1. Descriptive Statistics

The results of this study reveal that the annual average humidity 78.88%. Maximum monthly mean relative humidity was found in July (86.67%) and minimum was in April (22%). During this period (1976-2015), Monthly average relative humidity was found highest in Barisal (83.45%) divisions and lowest in Dhaka (74.59%) divisions. The following figures (2 & 3) and table: 1 shows the overall trends of humidity of Bangladesh during 1976-2015.

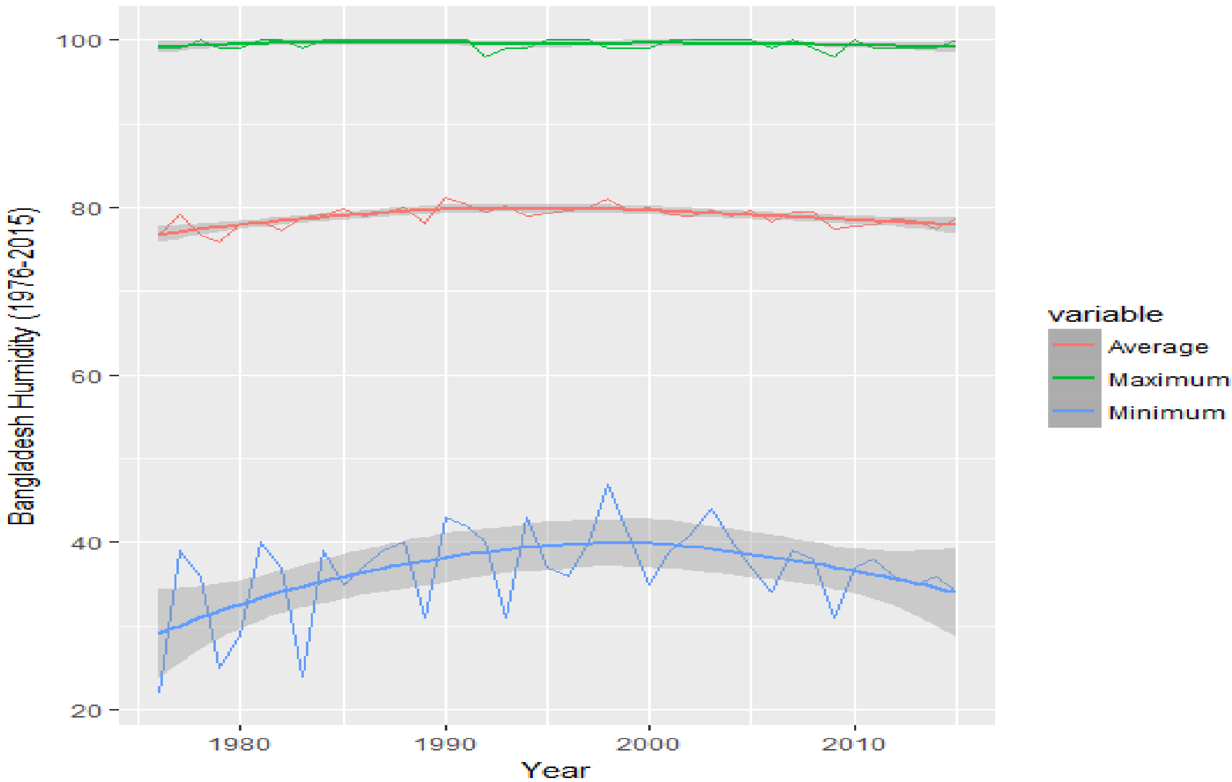

Figure 2: Trends of Bangladesh Humidity during 1976-2015

The following tables present some descriptive statistics on humidity of Bangladesh.

