Civil and Environmental Research www.iiste.org
ISSN 2224-5790 (Paper) ISSN 2225-0514 (Online)
Vol.7, No.1, 2015

Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model

S H BARI[1],MT RAHMAN[2], MM HUSSAIN[2], SOURAV RAY[2]

[1] Department of Civil Engineering, Leading University, Sylhet, Bangladesh
[2] Department of Civil & Environmental Engineering, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh
*Corresponding Author <sourav.ceesust@gmail.com>

ABSTRACT
In this study a seasonal ARIMA model was built using Box and Jenkins method to forecast long term rainfall in Sylhet. For this purpose rainfall data from 1980 to 2010 of Sylhet station were used to build and check the model. Rainfall data from 1980 to 2006 were used to develop the model while data from 2007 to 2010 were used to verify the prediction precision. Four basic chronological steps namely: identification, estimation, diagnostic checking, and forecasting were fitted out in developing the model. Validity of the model was tested using standard graphical explanation of residuals given by Box and Jenkins. As a second step of validation, forecasted values of monthly rainfall were checked using actual data series. After completion of necessary checking and forecast observation, the ARIMA(0, 0, 1) $(1,1, 1)_{12}$ was found to be the most effective to predict future precipitation with a 95% confidence interval. It is expected that this long term prediction will help decision makers in efficient scheduling of flood prediction, urban planning, rainwater harvesting and crop management.

Keywords: Nonlinear time series analysis, ARIMA model, rainfall forecasting, Sylhet.

INTRODUCTION
Rainfall is very non-linear in nature and very complicated to predict. Due to adverse effects of climate change rainfall pattern has also been changing rapidly Short term and long term forecast of rainfall have significant relevance to agricultural, tourism, flood prevention and management strategy and water body management which influence the economy of a country. To predict such event, numerous techniques including numerical and machine learning processes have been adopted based on historical time series and radar data (Chander et al. 2002, Ingsrisawang et al. 2008). Still, currently most common methodology for rainfall prediction uses radar image data available from various organizations and analyzing them to predict rainfall. However various statistical methods are often useful to predict rainfall (Bisgaard and Kulahci, 2011). Among which the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins (1976) and modified by Box, et al. (1994), also known as ARIMA (Autoregressive Integrated Moving Average).ARIMA has widely been exercised over the years to predict the rainfall trend (Mahsin et al. 2012,Kaushik and Singh 2008, Shamsnia et al. 2011, Thapaliyal 1981, Momani et al.2009), reservoir and river modeling (Dizon 2007, Cui 2011, Peng et al. 2000, Valipour et al. 2012), economics and production(Nochai et al. 2006),evapotranspiration(Valipour 2012).The method has some interesting features that made it more desirable for researchers. It eases the forecasting process allowing researchers to use only single variable time data series while also allow multiple for more complex cases.

Rainfall forecast study for Dhaka division of Bangladesh has been done by Mahsin et al. (2012). However, forecasting of rainfall for Sylhet hasn't been done yet. The area is one of the top tourist attracting locations in Bangladesh. Although the area is nearby the world's wettest place (Cherapunji), urban Sylhet and some other region including Surma river basin are facing a rapid ground water depletion rate. Besides, excessive iron and arsenic contamination has an adverse impact in drinking water supply. To pursue a sustainable alternative source of supply, rainwater has become a suitable option in many parts of the region. Therefore the study objective of this study is to focus on development of a reliable forecasting of rainfall over Sylhet to manage water resources as well as handle flash flood effective and timely.

METHODOLOGICAL APPROACH:

(i) Study location and data collection

Bangladesh meteorological department (BMD) collects rainfall data for Bangladesh through its 34 stations. Rainfall data for Sylhet station (Fig.:01) collected from BMD were used in the study. This station covers Sylhet District and it's nearby areas.

