Journal of Natural Sciences Research www.iiste.org
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.5, No.21, 2015

Year-Long Monthly Rainfall Forecasting for a Coastal Environment of Bangladesh

Ishtiak Mahmud[1] SH Bari[2] M.M. Hussain[1]
1. Dept. of Civil and Environmental Engineering, Shahjalal University of Science and Technology, Sylhet-3114
2. Dept. of Civil Engineering, Leading University, Sylhet

Abstract
Forecasting rainfall plays an important role to develop, planning and management a sustainable water resource system. In this study stochastic Seasonal Auto Regressive Integrated Moving Average (SARIMA) were used to forecast monthly rainfall of Teknaf for 12 month lead time. The best SARIMA (0, 0, 0) (1, 1, 1) model was selected based on Normalized BIC (Bayesian Information Criteria) and R-squared. Diagnostic check was then conducted for the best fitted model to check if the residuals are white noise. The predicted rainfall amount from the best fitted model was compared with the observed data. The predicted values shows reasonably good result. Thus the model can be used for future rainfall prediction.
Keywords: Bangladesh, Teknaf, Rainfall, ARIMA, Forecast

1. Introduction
Rainfall is the most significant natural phenomena. To find out a solution for several regional environmental problems, understanding the rainfall process is important. Further, due to emerging water crisis rain feed water supply is getting higher priority. In addition, in country like Bangladesh rainfall is the key component in agricultural production. Therefore, accurate rainfall prediction will be helpful for planning for future water resource related issues and crop pattern.

To predict rainfall amount numerous technique has been embraced. Among them Univariate ARIMA (Auto Regressive Integrated Moving Average) model is one of the most efficient and widely practiced methods to predict precipitation. (Johnson and Montgomery 1976) consider Box and Jenkins methodology as probably the most exact method to forecast time series information. (Gerretsadikan and Sharma 2011) mentioned that the Box-Jenkins method is universal means to forecast, unlike any other method. (Zakaria et al. 2012) used the ARIMA model to predict future rainfall trend. They showed that, ARIMA model can potentially detect future rainfall regime. (Dizon 2010) used ARMA model for reservoir modeling. (Bari et al. 2015) predicted monthly rainfall at Sylhet Division of Bangladesh Using Seasonal ARIMA model.

The main objective of the study is to develop a valid stochastic ARIMA model to forecast rainfall intensity of Teknaf. Rainfall forecast for the several stations of Bangladesh has been done. However, forecasting of rainfall of Teknaf has not been done yet. Beside Teknaf situated in coastal regions. Inundation occurs after some heavy rainfall. Moreover, the area suffers from severe salinity. Rainwater could be an alternative source of drinking water. Hence it was necessary to build up a forecasting model for a proper water resource management.

2. Study area
Teknaf is the southernmost region of Bangladesh. Teknaf is located at 20.8667°N to 92.3000°E. It is bounded by Ukhia upazila in the north, the Bay of Bengal in the south and west and Myanmar in the east. It has covered a total area of 388.66 km^2, Reserve forest area 227.6 km^2 and riverine areas of 1.36 km^2 (BBS 2013). Teknaf is one of the most cyclone prone areas of Bangladesh. Annual average highest temperature is 27.05^{0}C (April) and annual average lowest temperature is 19.56^{0}C (January). About 70% of the annual average rainfall occur between June to August.

3. Materials and Methods
Bangladesh meteorological Department (BMD) has 36 meteorological stations all over the country. Monthly rainfall data from 1981-2013 for Teknaf station were collected from BMD.

Quality control of data is a necessary step before handling with climatic parameter. The purpose of the quality control is to identify errors in data processing(Mortuza et al. 2013). Some measure has been adopted for this purpose, such as identification three or more successive months with the same quantity of rain, the precipitation value below 0mm, seasonal rainfall less than threshold value. Missing values have been screened before further analysis.

The inhomogeneity of the data series has been identified using four statistical methods, viz. Standard Normal Homogeneity Test (SNHT) for a single break (Alexandersson 1986), Von Neumann Ratio Test (Von Neumann 1941) Buishand Range test (Buishand 1981) and Pettitt test (Pettitt 1979). The data set was found homogenous. So no alteration of the data was required.

3.1. ARIMA model development

In the present study seasonal ARIMA model, proposed by (Box and Jenkins 1976) has been used for model building and forecasting of rainfall. The Box Jenkins SARIMA model has some advantages over other model due to its forecasting capabilities and richer information on time related changes (Mishra and Desai 2005). The SARIMA (Seasonal ARIMA) method also has a systematic approach for forecasting.