Table 1: Descriptive statistics of humidity according to division

Division	**Mean**	**Maximum**	**Minimum**
Dhaka	74.59438104	99	31
Chittagong	78.72585548	100	36
Khulna	80.29899925	100	36
Rajshahi	77.54298275	100	22
Barisal	83.44677114	100	46
Sylhet	78.80789278	100	38

 ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1241452

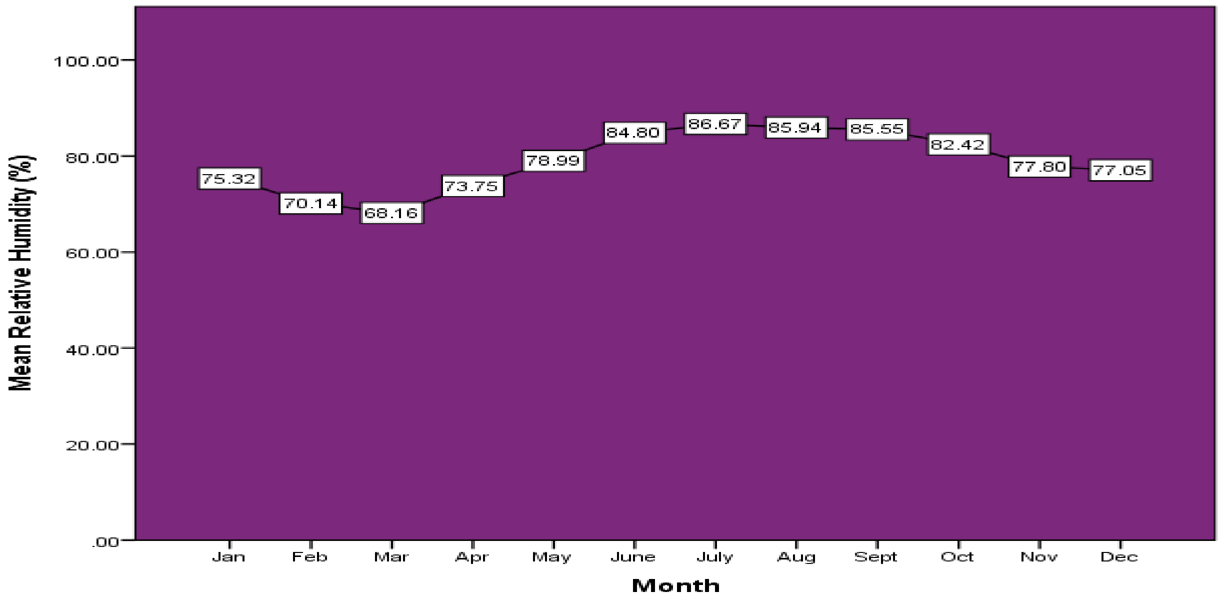

Figure 3: Mean monthly humidity pattern of Bangladesh

3.2. Identification of a Seasonal ARIMA Model

3.2.1. Variance Stability

A visual plot of monthly average humidity is plotted in Figure 4. From this plot it is seen that, there is no prominent trend is present in our data. Moreover it seems that the data are non-stationary in the mean only. So we do not need any transformation of the data to obtain stability in variance [8].

