TIME-SERIES ANALYSIS OF MONTHLY RAINFALL DATA FOR THE NUWARA ELIYA DISTRICT, SRI LANKA

NM Udumulla[1], MGAK Madhubashini[2], NM Hakmanage[3], IM Sandaruwani[4], DDM Jayasundara[5], RMKT Rathnayaka[6]

[1]Department of Statistics & Computer Science, University of Kelaniya, Kelaniya 11600, Sri Lanka
[6]Department of Physical Sciences & Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Srilanka.
[1]*nirangam2@gmail.com*

Abstract - One of the most important climatic element is Rainfall, that directly influences on the agriculture. As a real world practice, the rainfall data has a seasonal trend with short term and long term oscillations; especially, monthly rainfall forecasting is significant to make decisions in management of agricultural scheme and manage daily human activities.

In this current study monthly average rainfall of Nuwara eliya district, Sri Lanka since 1996 to 2015 were considered. According to the unit root study, original observations are non-stationery. So, we moved 1st difference for further study. Furthermore, Seasonal Autoregressive Integrated Moving Average (SARIMA) model was fitted for analyzing and validation 228 monthly observations. According to minimum Akaike information criteria (AIC), SARIMA (1, 0, 0) (1, 0, 2) (12) is selected as a best model for forecasting rainfall in this selected region.

Keywords - AIC, Rainfall, SARIMA models, Unit root

I. INTRODUCTION

Going back to the association of water, it is the source of all life on Earth necessary for maintaining daily life activities. Rainfall is a most significant way for maintaining the land water level. However, the need for accurate rainfall predictions is readily apparent when considering the many benefits such information would provide for river/riverbed control, reservoir operations, flash flood watches, etc. Furthermore, among climatic factors, rainfall is considered as the one of the most important climatic factor that affect in agriculture field. As a result, not only to planning and management of agricultural scheme but also to management of water resource systems, monthly rainfall forecasting plays a major role.

Harrison et.al (2014) carried out a study to forecast monthly rainfall data for the function relates to the level data in oscillatory of 12 periods. Furthermore, they adjudged stationary by the Augmented Dickey Fuller unit root test and correlogram gives an indication of stationarity as well as an involvement of the presence of a seasonal moving average component of order one and a seasonal autoregressive component of order two. These fundamental conditions with autocorrelation structure suggested a three multiplicative seasonal autoregressive Gadaref rainfall station in Sudan by using SARIMA methods. Their autocorrelation integrated moving average (SARIMA) models, namely: (0, 0, 0) (0, 1, 1)12 , (0, 0, 1) x(0, 1, 1)^{12} and (0, 0, 1)x(2, 1, 1)^{12} .

C. C. Nnaji et.al (2013) carried out a similar study based on SARIMA to forecast Nigeria's rainfall patterns. The Box-Jenkins methodology was used to build ARIMA model for the period from November 2003 to October 2013 with a total of 120 data points and suggested that SARIMA (1, 1, 1) (0, 0, 1)^{12} is best for forecasting rainfall data under the seasonality. In a similar study, develop a trend analysis to

Figure 1: Map of Nuwara eliya District

forecast rainfall deficit in the Blackland Prairie of Eastern Mississippi for a period of 120 years of annual, seasonal, and monthly periods.

Brath, et.al (1997) used time series analysis techniques in order to predict short-term rain fall for flood forecasting. For this study, linear stochastic Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) models, were used to express the future rainfall as a linear function of past data. By considering about the results we can see that the use of time series analysis techniques for rainfall forecasting may allow an extension of the lead-time up to which a reliable flood forecast may be issued, introducing a very quick prediction based on past data directly in the format required by the rainfall-run off transformation model. The current study attempted to develop linear stochastic models such as multiplicative seasonal autoregressive integrated moving average (SARIMA) and GRACH models for forecast monthly rainfall in Nuwara Eliya, Sri Lanka. The main objective of this study is to perform an analysis the monthly characteristics within the monsoon seasons and to show how these characteristics can be linked to the general circulation such that physical meaning for the cause of rainfall variability can be found. Furthermore, using 21-year period data, aims of this study is ascertaining the presence and range of Self-Organized Criticality in the monthly rainfall pattern of Nuwara eliya.