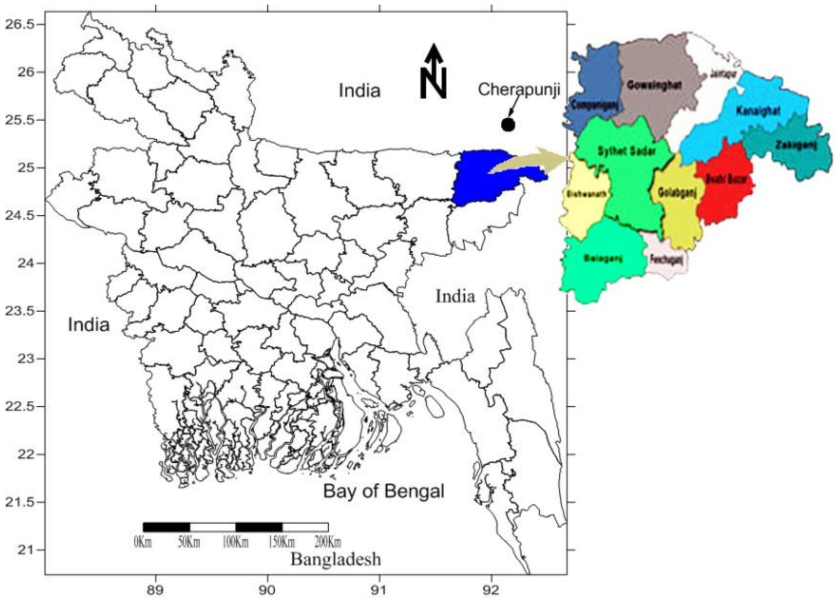


Figure01: Location of the study area in Bangladesh

The climate of Sylhet is tropical monsoon with a predominantly hot and humid summer and a relatively cool winter. The area is within the monsoon climatic zone that results a heavy rainfall in monsoon. Annual average highest temperatures is of 23°C (August to October) and average lowest temperature is 7°C (January). Nearly 80% of the annual average rainfall occurs between May and September (SCC 2012). Sylhet is located very cloase to Cherapunji, which receives the highest rainfall in the world. Cherapunji, received (Murata et al. 2008). A relatively shorter distance of 50 km from the world'swettest place (Cherapunji) influences heavily in the local rainfall. Comparative representation of normal annual and monthly average rainfall (Fig.02 and Fig.: 03) indicated higher intensity of rain in Sylhet. Monthly rainfall data from 1980 to 2010 that wereused for the studywere collected from Bangladesh meteorological department. Details of Rain gauge station and data used are shown in Table: 01.

Table 01: Data used in the Study with station location

Station	latitude(N)	longitude(E)	(elevation)	used data period		mean annual rainfall(mm)
				model development	model validation	
Sylhet	$24^0 54''$	$91^0 53''$	33.53	1980-2006	2007-2010	3963

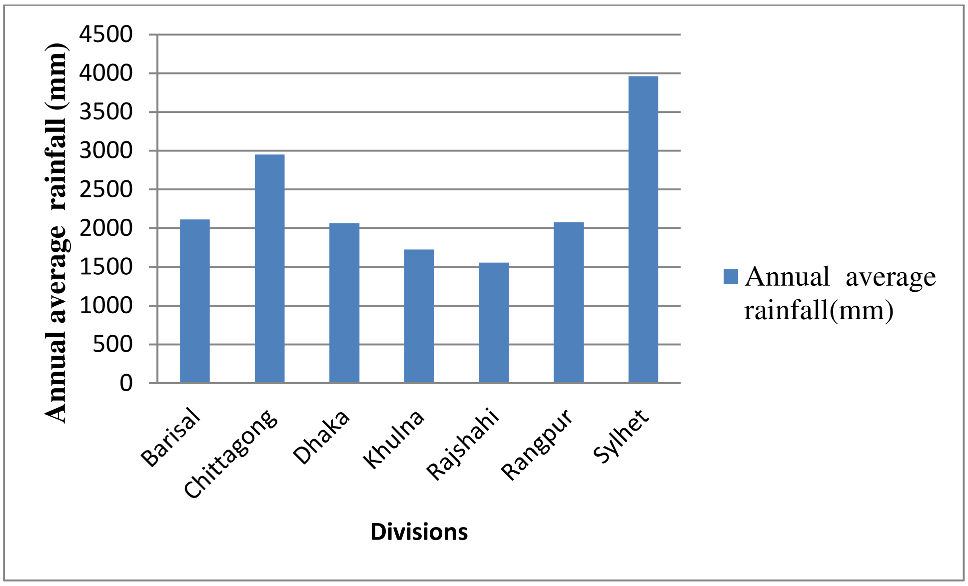


Figure02: A comparison of annual average rainfall in divisional cities in Bangladesh.

