INTERNATIONAL JOURNAL OF ENVIRONMENTAL SCIENCES Volume 6, No 6, 2016

Research article ISSN 0976 – 4402

On identifying the SARIMA Model to forecast the humidity of some selected stations in Bangladesh

Md. Moyazzem Hossain
Department of Statistics, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh
mmhmm.justat@gmail.com
doi: 10.6088/ijes. 6102

ABSTRACT

Bangladesh is predominantly an agricultural country. Crop agriculture is highly influenced by climatic change and majority of population is dependent on agricultural crop in Bangladesh. The cultivation mainly depends on natural calamities like rainfall, humidity and temperature. Humidity affects crops through evaporation, transpiration and condensation. This paper attempts to identify the appropriate SARIMA model and check the statistical properties of the fitted model for the humidity for selected stations. Finally, forecast some future monthly humidity of the stations considered in this study. The forecasted series is really a better representation of the original humidity series of the selected stations in Bangladesh. Thus the fitted models are used to forecasting the monthly humidity for the upcoming two years to help decision makers to establish priorities in terms of water demand management.

Keywords: Humidity, Model Selection, Jarqu-Bera Test, Ljung-Box Test, Bangladesh.

1. Introduction

Bangladesh is predominantly an agricultural country. Crop agriculture is highly influenced by climatic change and majority of population is dependent on agricultural crop in Bangladesh. Crop agriculture is the mainstay of Bangladesh and will continue to be so in the foreseeable future. About 60.1% of the area is presently under agriculture and the sector contributes about 22% to the GNP. Any unfavorable change in future climate could have a devastating impact on agriculture and the economy of the country (Mondal, et al., 2012). The cultivation mainly depends on natural calamities like rainfall, humidity and temperature. Humidity affects crops through evaporation, transpiration and condensation. Humidity is the amount of Water vapor in the air, and relative humidity considers the ratio of the actual vapor pressure of the air to the saturated vapor pressure which is usually expressed in percentage. Atmospheric water vapor is a complex meteorological element. It is a fundamental component in the climate system as the most significant greenhouse gas and a key driver of many atmospheric processes. Water vapor and its transport around the atmosphere is a fundamental component of the hydrological cycle. The vapor plays a key role in determining the dynamic and radioactive properties of the climate system. Atmospheric water vapor accounts for only about 1/10,000th of the total amount of water in the global hydrological cycle. Nevertheless, atmospheric water vapor is one of the most important factors in determining Earth's weather and climate, because of its role as a greenhouse gas and because of the large amounts of energy involved as water changes between the gaseous (vapor) phase and liquid and solid phases. Humidity is very important as an environmental condition which influences the growth of the plants, health, pollution etc. For example, plants also respond to changes in humidity.

The prediction of atmospheric parameters is essential for climate monitoring, drought detection, severe weather prediction, agriculture and production, planning in energy and industry, communication, pollution dispersal etc. Accurate prediction of Humidity is a difficult task due to the dynamic nature of the atmosphere. It is very much essential to know the nature of changes of Humidity. Thus, this paper attempts to identify the appropriate SARIMA model and check the statistical properties of the fitted model for the humidity for selected stations. Finally, forecast some future monthly humidity of the stations considered in this study.

2. Literature review

This is very important area of research is well explored and a good number of works are made available. Islam, (2014) carried out a study to find Trends, periodicities and frequency distribution of the annual average humidity by using the standard statistical techniques. He considered the annual average humidity of 30 meteorological stations of Bangladesh over the period 1981 to 2008. He found that the frequency distribution of most of the stations of Bangladesh follow normal distribution. Positive trends are shown for the data of Dinajpur, Rajshahi, Mymensingh, Ishurdi, Jessore, Madaripur, Satkhira, Hatiya, Sitakunda, Teknaf & Patuakhali, while Dhaka the capital of Bangladesh has negative trend. The periodigram analyses of the annual average humidity of most of the stations show a significant cycle of range 8 to 12 years. Abu-Taleb, et al., (2007) examines the recent changes in annual and seasonal relative humidity variations in Jordan. Their analysis indicates an increasing trend in relative humidity at different stations. Their analysis also shows a significant increasing trend at Amman Airport Meteorological (AAM) station with a rate of increase 0.13% per year. These increasing trends are statistically significant during summer and autumn seasons. Finally, they found that a major change point in the annual relative humidity occurred in 1979 at AAM station.