The short notation the SARIMA model described as ARIMA (p, d, q) * (P, D, Q) $_s$ where (p, d, q) is the nonseasnal part and (P, D, Q)$_s$ is the seasonal part of the model. Which could be written as:

$$\varphi_p(B)\Phi_p(B^s)\,\nabla^d\,\nabla_s^D z_t = \theta_q(B)\Theta_Q(B^S)a_t \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where, 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 = seasonal length (Box, et al., 1994).

$\Phi p(B)$ *AR polynomial of B of order p,* $\Phi p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$

$\Theta q(B)$ *MA polynomial of B of order q,* $\theta q(B) = 1 - \vartheta_1 B - \vartheta_2 B^2 - \ldots - \vartheta_q B^q$

$\Phi P(Bs)$ *Seasonal AR polynomial of BS of order P,* $\Phi p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{s2} - \ldots - \Phi_p B^{sp}$

$\Theta Q(Bs)$ *Seasonal MA polynomial of BS of order Q,* $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{s2} - \ldots - \Theta_Q B^{sQ}$

Δ = Differencing operator $\Delta = (1 - B)^d\,(1 - B^s)^D$

B =Backward shift operator with $B\,Y_t = Y_{t-1}$ and $B\,a_t = a_{t-1}$

The ARIMA model consists of four systematic stages (identification, estimation, diagnostic check and application or forecast). The identification stage involves with the improvement of the stationarity and normality of the data. At this stage, the general form of the model has been estimated. Model parameters are estimated using the method of maximum likelihood. Then the diagnostic checks were performed to reveal the possible inadequacies and to select the best model. Ultimately, the prediction of the rainfall time series has been executed.

4. Result and discussion:

The forecasting procedure could be seriously deficient if the model either inadequate or unnecessarily prodigal with the use of parameter (Box and Jenkins 1976). So the selected model should be adequate and parsimonious.

4.1. Model Identification

The first step of the ARIMA method is model identification. This step target to improve the stationarity and normality of the time series data. The time series plot, ACF and PACF has been (Figure: 1) examined to check stationarity and seasonality.

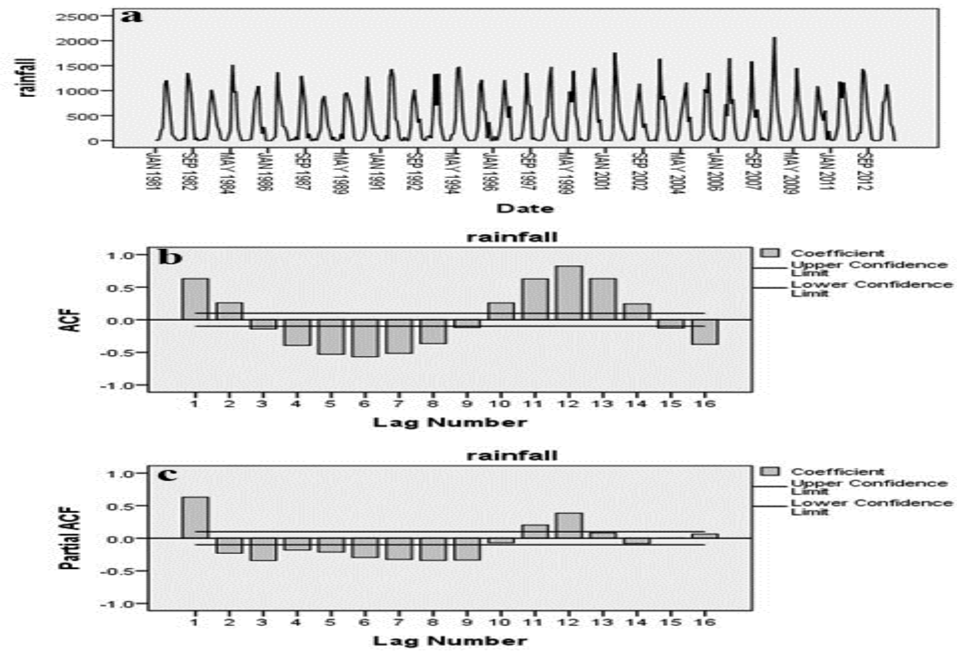

Figure 2: time series (a), ACF (b) and PACF (c) plot of rainfall for Teknaf station

The Figure 1 indicate a clear seasonality. Thus a first order seasonal differencing has been performed to remove seasonality. The ACF and PACF (Figure 2 :) of the transformed series shows that a first order seasonal differencing is adequate.