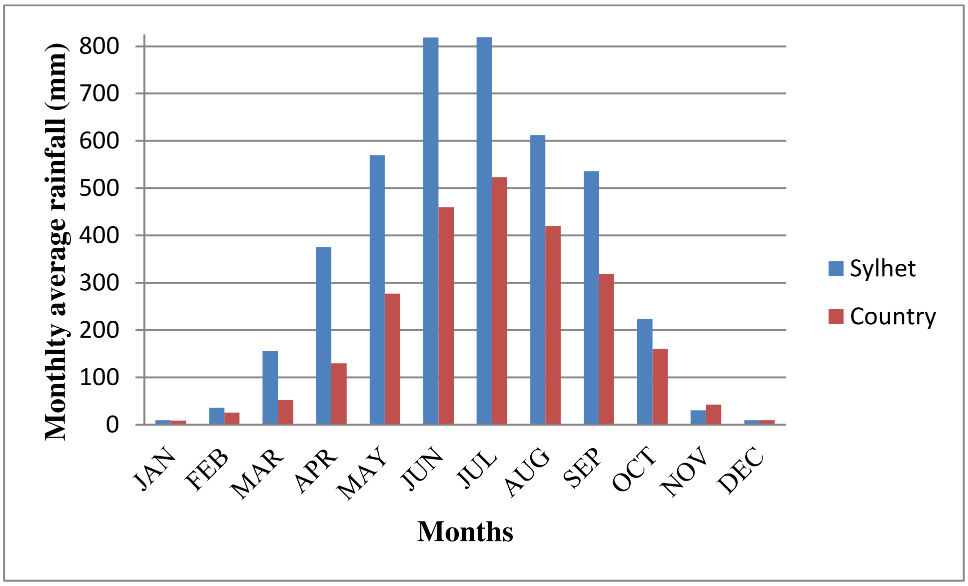


Figure 03:A comparison of normal monthly average rainfall (mm)between sylhet and Bangladesh

(ii) ARIMA model development

For developing an ARIMA model it requires to follow four basic steps: identification, estimation, diagnostic checking, and forecasting (Bowerman and O'Connell, 1993). 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) × $(P, D, Q)_S$, with 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 = time span of repeating seasonal pattern (Box et al. 1994).

The following flow diagram demonstrates basic methodology of ARIMA development:

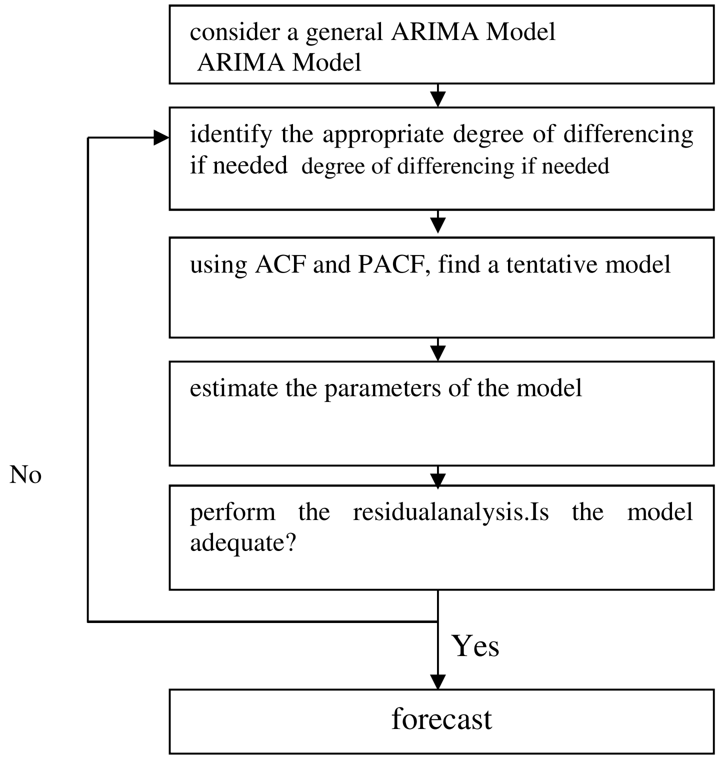


Figure 04: ARIMA modeldevelopment

RESULTS AND DISCUSSION

(i) Model identification

The first step is to check whether there is any seasonality exists in the observed data and if the data is stationary. Time Series plot (Fig.: 05) shows that there is a clear seasonality with periodicity of one year (twelve month) in the data set. The ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) are significant in identifying stationary of the data set. Both function (Fig.: 06) indicated that, the data is non-stationary.