Syeda (2012) investigates the trend and variability pattern for decadal, annual and seasonal (three crop seasons) average relative humidity (ARH) of six divisional stations in Bangladesh: Dhaka, Rajshahi, Khulna, Barisal, Sylhet and Chittagong. She examined the rates of linear trend (LT) for minimum, maximum and range humidity. To forecast the monthly ARH for 2009-2012 she used the univariate Box-Jenkin’s ARIMA modelling technique. The findings of her research were as follows: the rates of LT for annual ARH were found negative for Dhaka and Chittagong but positive for others. The rates were found negative for all the coefficient of variations (CVs). The rate for annual minimum humidity was positive for Dhaka but negative for others. The rates for annual maximum and range humidity were negative for Dhaka and Chittagong but positive for others. The rates for seasonal ARH were negative for Dhaka while positive for Rajshahi and Barisal in all the three seasons. It was negative for Kharif season, whereas positive for Prekharif and Rabi seasons for Khulna and Sylhet. It was negative for Kharif and Prekharif seasons, as the same time as positive for Rabi season for Chittagong.

Keka, et al., (2013) found the characteristics of different climatic parameters and the recent climatic trends in Bangladesh by using the surface climatological data at 30 stations for the period of 34 years from 1971 to 2004. Mondal, et al., (2013), shown that the relative humidity at Khulna has increasing trends of 2.3%, 1.3% and 0.3% per decade in the winter, post-monsoon and pre-monsoon seasons, respectively. Chowdhury and Hossain (2011) were carried out a study during 2005-08 to find the effect of weather prevalence of seedling diseases of jackfruit in different areas of Bangladesh and develop an environment friendly disease management practice. They found that occurrence of seedling diseases was significantly influenced by temperature, rainfall and relative humidity.

3. Material and methods

3.1 Study area

This study purposively selected eight stations from the different part of Bangladesh. They are Barisal, Bogra, Chittagong, Dhaka, Khulna, Rajshahi, Rangpur and Sylhet.

3.2 Data source

The necessary secondary data have been taken from the website of Bangladesh Agricultural Research Council (BARC) and http://en.tutiempo.net/climate/10-1982/ws-418830.html for measuring the humidity pattern.

3.3 Study period

The necessary data of monthly humidity were collected for the selected meteorological stations in Bangladesh over the period January, 1972 to July, 2015.

3.4 SARIMA model

Box and Jenkins (1976) suggested the use of seasonal autoregressive (SAR) and seasonal moving average (SMA) terms for monthly or quarterly data with systematic seasonal movements. The Box-Jenkins approach for modeling and forecasting has the advantage in analyze the seasonal time series data. In this case the seasonal components are included and the model is called seasonal ARIMA model or SARIMA model. A seasonal ARIMA model is classified as $ARIMA(p,d,q)\times(P,D,Q)_m$ model, where, $p=$ is the order of the Autoregressive (AR) term, $d=$ is the degree of differencing, $q=$ is the order of the Moving-average (MA) term and $P=$ is the number of seasonal autoregressive (SAR) terms, $D=$ is the number of seasonal differences, $Q=$ is the number of seasonal moving average (SMA) terms and $m=$ the number of time periods until the pattern repeats again. In the Seasonal ARIMA model the lowercase for non-seasonal part meanwhile the uppercase for seasonal part can be written in form of

$$\Phi_P\left(B^S\right)\phi_p\left(B\right)\nabla_S^D\nabla^d y_t=\Theta_Q\left(B^S\right)\theta_q\left(B\right)u_t$$

where, $\Phi_P\left(B^S\right)=1-\Phi_1B^S-\Phi_2B^{2S}-\ldots-\Phi_PB^{PS}$,

$\Theta_Q\left(B^S\right)=1-\Theta_1B^S-\Theta_2B^{2S}-\ldots-\Theta_QB^{QS}$,

$\phi_p\left(B\right)=1-\phi_1B-\phi_2B^2-\ldots-\phi_pB^p$,

$\theta_q\left(B\right)=1-\theta_1B-\theta_2B^2-\ldots-\theta_qB^q$, $\nabla_S=1-B^S$

3.5 Ljung-Box test

Ljung-Box (Ljung and Box, 1978) test can be used to check autocorrelation among the residuals. In this case the null hypothesis is $H_0:\rho_1(e)=\rho_2(e)=\ldots=\rho_k(e)=0$ is tested with the Ljung- Box statistic, $Q^*=N(N+1)\sum_{i=1}^{k}(N-k)\rho_k^2(e)$, where, N is the number of observation used to estimate the model. This statistic Q^* approximately follows the chi-square distribution with $(k-q)$ degrees of freedom, where q is the number of parameter should be estimated in the model. If Q^* is large (significantly large from zero), it is said that the residuals of the estimated model are probably auto-correlated. Thus, one should then consider reformulating the model.