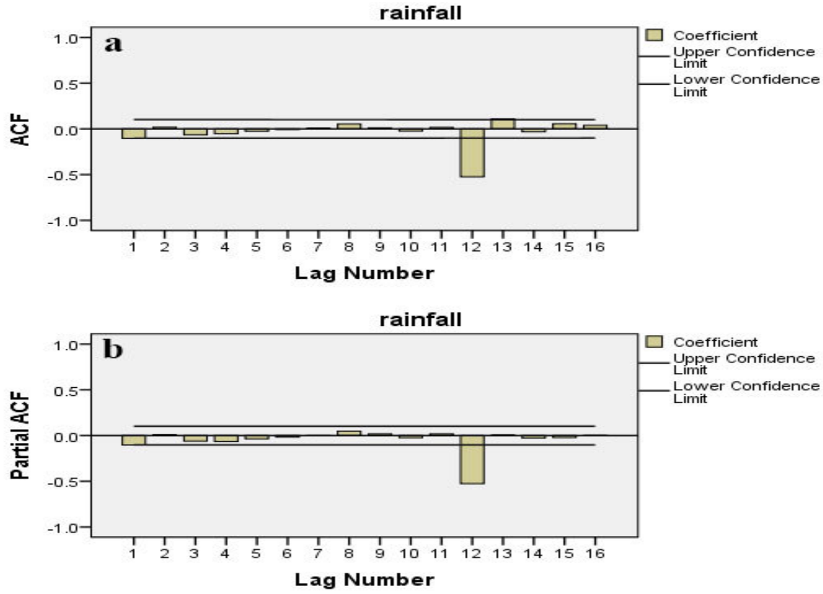

Figure 3: ACF and PAF plot of rainfall transformed series (Teknaf)

The autocorrelation is negative in the first lag, indicate that the series is a little bit over differentiate. A significant spike is present in the 13th lag of autocorrelation function. The PACF is damping out with a significant negative spike in the first lag. Which indicate the model could be a combination of AR and MA process. So the possible SARIMA model with P=0-2, Q=0-2 and p =0-1, q= 0-1 can be preliminarily tested. All the possible combination were tried to find out the best model. The model that yields the minimum Normalized BIC and maximum R-squared is selected as the best fit model. The identification of best model for rainfall series on the basis of R-squared value and Normalized BIC criteria is shown in table1. Based on the above mentioned criterion Seasonal ARIMA (0, 0, 0), (1, 1, 1) has been found best model.

Table 1: comparison of R-squared and Normalized BIC for selected candidate models

model	R-squared	Normalized BIC
ARIMA(0,0,0)(0,1,1)	.880	10.265
ARIMA(0,0,0)(1,1,1)	.888	10.232
ARIMA(1,0,0)(0,1,1)	.883	10.257
ARIMA(1,0,1)(0,1,1)	.883	10.272
ARIMA(1,0,1)(0,1,0)	.839	10.704
ARIMA(0,0,1)(0,1,1)	.882	10.259
ARIMA(0,0,1)(0,1,2)	.884	10.261

4.2. Parameter Estimation:

The preliminary estimation of parameter is done by using the primary values, estimated on identification stage. Parameter estimation is done by maximum likelihood method. The value of parameters, standard error, t-ratios and p-value are shown in table 2: standard error calculated for the model parameter were smaller than the associated parameters. Hence, for the maximum case the estimated parameter is statistically significant.

Table 2: statistical analysis of model parameter

Model	Parameter		Estimate	SE	t	Sig.(p) <0.05
ARIMA(0,0,0)(1,1,1)	AR, seasonal	Lag 1	-.170	.056	-3.066	.002
	MA, Seasonal	Lag 1	.968	.064	15.241	.000

4.3. Diagnostic Check

After the selection process diagnostic check has been performed to check out the model adequacy. All the validation checks has been implemented on residuals. For a good forecasting model the residuals should be white noise (Box and Jenkins 1976). The diagnostic checks are summarized as below:

4.3.1. ACF and PACF

The ACF and PACF (Figure 3a) of residuals shows that the values are randomly distributed and maximum values lie within the confidence limit except at lag 24. Which indicates that the residuals are white noise.

4.3.2. Histogram of residual

Histogram of the residual of rainfall series for Teknaf shown in Figure 3b: indicate that the residuals are normally distributed. This also implies the residuals are white noise.

4.3.3. Normal Probability of Residual

The cumulative distribution for the residual data normally appears as a straight line when plotted on normal probability paper(Chow, Ven T., David R. Maidment 1988). The normal probability plot (Figure 3c :) shows that the residual is linear enough.