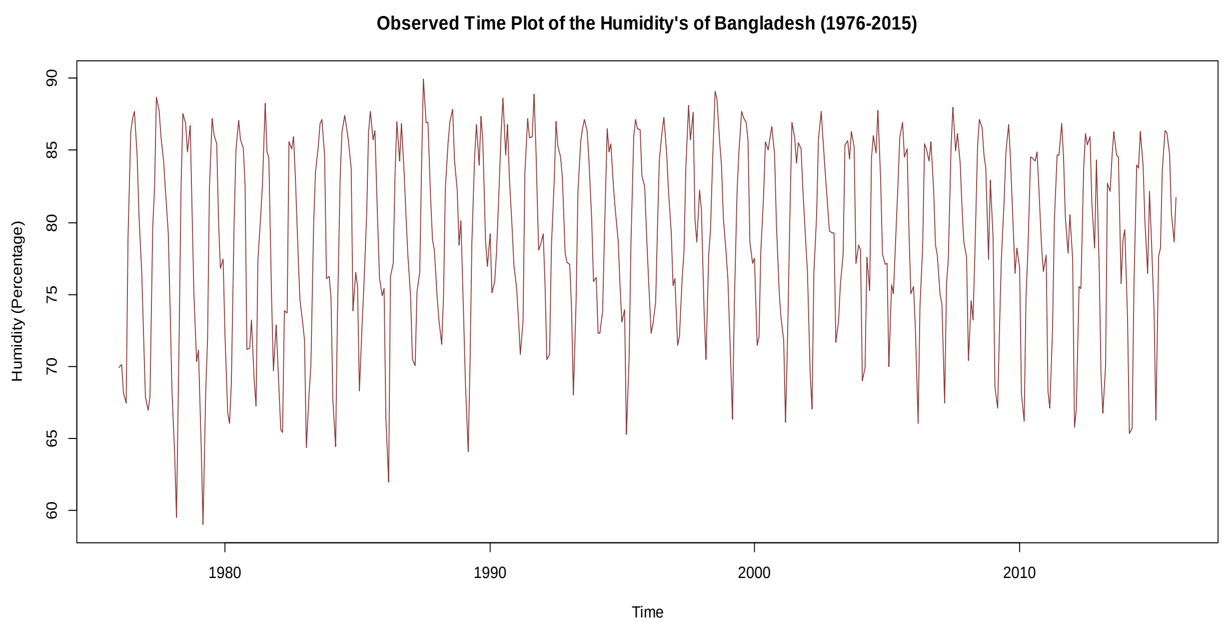

Figure 4: Observed time plot of the mean humidity of Bangladesh

 ISSN- 2350-0530(O), ISSN- 2394-3629(P)
(Received: Jan 03, 2018 - Accepted: Apr 16, 2018) DOI: 10.5281/zenodo.1241452

3.3. Checking the Stationarity

3.3.1. Augment Dickey-Fuller (ADF) Test about Stationarity

To test the stationarity of the non-seasonality parameter $(d = 0)$, the familiar Augmented Dickey-Fuller test was used. The value of the Augmented Dickey-Fuller test statistic is found -13.6with lag order 7. The p-value is calculated as 0.01. It indicates that null hypothesis may be rejected (H_0: Data is non-stationary) and the data set is stationary in mean. So differencing or transformation is not necessary to achieve stationarity.

To test the stationarity of the seasonal parameter $(D = 1)$, again Augmented Dickey-Fuller test was used. The value of the test statistic is found is -7.2with lag order 7. The p-value is found to be 0.01which indicates that the data is stationary in mean. Finally, we can conclude that non-seasonality difference parameter is $(d = 0)$ and seasonality difference parameter $(D = 1)$.

3.3.2. Model Selection

Following table shows the AIC and BIC values for different combinations of non-seasonal (p, q) and seasonal (P, Q) parameters, that is, for different ARIMA (p, 0, q) (P,1, Q)$_{12}$ models.

Table 2: AIC and BIC values of the fitted models

Models	**AIC**	**BIC**
ARIMA(1,0,1)(0,1,1)$_{12}$	2135.105	2151.699
ARIMA(0,0,1)(1,1,1)$_{12}$	2139.583	2156.177
ARIMA(1,0,0)(0,1,1)$_{12}$	2134.706	2147.151
ARIMA(1,0,0)(1,1,0)$_{12}$	2243.183	2255.629
ARIMA(1,0,0)(1,1,1)$_{12}$	2136.062	2152.656
ARIMA(1,0,0)(2,1,0)$_{12}$	2180.920	2197.514
ARIMA(1,0,0)(2,1,1)$_{12}$	2135.313	2156.055
ARIMA(1,0,0)(2,1,2)$_{12}$	2137.018	2161.903
ARIMA(1,0,1)(2,1,1)$_{12}$	2136.535	2161.274
ARIMA(1,0,1)(1,1,0)$_{12}$	2243.413	2260.006
ARIMA(1,0,1)(1,1,1)$_{12}$	2136.710	2157.452
ARIMA(1,0,1)(2,1,1)$_{12}$	2136.535	2161.426
ARIMA(1,0,1)(0,1,1)$_{12}$	2135.105	2151.699
ARIMA(1,0,2)(0,1,1)$_{12}$	**2130.729**	2151.471
ARIMA(1,0,2)(1,1,0)$_{12}$	2244.432	2265.175
ARIMA(1,0,2)(1,1,1)$_{12}$	**2131.785**	2156.576
ARIMA(1,0,2)(2,1,1)$_{12}$	**2130.213**	2159.253