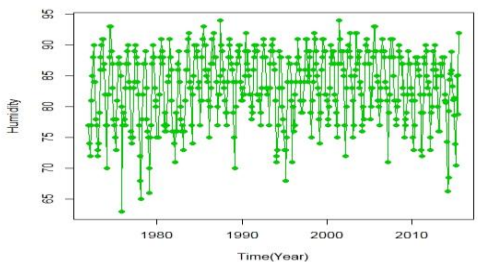
3.6 Jarque-Bera test

The normality assumption is checked by using Jarque and Bera (1987) test, which is a goodness of fit measure of departure from normality, based on the sample kurtosis (k) and skewness (s). The test statistics Jarque-Bera (JB) is defined as $JB=\frac{n}{6}\left(s^2+\frac{(k-2)^2}{4}\right)$, where n is the number of observations and k is the sample kurtosis an s is the sample skewness. The statistic JB has an asymptotic chi-square distribution with 2 degrees of freedom, and can be used to test the hypothesis of skewness being zero and excess kurtosis being zero, since sample from a normal distribution have expected skewness of zero and expected excess kurtosis of zero.

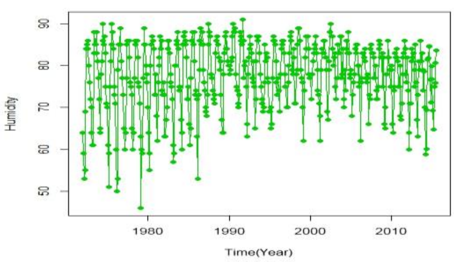
4. Result

For the purpose of analysis firstly we need to check the stationary of the series. Since our data are monthly so there may be exist seasonal variations. The Time Series plot of humidity fo the selected stations of this study are presented in Figure 1 (a)-(h).