4.3.4. Residual vs Prediction

Residual vs prediction plotted in Figure 3d: exhibit the residuals are evenly distributed around the mean indicating a well fitted model.

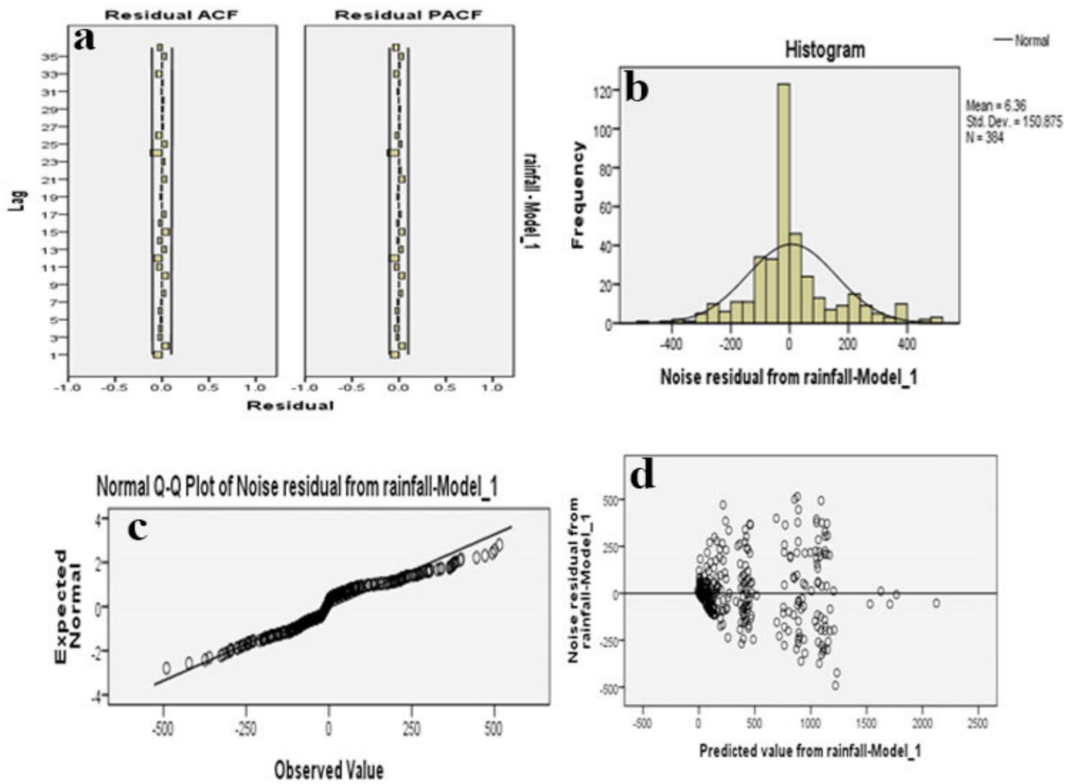

Figure 4: diagnostic check for best fitted model for rainfall series of Teknaf

4.3.5. Lack of Fit test

The modified Ljung Box statistics proposed by (Ljung and Box 1978) is used to check the null hypothesis, that the model is correctly specified. The test statistics and associated p-value in table 3 specify the model is well fitted.

Table 3: Lack of fit (L-jung Box) test for rainfall series of Teknaf

		Ljung-Box Q(18)	
model	Statistics	DF	Sig.
ARIMA(0,0,0)(1,1,1)	17.354	16	363

4.4. Rainfall forecast

The forecast has been done for 12 months using the best fit model. The plot between observed and predicted data