From the above table-2, AIC values of ARIMA (1, 0, 2) (2, 1, 1)$_{12}$ is found lowest than the other model. So, **ARIMA (1, 0, 2) (2, 1, 1)$_{12}$**model is selected.

(Received: Jan 03, 2018 - Accepted: Apr 16, 2018)
ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1241452

3.3.3. Estimation and Diagnostic Checking

ARIMA (1, 0, 2) (2, 1, 1)$_{12}$, includes non-seasonal AR and seasonal MA. The coefficient of this model and corresponding *p* values are given in the following table. All of the parameters are found highly significant.

Table 3: The significance test of the parameter

Parameter	Estimate	Std. Error	P-Values	Decision
ar1	0.9803	0.0192	0.00	Highly Significant
ma1	-0.6807	0.0507	0.00	Highly Significant
ma2	-0.2482	0.0494	0.00	Highly Significant
sar1	-0.1194	0.0679	0.00	Highly Significant
sar2	-0.1178	0.0614	0.00	Highly Significant
sma1	-0.7629	0.0552	0.00	Highly Significant

3.3.4. Shapiro-Wilk Test for Checking Normality Assumption of Residuals

Shapiro-Wilk test is performed to check that the residuals are normally distributed or not. Here, after conducting the test, p-value is found to be 0.41. Thus, the null hypothesis (H_0 : residuals are normally distributed) cannot be rejected. Hence, the residuals can be concluded as normally distributed [9]. This test can be interpreted using normal *Q-Q* plot is given in below.

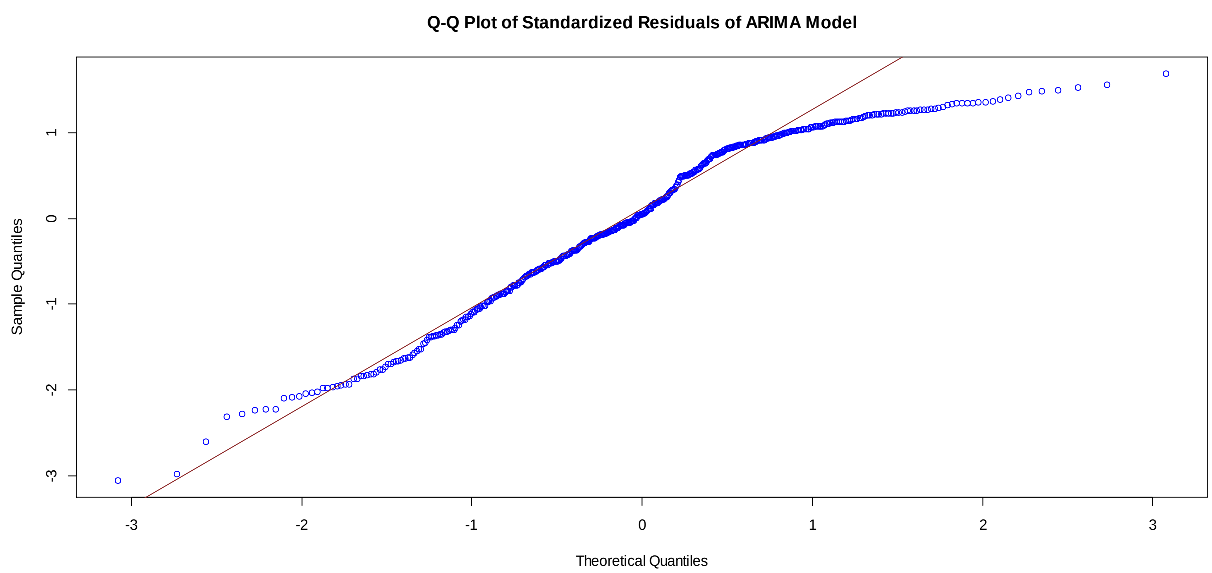

Figure 5: Q-Q Plot of standardized residuals of ARIMA (l, 0, 2) (2, l, 1)$_{12}$ model