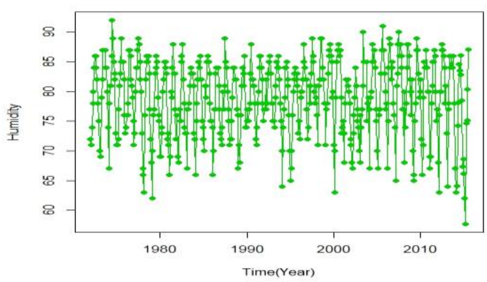
(a) Barisal



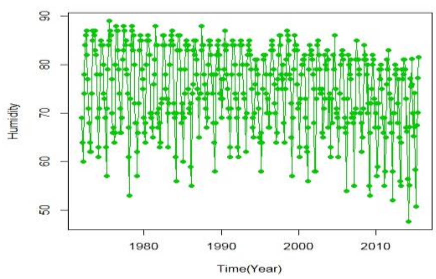
(b) Bogra



(c) Chittagong

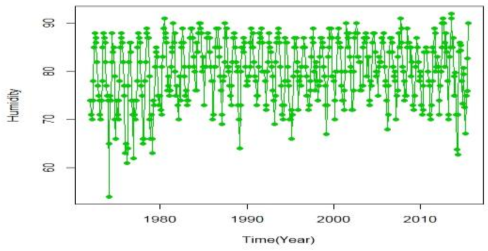


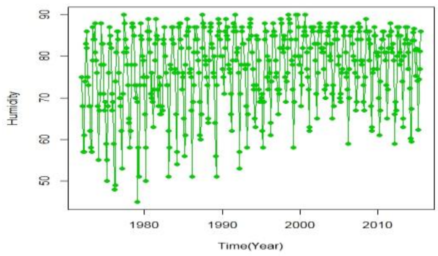
(d) Dhaka



(e) Khulna

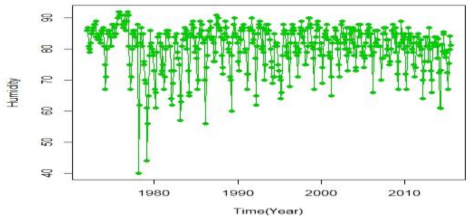
(f) Rajshahi





(g) Rangpur

(h) Sylhet



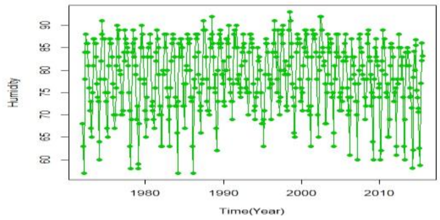


Figure 1 (a)-(h): Time Series Plot of Humidity for the Selected Stations

It is observed from the Time Series plot that there is a variation among the monthly humidity for all stations considered in this study over the study period. We also observed that in every month Humidity are changed i.e., either it increased or decreased. However, it is not observed any upward or downward trend in humidity for any station. Hence, it may be conclude that the Time Series is stationary. The ADF test is used to check the stationarity of the humidity Time Series for all stations considered in this study. The results of ADF test are given in Table 1.

Table 1: Results of ADF Test

Station	Test Statistic	P value	Decision
Barisal	-8.7253	0.001	Stationary
Bogra	-9.3236	0.001	Stationary
Chittagong	-9.077	0.001	Stationary
Dhaka	-9.9599	0.001	Stationary
Khulna	-8.754	0.001	Stationary
Rajshahi	-9.301	0.001	Stationary
Rangpur	-10.427	0.001	Stationary
Sylhet	-9.3986	0.001	Stationary

According to ADF test we may conclude that the Time Series of humidity for all stations selected for this study are stationary at 5 percent level of significance. To check the seasonal variation we compute the seasonal index. The values of seasonal index are given in Table 2.

Table 2: Seasonal Index for Selected Stations

Month	Seasonal Index							
	Barisal	Bogra	Chittagong	Dhaka	Khulna	Rajshahi	Rangpur	Sylhet
January	96.61	98.16	93.79	94.87	96.75	99.24	101.24	95.44
February	93.13	90.92	90.56	88.02	92.54	92.32	94.58	88.76
March	91.70	86.32	93.63	86.47	91.18	83.91	87.24	87.88
April	96.04	92.11	98.12	95.49	95.43	87.31	92.58	96.65

May	99.70	99.87	100.77	102.0	98.89	97.66	99.70	102.4
June	104.50	106.10	105.23	107.9	104.9	106.11	104.42	107.9
July	106.31	108.34	107.09	109.1	107.5	110.12	105.15	108.4
August	105.87	107.66	106.61	108.3	106.7	109.27	104.51	107.6
September	105.25	108.10	105.37	108.2	106.7	108.75	105.77	107.3
October	103.16	104.38	103.26	103.8	103.1	105.11	103.69	103.6
November	99.89	98.97	98.97	98.12	98.59	100.21	100.24	97.96
December	97.85	99.10	96.61	97.81	97.67	99.99	100.88	96.10

Seasonal index for all stations suggest that there is a seasonal effect on humidity (Table 2). It indicates that we have to adjust the data to remove the seasonal effect for building the Time Series model which is used to forecast the humidity of the selected stations. After adjusting the seasonal effect we find the ACF and PACF. Form the ACF and PACF we identify the tentative models for the selected stations. We identify the appropriate model by using model selection criteria such as AIC, AIC_C and BIC. The lower value indicates the better model. We change the parameters of the model and record the values of the model selection criteria. Among the models we select the best model. The values of the model selection criteria for the selected models are given in Table 3.

Table 3: Values of Different Model Selection Criteria

Station	AIC	AIC_C	BIC
Barisal	2651.09	2651.25	2676.65
Bogra	2911.14	2911.31	2936.69
Chittagong	2816.72	2816.94	2846.54
Dhaka	2941.48	2941.76	2975.54
Khulna	2711.56	2711.72	2737.11
Rajshahi	3169.70	3169.78	3186.73
Rangpur	2989.88	2990.16	3023.95
Sylhet	2992.38	2992.5	3013.68

The estimates of the parameter for different stations considered in this study are presented in the following Table 4. The values in parenthesis indicate the standard error of the parameters. Form Table 4, it is observed that for different model the estimated models are different.