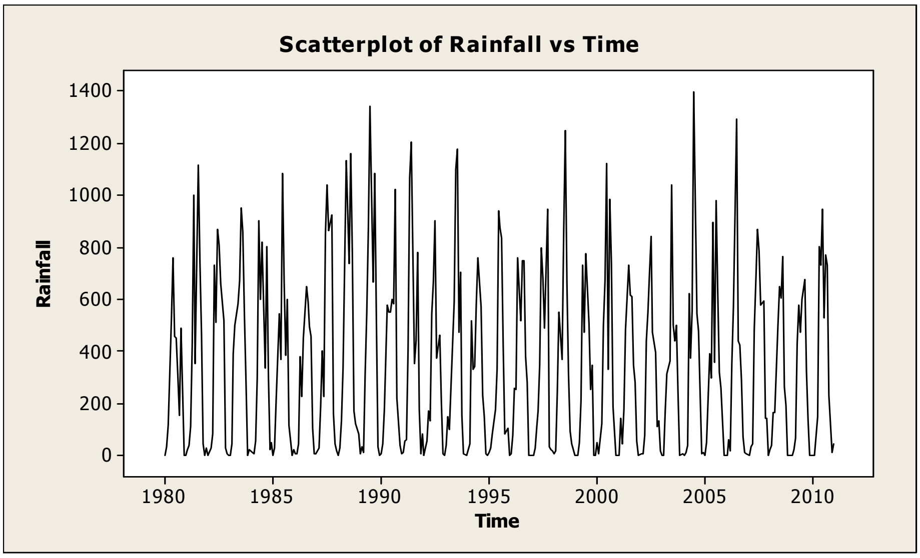
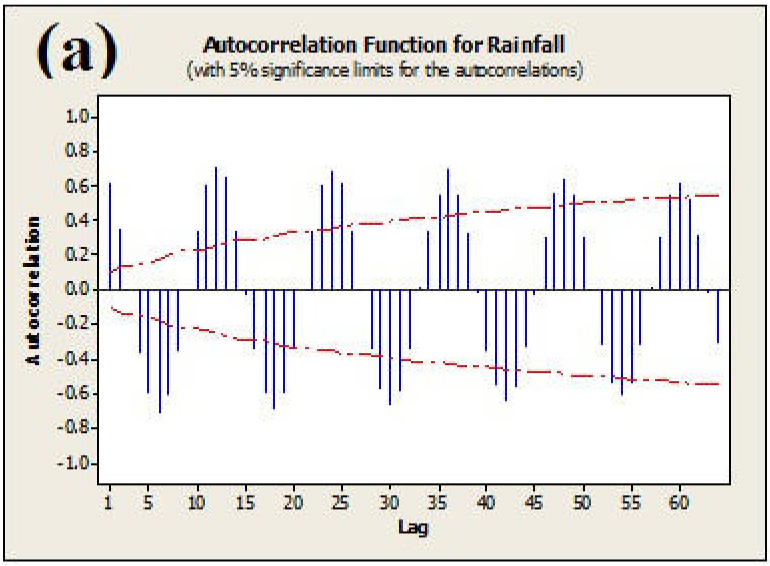


Figure 05: Time series plot of observed data.