3.3.5. Ljung-Box Test for Checking White Noise Assumption

To ensure that the chosen seasonal ARIMA model fits the data well, the assumption of white noise tested by Ljung-box test. Figure- 6 shows the behavior of the residuals left over after fitting the ARIMA (1, 0, 2) (2, 1, 1)$_{12}$ model. The plot of the standardized residuals shows that most of the standardized residuals are within 95 percent limit. The plot of ACF of residuals is given. In both cases, all the spikes are in 95 percent limit and near to zero. In order to check the residuals are

(Received: Jan 03, 2018 - Accepted: Apr 16, 2018)
ISSN- 2350-0530(O), ISSN- 2394-3629(P)
DOI: 10.5281/zenodo.1241452

white noise or not, Ljung-Box test is conducted. The p value of Ljung-box test is found 0.1955 indicates that the residuals are white noise.

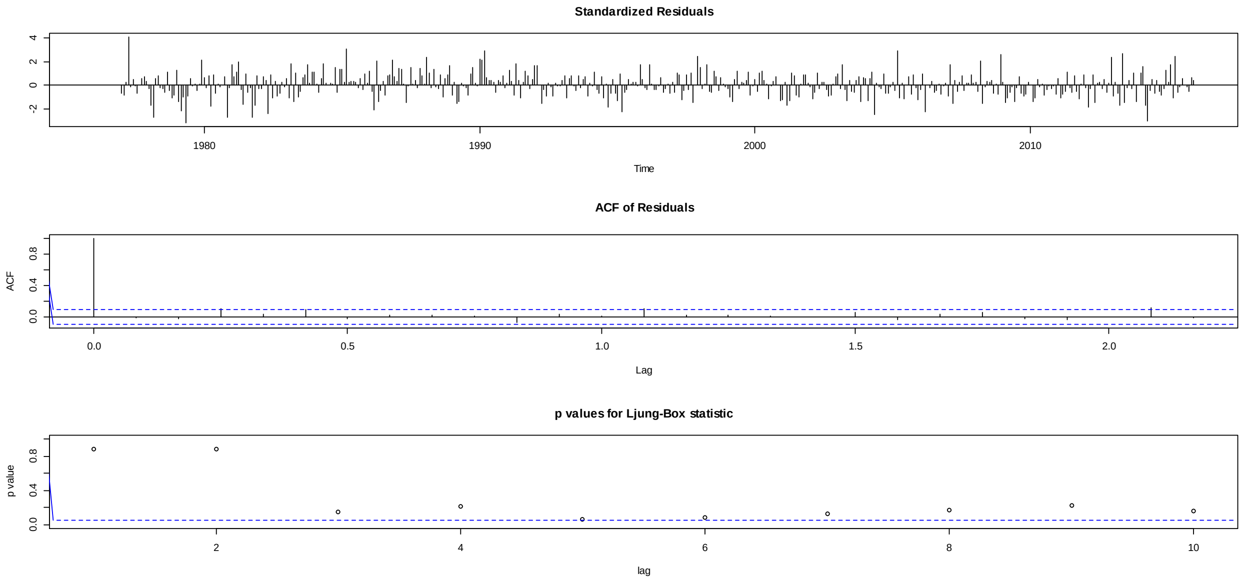

Figure 6: Diagnostic plot of AR1MA(1,0,2)(2,1,1)$_{12}$ model

3.3.6. Actual and Fitted Plot

A plot of the actual values and the fitted values using the model is given in figure- 7. The red line denotes the fitted values and the blue line denotes the actual values. From the plot, it is seen that the model has a much close fit. We can conclude that the actual and fitted values are very close to each other.