Table 4: Estimates of the Parameters of the Final Model for Different Stations

Barisal Station $ARIMA(2,0,0)(2,1,0)_{12}$				
ar1	ar2	sar1	sar2	Intercept
0.4202 (0.0479)	0.0607 (0.0441)	0.4997 (0.0441)	0.2694 (0.0443)	82.9256 (0.9923)
Bogra Station $ARIMA(2,0,1)(2,1,0)_{12}$				
ar1	ar2	ma1	sar1	sar2
0.3823 (0.0473)	0.029 (0.0453)	-0.9747 (0.0136)	0.4675 (0.0417)	0.3235 (0.044)
Chittagong Station $ARIMA(3,0,0)(2,1,0)_{12}$				

ar1	ar2	ar3	sar1	sar2	Intercept
0.3611 (0.0491)	0.0719 (0.0465)	0.0089 (0.0463)	0.4855 (0.0443)	0.2866 (0.0435)	78.5895 (1.091)

Dhaka Station $ARIMA(5,0,1)(1,1,0)_{12}$

ar1	ar2	ar3	ar4	ar5	ma1	sar1
0.445 (0.0529)	0.0113 (0.0468)	0.0197 (0.0486)	-0.0435 (0.0495)	-0.1242 (0.052)	-0.9877 (0.006)	0.7018 (0.0589)

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

ar1	ar2	sar1	sar2	Intercept
0.4165 (0.0457)	0.1228 (0.0444)	0.3771 (0.0387)	0.4495 (0.0401)	79.7056 (1.4913)

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

ar1	sar1	sar2
-0.2284 (0.0446)	0.3989 (0.0406)	0.4225 (0.0407)

Rangpur Station $ARIMA(4,0,0)(2,1,0)_{12}$

ar1	ar2	ar3	ar4	sar1	sar2	Intercept
0.5037 (0.0458)	-0.0193 (0.0495)	0.0634 (0.049)	0.0677 (0.045)	0.476 (0.042)	0.264 (0.04)	80.8113 (1.6544)

Sylhet Station $ARIMA(1,0,1)(1,1,0)_{12}$

ar1	ma1	sar1	Intercept
0.488 (0.1159)	-0.188 (0.1359)	0.767 (0.0332)	78.3405 (1.1497)

Note: Values in parenthesis indicate the standard error

In order to forecast the future value by a model, it is needed to check the accuracy of the model. The accuracy of a model is done by justifying the statistical properties of the error of the fitted model. We check the normality and autocorrelation of the error of the fitted model. Ljung-Box test is used to check the autocorrelation and Jarque-Bera test is used to check the normality assumption of the error of fitted model. The values of the test statistic for the fitted model of different locations are presented in Table 5.

Table 5: Value of Test Statistic

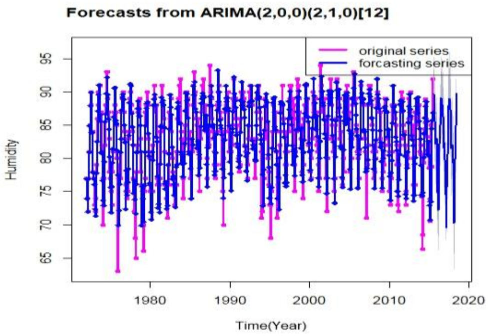
Station	Ljung-Box		Jarque-Bera	
	Test Statistic	P value	Test Statistic	P value
Barisal	26.1009	0.01039	100.947	0.0
Bogra	16.356	0.1755	147.361	0.0
Chittagong	25.3022	0.0134	35.7032	0.0
Dhaka	22.293	0.034	20.423	0.0
Khulna	21.525	0.0432	86.177	0.0
Rajshahi	14.474	0.2714	301.608	0.0
Rangpur	14.868	0.2487	6453.891	0.0
Sylhet	25.459	0.0128	29.1772	0.0

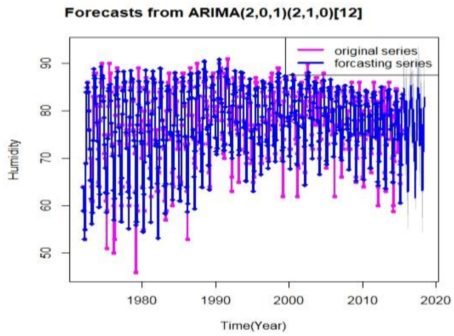
From Table 5, it is observed that the errors of the fitted model are normally distributed at 5 percent level of significance for all stations considered in this study. Also, there is no autocorrelation among the residuals of the fitted model for Barisal, Chittagong, Dhaka, Khulna and Sylhet. For Bogra, Rajshahi and Rangpur the errors are autocorrelated. The graphical comparison of the observed and predicted value is given in the following Figure 2 of all stations considered in this study.