(a)

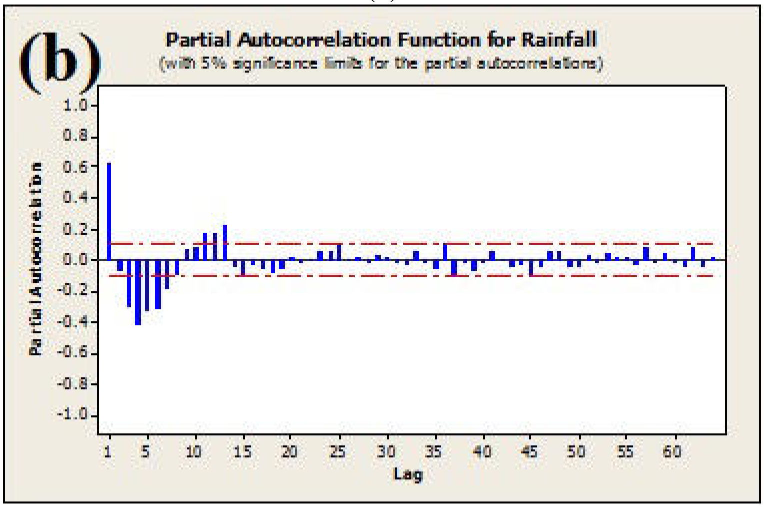
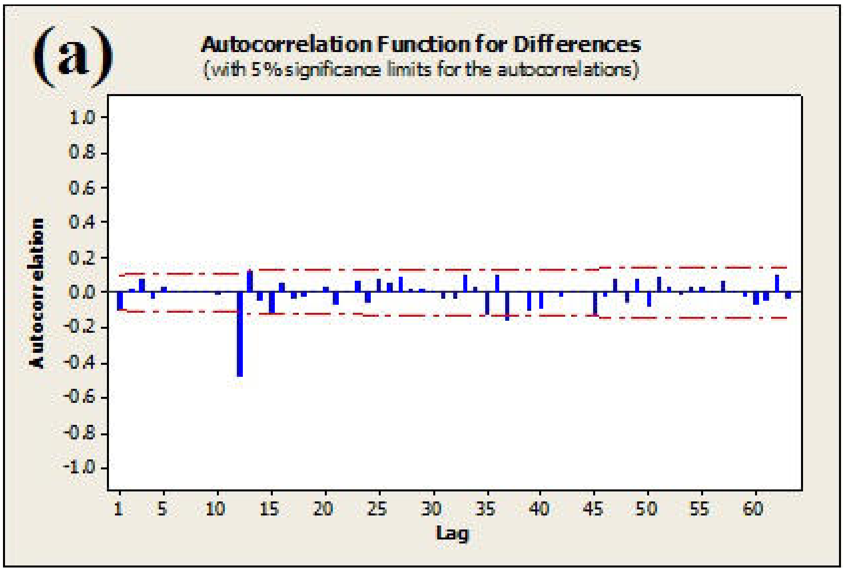


Figure 06: (a) Plot of ACF, (b) plot of PACFfor observed rainfall data.

In order to fit an ARIMA model stationary data in both variance and mean are needed. Stationarity could be attained in the variance by having log transformation and differencing of the original data to attain stationarity in the mean. Since data series contain zero values straight forward log transformation is not possible. In this data series, a seasonal first difference (D = 1) of the original data was done in order to obtain stationarity. Thereafter, ACF and PACF for the differenced series were tested to check stationarity. The ACF and PACF (Fig.: 07) show that one order seasonal differencing is adequate. Although further differencing shows a similar result but first seasonal differencing has a minimum standard deviation. Therefore, one order difference is enough for the data series. From this, a preliminary ARIMA (p, 0, q) × (P, 1, Q)$_{12}$ was selected.