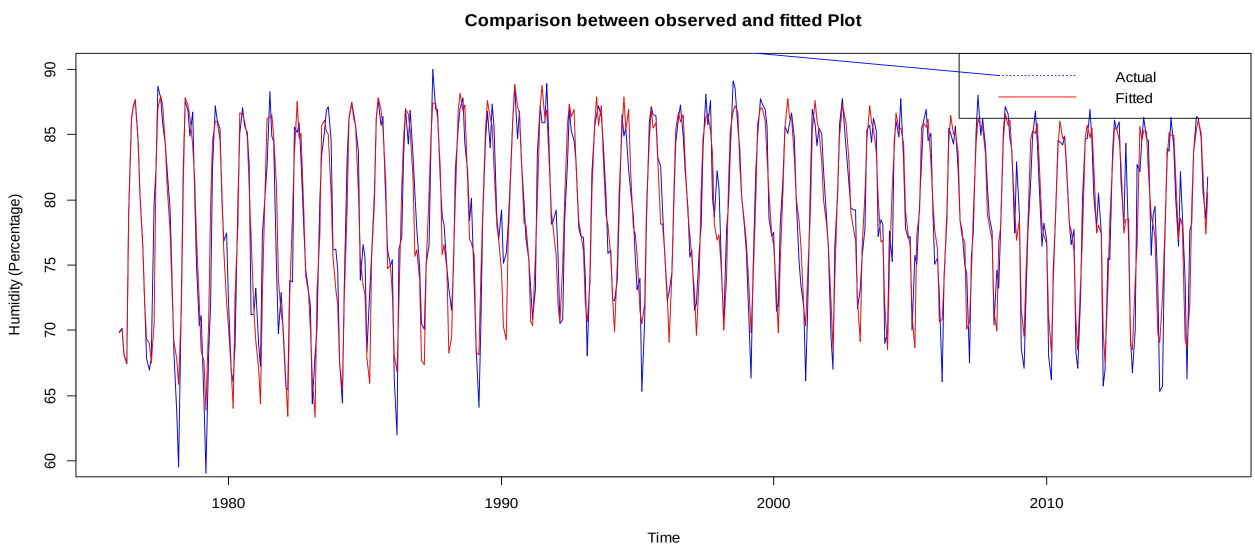

Figure 7: Comparison between observed and fitted plot

All diagnostic check support that AR1MA(1,0,2)(2,1,1)$_{12}$ model not only has the smallest AIC but also satisfies all assumptions of residuals. Now, this model is used for forecasting humidity of Bangladesh.

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
(Received: Jan 03, 2018 - Accepted: Apr 16, 2018)
DOI: 10.5281/zenodo.1241452

3.3.7. Forecasting

The point forecast with 95% confidence interval on average humidity of Bangladesh for the month January, 2016 to December, 2020 by using the selected model is given in Table-4.

Table 4: Forecasts of humidity for next 60 months (2016-2020)