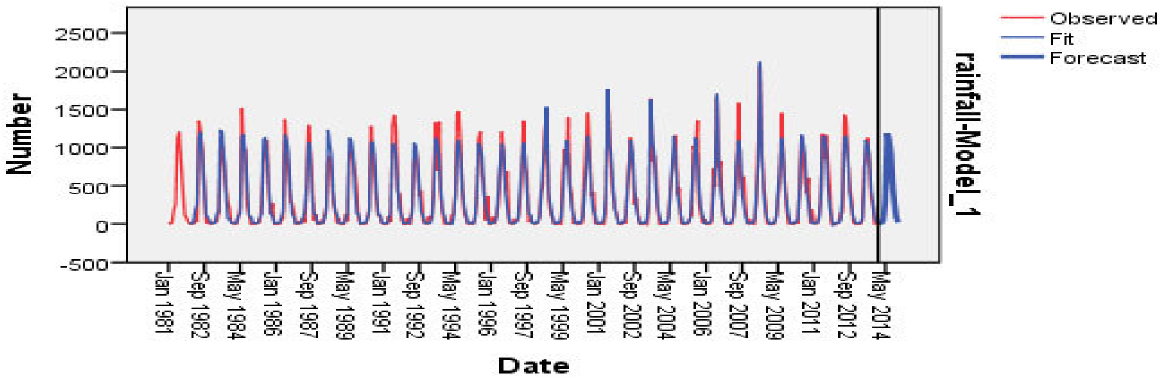

Figure 5: comparison of observed data with predicted data using best ARIMA model

(Figure 4) shows that the prediction is close enough to the observed value. The performance criteria shows in table 4 for the selected model, which indicate the model is good enough and it could be used to rainfall forecast.

Table 4: performance criteria of selected best fit model for rainfall sereis

model	R-squared	Model Fit statistics		
		RMSE	MAPE	Normalized BIC
ARIMA(0,0,0)(1,1,1)	.888	153.021	143.861	10.232

5. Conclusion

Rainfall is one of the most important and creeping phenomena than other climatic event. It became difficult to predict rainfall in a region where rainfall variability is higher. Like many other coastal area forecasting rainfall of Teknaf has great importance in planning and optimal operation of irrigation system as agriculture is the primary activity in the area. This work focused on rainfall forecast using seasonal ARIMA model for one year lead time. The stochastic model developed to forecast rainfall found effective. Hence it is recommended that stochastic model can be used in Teknaf and similar area to forecast rainfall. This model developed for Teknaf can be applied to determine possible future strategy and to develop sustainable water resource planning.

Reference

Alexandersson, Hans. 1986. “A Homogeneity Test Applied To Precipitation” 6: 661–75.

Bari, S H, M T Rahman, M M Hussain, and Sourav Ray. 2015. “Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model” 7 (1): 69–78.

BBS. 2013. “District Statistics 2011 Cox’s Bazar.” Dhaka. http://www.bbs.gov.bd/WebTestApplication/userfiles/Image/District Statistics/Cox`s Bazar.pdf.

Box, George E P, and Gwilym M Jenkins. 1976. *Time Series Analysis: Forecasting and Control, Revised Ed.* Holden-Day.

Buishand, Tjerk Adriaan. 1981. *The Analysis of Homogeneity of Long-Term Rainfall Records in the Netherlands*. KNMI.

Chow, Ven T., David R. Maidment, and Larry W. Mays. 1988. *Applied Hydrology*. New York: Mcgraw-Hill Book Company.

Dizon, Cornelio Q. 2010. “Arma Modeling of a Stochastic Process Appropriate for the Angat Reservoir” 28 (1): 1–20.

Gerretsadikan, Amaha, and M K Sharma. 2011. “Modeling and Forecasting of Rainfall Data of Mekele for Tigray Region (Ethiopia) Material and Methods Material Methodology” 9 (1976): 31–53.

Johnson, Lynwood A, and Douglas C Montgomery. 1976. *Forecasting and Time Series Analysis*. New York, Mc Graw Hill Book Company.

Ljung, G M, and G E P Box. 1978. “On a Measure of Lack of Fit in Time Series Models.” *Biometrika* 65 (2). Biometrika Trust: 297–303. doi:10.2307/2335207.

Mishra, a. K., and V. R. Desai. 2005. “Drought Forecasting Using Stochastic Models.” *Stochastic Environmental Research and Risk Assessment* 19 (5): 326–39. doi:10.1007/s00477-005-0238-4.

Mortuza, Md. Rubayet, Saranah Selmi, Md. Mohsan Khudri, Asif Khan Ankur, and Md. Mujibur Rahman. 2013. “Evaluation of Temporal and Spatial Trends in Relative Humidity and Dew Point Temperature in

Bangladesh.” *Arabian Journal of Geosciences* 7 (12): 5037–50. doi:10.1007/s12517-013-1139-3.

Pettitt, A N. 1979. “A Non-Parametric Approach to the Change-Point Problem.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (2). Wiley for the Royal Statistical Society: 126–35. doi:10.2307/2346729.

Von Neumann, John. 1941. “Distribution of the Ratio of the Mean Square Successive Difference to the Variance.” *The Annals of Mathematical Statistics* 12 (4). JSTOR: 367–95.

Zakaria, Saleh, Nadhir Al-ansari, Sven Knutsson, and Thafer Al-badrany. 2012. “ARIMA Models for Weekly Rainfall in the Semi-Arid Sinjar District at Iraq.” *Journal of Earth Sciences and Geotechnical Engineering* 2 (3): 25–55.