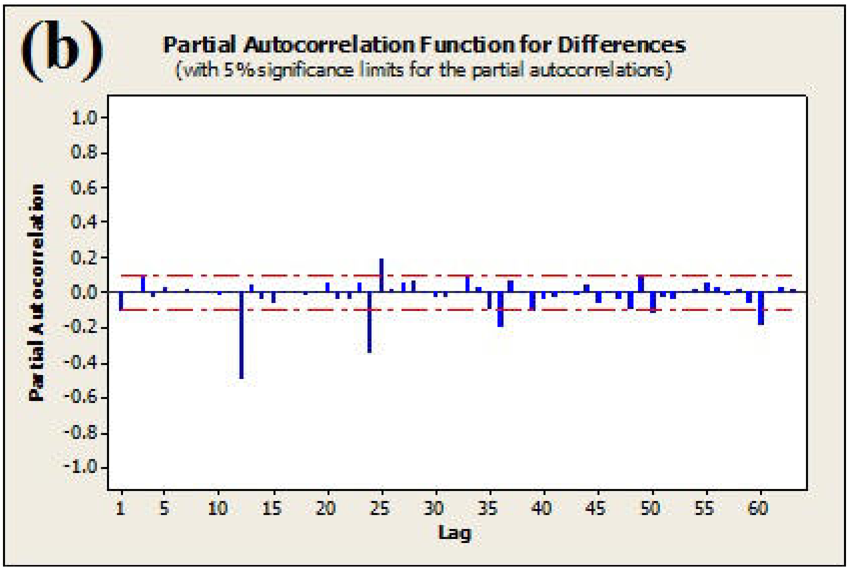


Figure 07: First order seasonal differencing and de-seasonalized original rainfall data for, (a) ACF, (b) PACF.

(ii) Model Estimation

Since the orders P, p and Q, q necessary to adequately model for a given problem is not known to us. It is required to determine the model that best fits the data based on observing the ACF and the PACF of the differenced data. After carefully examining ACF and PACF, following five models were identified for test. These models are: ARIMA $(0,0,1)(1,1,1)_{12}$, ARIMA $(0,0,0)(1,1,1)_{12}$, ARIMA $(0,0,0)(0,1,1)_{12}$, ARIMA $(0,0,1)(1,0,1)_{12}$, ARIMA $(0,0,0)(1,1,1)_{12}$(Fig. 07).

(iii) Model Diagnostic Checking

Once the models have been fitted to the data, a number of diagnostic checks were initialized. If the model fits well, the residuals should be uncorrelated with constant variance. Moreover, in developing model this is often assumed that the errors are normally distributed. Hence, we expect the residuals to be more or less normally distributed.

Standard checks for ARIMA is to compute the ACF and the PACF of the residuals. Further diagnostics checking can be done by looking at the residuals in various ways (Figure: 08).If the residuals are normally distributed, they should all more or less lie on a straight upward sloping line (Bowerman, and O'Connell, 1993). After accomplishes the above series checking the ARIMA (0, 0, 0) $(1, 1, 1)_{12}$ and ARIMA (0, 0, 1) $(1, 1, 1,)_{12}$ were found to give significant results. But compared to the forecasting accuracy alongside the diagnostic checking ARIMA (0, 0, 1) $(1, 1, 1)_{12}$ was found to be our desired model.

(iv) Forecasting

ARIMA (0, 0, 1) $(1, 1, 1)_{12}$ was applied to forecast the monthly rainfall data from January 2007 to December 2012. Forecasted values of January 2007 to December 2010 ware used to compare the observed and forecasted values. The forecasted value for rainfall shows significant result. Forecasted time series and the observed time series with the 95% confidence level error bound are plotted in Fig.: 09.It is observed that measured monthly values fall within the error bound, and the forecasted track of the seasonal pattern fits reasonably well.

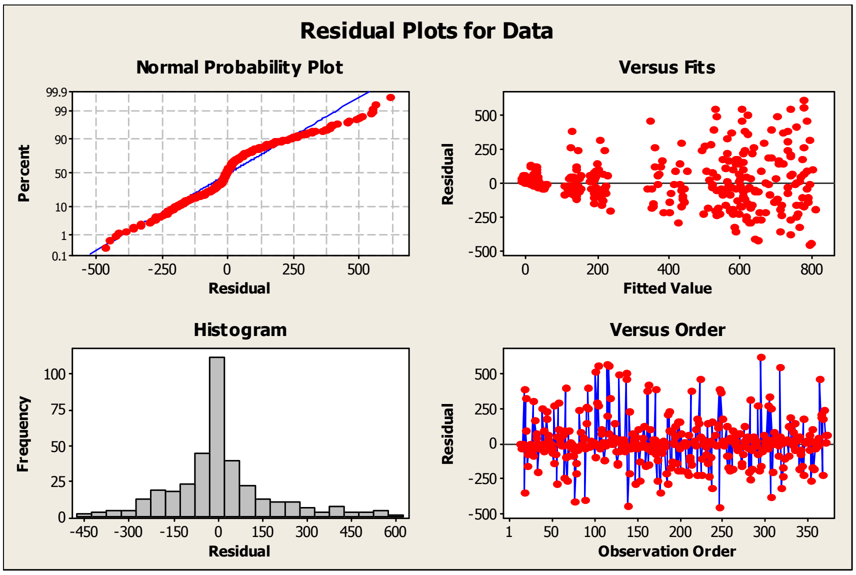


Figure 08: Diagnostics for the ARIMA (0, 0, 1) $(1, 1, 1)_{12}$ fit on the rainfall data for Sylhet.