Time Period	Point Forecasts	95% Confidence Intervals
January-2016	77.96663	(73.48441, 82.44886)
February-2016	69.98481	(65.30575, 74.66387)
March-2016	67.45935	(62.77584, 72.14286)
April-2016	73.20095	(68.51317, 77.88874)
May-2016	78.62566	(73.93377, 83.31755)
June-2016	83.78104	(79.08521, 88.47687)
July-2016	85.42041	(80.72079, 90.12002)
August-2016	85.99354	(81.29030, 90.69678)
September-2016	85.00947	(80.30274, 89.71621)
October-2016	82.21782	(77.50773, 86.92790)
November-2016	77.67652	(72.96322, 82.38982)
December-2016	80.27103	(75.55463, 84.98742)
January-2017	77.91647	(73.14910, 82.68384)
February-2017	70.38734	(65.60909, 75.16559)
March-2017	67.20783	(62.42600, 71.98965)
April-2017	72.32555	(67.54029, 77.11081)
May-2017	78.26389	(73.47534, 83.05245)
June-2017	83.80098	(79.00926, 88.59270)
July-2017	85.22546	(80.43069, 90.02022)
August-2017	86.03330	(81.23561, 90.83098)
September-2017	84.88755	(80.08706, 89.68804)
October-2017	81.96606	(77.16287, 86.76924)
November-2017	77.53297	(72.72720, 82.33874)
December-2017	80.49904	(75.69078, 85.30730)
January-2018	78.04027	(73.19176, 82.88879)
February-2018	70.79173	(65.93455, 75.64891)
March-2018	67.09872	(62.23867, 71.95877)
April-2018	72.95336	(68.09054, 77.81617)
May-2018	78.26041	(73.39494, 83.12587)
June-2018	83.78364	(78.91562, 88.65166)
July-2018	85.36045	(80.48999, 90.23092)
August-2018	86.05628	(81.18347, 90.92909)
September-2018	84.86841	(79.99334, 89.74348)
October-2018	81.80802	(76.93078, 86.68526)
November-2018	77.65945	(72.78013, 82.53878)
December-2018	80.64412	(75.76279, 85.52544)
January-2019	78.03232	(73.03263, 83.03200)
February-2019	70.69694	(65.67987, 75.7140)
March-2019	67.14224	(62.12213, 72.16235)
April-2019	72.98240	(67.95937, 78.00543)
May-2019	78.30428	(73.27844, 83.33011)
June-2019	83.78418	(78.75565, 88.81271)
July-2019	85.36811	(80.33700, 90.39923)
August-2019	86.04965	(81.01604, 91.08325)

September-2019	84.88583	(79.84984, 89.92182)
October-2019	81.85730	(76.81901, 86.89558)
November-2019	77.66201	(72.62152, 82.70250)
December-2019	80.60067	(75.55807, 85.64328)
January-2020	78.01940	(72.87110, 83.16770)
February-2020	70.66132	(65.49685, 75.82580)
March-2020	67.15059	(61.98304, 72.31814)
April-2020	72.90566	(67.73515, 78.07617)
May-2020	78.30011	(73.12676, 83.47346)
June-2020	83.78681	(78.61073, 88.96289)
July-2020	85.35193	(80.17323, 90.53063)
August-2020	86.04835	(80.86714, 91.22957)
September-2020	84.88662	(79.70298, 90.07025)
October-2020	81.87063	(76.68467, 87.05659)
November-2020	77.64739	(72.45920, 82.83559)
December-2020	80.58934	(75.39901, 85.77968)

Actual and Fitted Forecasting Plot

The plot of forecasts with the 95% confidence intervals is shown in figure-8.

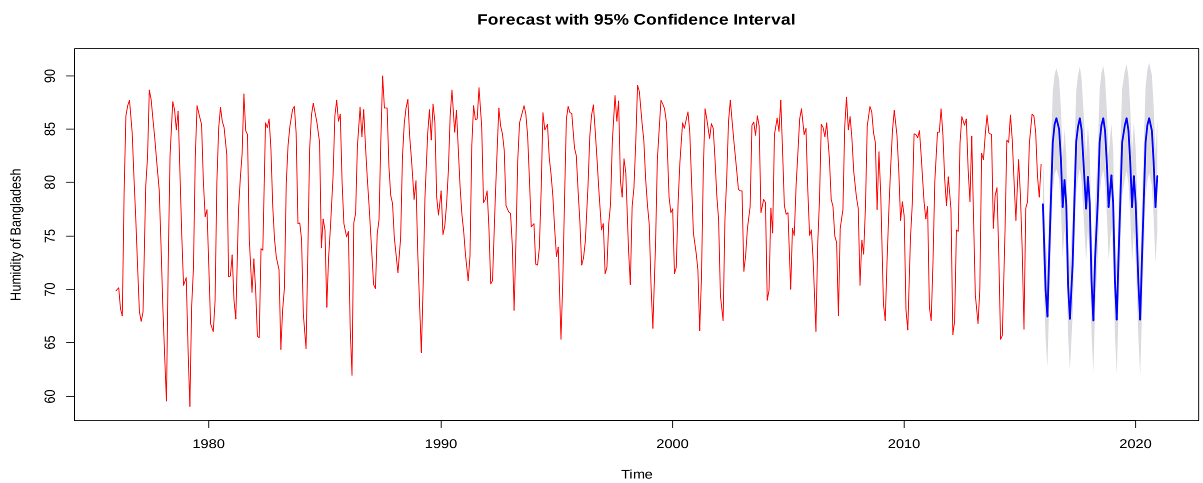

Figure 8: Forecasts with 95% confidence interval using ARIMA (1, 0, 2) (2, 1, 1)$_{12}$ model