It is observed that the forecasted series (blue-color) fluctuated from the original series (pink-color) with a very small amount which shows the fitted series has the same manner of the original series for all stations considered in this study (Figure 2). Therefore, the forecasted series is really a better representation of the original humidity series of the selected stations in Bangladesh.

(a) Barisal

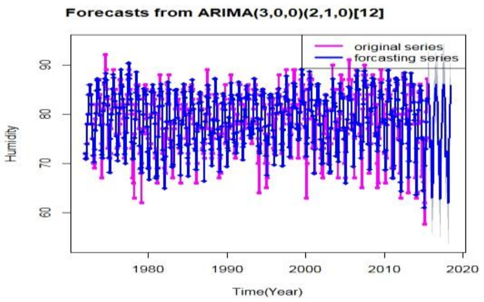
(b) Bogra

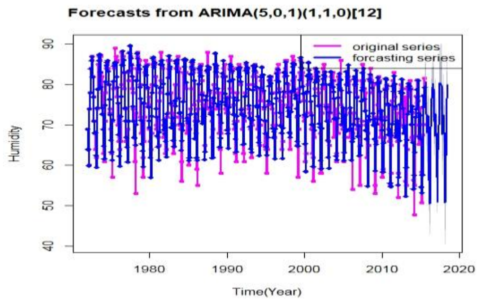




(c) Chittagong

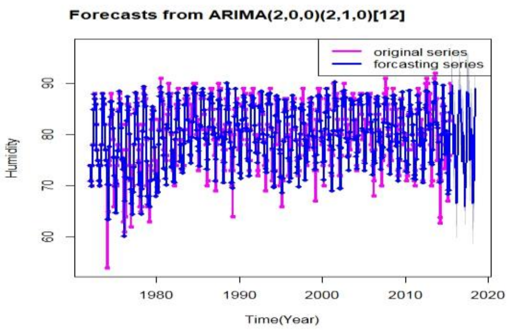
(d) Dhaka

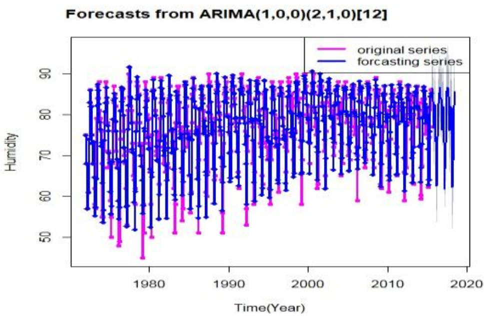




(e) Khulna

(f) Rajshahi





(g) Rangpur

(h) Sylhet

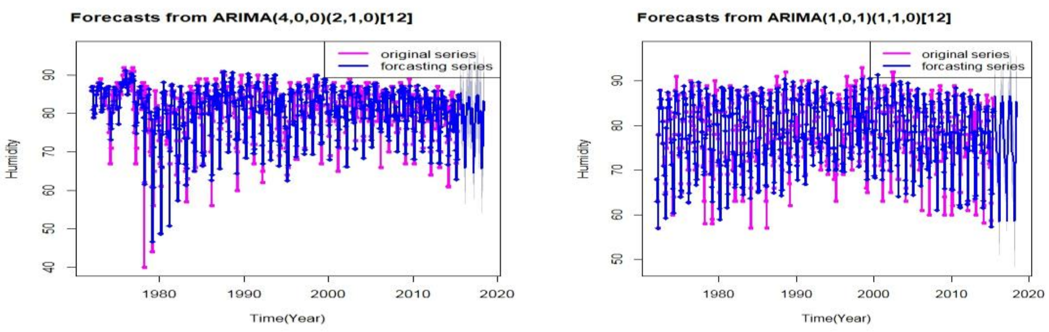


Figure 2(a)-(h): Comparison of Observed and Predicted Values of Humidity

5. Conclusion

The prediction of atmospheric parameters is essential for climate monitoring, drought detection, severe weather prediction, agriculture and production, planning in energy and industry, communication, pollution dispersal etc. The original Time Series of humidity for all stations selected for this study are stationary at 5 percent level of significance. Also, to check the seasonal variation we compute the seasonal index. From the seasonal index we may conclude that there is a seasonal variation of the humidity series for all stations considered in the study. Thus, the seasonal variations are adjusted. After that, this paper identifies the $ARIMA(2,0,0)(2,1,0)_{12}$, $ARIMA(2,0,1)(2,1,0)_{12}$, $ARIMA(3,0,0)(2,1,0)_{12}$, $ARIMA(5,0,1)(1,1,0)_{12}$, $ARIMA(2,0,0)(2,1,0)_{12}$, $ARIMA(1,0,0)(2,1,0)_{12}$ and $ARIMA(1,0,1)(1,1,0)_{12}$ model are found to be suitable models for Barisal, Bogra, Chittagong, Dhaka, Khulna, Rajshahi, Rangpur and Sylhet station respectively. These models are selected on the basis of model selection criteria like AIC, AIC_C and BIC. We check the normality and autocorrelation of errors of the fitted model. Ljung-Box test is used to check the autocorrelation and Jarque-Bera test is used to check the normality assumption of the error of fitted model. It is observed that the errors of the fitted model are normally distributed at 5 percent level of significance for all stations considered in this study. Also, there is no autocorrelation among the residuals of the fitted model for Barisal, Chittagong, Dhaka, Khulna and Sylhet at 5 percent level of significance. The autocorrelation is present in Bogra, Rajshahi and Rangpur stations.