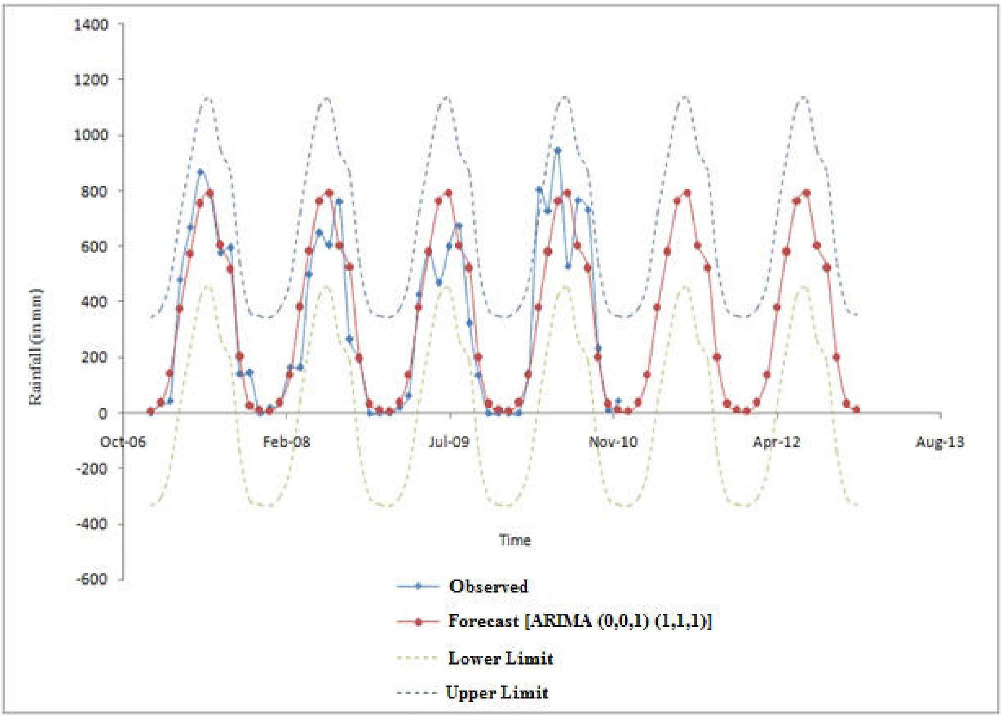


Figure 09: Observed and forecasted data with 95% confidence limit

APPLICATION

Sylhet is a divisional city with a rapid growing economy and tourist attractions. The developed model could be applied for the following fields for managing water resources problems efficiently.

(i) Flood Management: Devastating floods due to excessive rainfall frequently creates catastrophe to human livelihood and economy. Developed model can be used to predict such extreme events.

(ii) Rainwater Harvesting: Although Sylhet receives great amounts of annual rainfall; unfortunately ground water is either Arsenic or Iron contaminated along with a fast depletion rate. A suitable alternative for this is Rainwater Harvesting. Moreover, Bangladesh government has already added a provision for implementing Rainwater Harvesting system in urban areas. This model could be used for future feasibility study of such activities.

(iii) Haor Basin Management: Efficient rainfall fore casting will help to develop a result oriented Haor (Large bowled shape depression) management plan.

(iv) Tourism: Sylhet is one of the top tourist attraction places now days. Travel plan could be advised according to the forecasted rainfall events.

(v) Crop Calendar Adjustment: Crop production has now been facing severe unfavourable events due to climate change effects. Cropping pattern can be adjusted according to the future rain occasions.

(vi)Urban Planning: A sustainable urban development cannot decline a proper storm water management process. Long term forecasting of rainfall can help to mak an effective plan.