4. Conclusions and Recommendations

The main objective of this study is to modeling and forecasting monthly humidity using SARIMA model. To apply SARIMA model, we estimate 6 parameters. Initially we check stationarity of our humidity data set and get an idea about non-seasonal and seasonal parameters of seasonal ARIMA model using ACF, PACF and Augmented Dickey-Fuller test. Since the observed data does not follow any trend but seasonality. Hence, seasonal differencing was made to make the time series seasonally stationary. Using the model selection criterion, AIC, ARIMA (1,0,2)(2,1,1)$_{12}$ model is found to be the best model for humidity data set. The parameters of this model have been estimated using the maximum likelihood method. Then we test the significance of all the parameters of this

 ISSN- 2350-0530(O), ISSN- 2394-3629(P)
(Received: Jan 03, 2018 - Accepted: Apr 16, 2018) DOI: 10.5281/zenodo.1241452

model and have been found to be significant. The assumption is made on normality and independence of the residuals has been checked using different plots and test. The plot comparing actual values and fitted values using the model shows much close fit. Then the model has been used for forecasting humidity from June 2017 to December 2035. From this study we have found that annual average humidity of Bangladesh is 78.88%. The forecasting plot also shows that humidity fluctuation trend will be uniform, steady and regular in the upcoming year.

We hope that the findings of this study can help the policy makers to a great extent in an environmental issue and also helps to set up a fruitful policy for maintaining the ecological balance of Bangladesh.

5. Limitations of the Study

There are 34 meteorological stations in Bangladesh. But due to time and cost constraints data from 6 divisional meteorological stations has been considered for all relevant analysis in this study.

References

[1] Basak J. K., Mahmud R. A., Dey N. C., (2013), Climate Change in Bangladesh: A Historical Analysis of Temperature and Rainfall Data, Journal of Environment, 2, 41-46.
[2] Tibbitts T. W., Kozlowski T. T., (1979), Controlled Environment Guidelines for Plant Research: Proceedings of the Controlled Environments Working Conference Held at Madison, Wisconsin, March 12-14.
[3] DEW-DROP, (2016), A Scientific Journal of Meteorology and Geo-Physics, Bangladesh Meteorology Department, Volume 2, Number 1.
[4] Haque, R., Rahman, S. H., Abedin, M. Z. (2011). Recent Climate Change Trend Analysis and Future Prediction at Satkhira District, Bangladesh. IOP Publishing.
[5] Haque, R., Rahman, S. H. Karrnakar, S., Hussain, S. G., (2011). Trend analysis of Climate Change and Investigation on its Probable Impacts on Rice Product at Satkhira, Bangladesh, Pakistan Journal of Meteorology, 6.
[6] Brockwell, P. J. and Davis, R. A. (1996) Introduction to Time Series and Forecasting, 2nd Edition, Springer, New York.
[7] Gallop, C., Tseand, C., and Zhao, J. (2012) A Seasonal Autoregressive Model of Vancouver Bicycle Traffic Using Weather Variables, TRB Annual Meeting
[8] Gujarati D. N., Poter D. C., Gunasekar S, (2012), Basic Econometrics, 5th edition, McGraw Hill Book companies, inc. New York
[9] José A. Alva V., Estrada E. G., (2009), A Generalization of Shapiro–Wilk's Test for Multivariate Normality, Communications in Statistics - Theory and Methods, 38:11, 1870-1883

*Corresponding author.
E-mail address: ijobaidul@ mbstu.ac.bd