From the graphical comparison, it is observed that the forecasted series fluctuated from the original series with a very small amount which shows the fitted series has the same manner of the original series for all stations considered in this study. Therefore, the forecasted series is really a better representation of the original humidity series of the selected stations in Bangladesh. Thus the fitted models are used to forecasting the monthly humidity for the upcoming two years to help decision makers to establish priorities in terms of water demand management.

Acknowledgement

The authors are grateful to University Grants Commission of Bangladesh for providing the financial support through the faculty of physical and mathematical science, Jahangirnagar University, Bangladesh doing this research.

6. References

1. Abu-Taleb A.A., Alawneh A.J., and Smadi M.M., (2007), Statistical Analysis of Recent Changes in Relative Humidity in Jordan, American Journal of Environmental Sciences, 3(2), pp 75-77.

2. Box G.E.P., and Jenkins G.M., (1976), Time Series Analysis: Forecasting and Control, Revised Edition, San Francisco: Holden Day.

3. Chowdhury M.S.M., and Hossain I., (2011), Effects of Temperature, Rainfall and Relative Humidity on Leaf Spot of Jackfruit Seedling and its Eco-friendly Management, The Agriculturists, 9(1&2), pp 126-136.

4. Islam M.M., (2014), Regional Differentials of Annual Average Humidity over Bangladesh, ASA University Review, 8(1), pp 1-14.

5. Jarque C.M., and Bera A.K., (1987), A test for normality of observations and regression residuals, International Statistical Review, 55(2), pp 163–172.

6. Keka I.A., Matin I., and Banu D.A., (2013), Recent Climatic Trends in Bangladesh, Daffodil International University Journal of Science and Technology, 8(2), pp 43-50.

7. Ljung G.M., and Box G.E.P., (1978), On a Measure of a Lack of Fit in Time Series Models, Biometrika, 65 (2), pp 297–303.

8. Mondal M.S., Islam A.K.M.S., and Madhu M.K., (2012), Spatial and Temporal Distribution of Temperature, Rainfall, Sunshine and Humidity in Context of Crop Agriculture, Institute of Water and Flood Management Bangladesh University of Engineering & Technology, Bangladesh.

9. Mondal M.S., Jalal M.R., Khan M.S.A., Kumar U., Rahman R., and Huq H., (2013), Hydro-Meteorological Trends in Southwest Coastal Bangladesh: Perspectives of Climate Change and Human Interventions, American Journal of Climate Change, 2, pp 62-70.

10. Syeda J.A., (2012), Variability Analysis and Forecasting of Relative Humidity in Bangladesh, Journal of Environment Science & Natural Resources, 5(2), pp 137 – 147.