CONCLUSION

Here an ARIMA model has been developed. Comparing the observed and forecasted values with 95% confidence limit,the presented model gives a reasonable result. Therefore, this model could help to determine possible future strategy in the respective field for the Sylhet City and its nearby areas. Model diagnostic checking presented that ARIMA(0,0,0) $(1,1,1)_{12}$ and ARIMA(0,0,1) $(1,1,1)_{12}$ should have significant results.

ACKNOWLEDGEMENT

A partial funding from SUST Research Grant-2012 related with logistic support of this research is sincerely acknowledged.

LIST OF REFERENCES

Bowerman, B.L. and O'Connell, R.T. 1993. Forecasting and Time Series: An Applied Approach, Duxbury Press, Belmont, CA.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. Time Series Analysis, Forecasting and Control, Englewood Cliffs, N.J: Prentice Hall.

Box, G.E.P. and Jenkins,G.M. 1976. Time Series Analysis: Forecasting and Control, Revised Edn., Hoden-Day, San Francisco.

Bisgaard , S. and Kulahci,M. 2011.Time series analysis and forecasting by example, Hoboken, New Jersey: John Wiley & Sons, Inc.

Chander, R.E. and Wheater, H.S. 2002. Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, Water Resour. Res., 38(10), 10-1– 10-11.

Cui, F. 2011. ARIMA models for bank failures: Prediction and comparison, 2011,University of Nevada, Las Vegas.

Dizon, Cornelio Q. 2007. Arma modeling of a stochastic process appropriate for the angat reservoir ,Philippine Engineering Journal, Vol. 28, No. 1:1-20.

Ingsrisawang, L., Ingsriswang, S., Somchit, S., Aungsuratana, P. and Khantiyanan, W. 2008. machine learning techniques for short-term rain forecasting system in the Northeastern part of Thailand,Proceedings of World Academy of Science, Engineering and Technology, 31, 248 -253.

Kaushik, I.. andSingh, S.M. 2008. Seasonal ARIMA model for forecasting of monthly rainfall and temperature,Journal of Environmental Research and Development Vol. 3(2).

Mahsin, M., Akhter, Y. and Begum, M. 2012. Modeling rainfall in Dhaka division of Bangladesh using time series analysis, Journal of Mathematical Modelling and Application, Vol. 1, No.5, 67-73.

Momani, P.E., Naill M. 2009.Time series analysis model for rainfall data in Jordan: Case Study for using time series analysis,American Journal of Environmental Sciences 5 (5): 599-604.

Murata,F., Terao,T., Hayashi T., Asada H. and Matsumoto J. 2008.Relationship between atmospheric conditions at Dhaka, Bangladesh and rainfall at Cherrapunjee, India, Nat Hazards ,44:399–410.

Nochai R. and Nochai, T. 2006.ARIMA model for forecasting oil palm price, Proceedings of the 2nd IMT-GT regional conference on mathematics, statistics and application, Malaysia, penang, June 13-15.

Peng, M.,Liu, J., and Shih, T. 2000. Groundwater level forecasting with time series analysis,ACRS(online). Available via: http://www.a-a-r-s.org/acrs/proceeding/ACRS2000/Papers/WR00-5.htm. [Accessed 20 August 2012].

SCC, 2012. http://www.sylhetcitycorporationbd.com [Accessed on 16thAugust 2012]

Shamsnia, S. A., Shahidi, N., Liaghat, A., Sarraf, A., and Vahdat, S. F. 2011. Modeling of weather parameters using stochastic methods (ARIMA Model)(Case Study: Abadeh Region, Iran),Proc. of Int. Conf. on Environment and Industrial Innovation, IPCBEE vol.12, IACSIT Press, Singapore.

Thapaliyal, V. 1981. ARIMA model for longrange prediction of monsoon rainfall in Peninsular India. India Meteorol Dept Monogr Climatology, 12/81.

Valipour, M. 2012. Ability of Box-Jenkins models to estimate of reference potential evapotranspiration (A Case Study: Mehrabad Synoptic Station, Tehran, Iran), IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS), Volume 1, Issue 5 , 01-11

Valipour, M., Banihabib, M. E. and Behbahani S. M. R. 2012. Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting, Journal of Mathematics and Statistics 8 (3): 330-338.