II. MATERIALS AND METHODS

In this research, our main objective was to find a model to efficiently forecast the monthly rainfall in Nuwara Eliya by applying Box and Jenkins method.

a. Data

For this study, average monthly rainfall data of Nuwara eliya district was considered from period 1996-2015. The rainfall gauge contains 240 observations. We have used 228 observations for build appropriate of Auto regressive moving average model and 12 observations are rest for validation procedure. The average monthly data of Nuwara eliya district show the highest density of the rainfall in period June- November. Thus most of the peaks were reached in October and November months.

b. Modeling by SARIMA

The combination of autoregressive (AR) and moving average (MA) models called ARMA which should be used for forecasting stationery time series. If the time series observations are non-stationery, then we can allow the stationery condition using differencing. Thus the general non seasonal model of ARIMA is shows with 3 parameters such as ARIMA (p, d, q) where p is the order of AR and q is the order of MA and d is the differencing order.

Non seasonal ARIMA model can be written as

$$\emptyset_p(B)\nabla^d Y_t = \theta_q(B)\varepsilon_t \quad \text{-----}(1)$$

Where $\emptyset_p$ and θ_q are the polynomials of order p and q respectively and ε_t is the white noise term. If the time series is with seasonally, the Box-Jenkins (1976) proposed a SARIMA model. In short SARIMA model described as ARIMA (p, d, q)×$(P,D,Q)_s$ which is

$$\emptyset_p(B)\varphi_p(B^s)\nabla^d\nabla^d{}_sY_t = \theta_q(B)\theta_q(B)_s\varepsilon_t \text{ ......} \quad (2)$$

Where p and q is the non-seasonal autoregressive and moving average orders and d is the number of regular differencing and P and Q is the seasonal autoregressive and seasonal moving average orders respectively. Also, D is the seasonal differencing order. It is based on the least squares optimization criteria.

c. Model Selection

The fitting of model (2) begins with order selection. The seasonality period can detect using time series plot or correlogram of the object. For this section, rainfall is seasonal time series with s=12 months. The non-seasonal and seasonal AR orders p and P are estimated by the non-seasonal and the seasonal cut-off lags of the partial autocorrelation and similarly, Q and q can determine using seasonal and non-seasonal ACF cut-off points. The minimum **Akaike information criteria (AIC)** values gives the most effective models. For this study, the econometric software Eviews-8 and R-software was used for all analytical work.

d Model Validation

Model validation provides the efficiency of the model while forecasting. Model validation totally based on mean absolute percentage error (MAPE). The value of MAPE classify as follows.

Table 5: MAPE classification

MAPE	Judgment of Accuracy
< 10%	Highly Accurate
11% to 20%	Good Forecast
21% to 50%	Reasonable Forecast
> 51%	Inaccurate forecast

III. RESULTS & DISCUSSION

e. Preliminary analysis

Fig. 1 shows that the time series plot of the original observations from period 1996-2015. The data patterns indicated that series is stationary and the mean of the time series is relatively constant with respect to the time. We were performed unit root test and evaluated the original series are stationery over time. Unit root test results are given below.

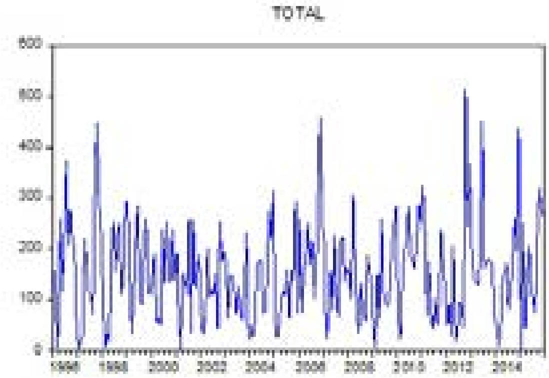

Figure 2-Time series plot of the original observations

Table 2: Unit root test

Null Hypothesis: TOTAL has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-11.69503	0.0000
Test critical values:	1% level	-3.457630	
	5% level	-2.873440	
	10% level	-2.573187	

*MacKinnon (1996) one-sided p-values.

After testing unit root test, we were concluded original series is stationery. Obviously rainfall data shows seasonal time series with s = 12 months. Figure 2 shows the clear seasonal pattern in 12 months' period. Since data should be seasonal in 12 months' time period.

Thus Figure 3 Shows the ACF graph of observations.

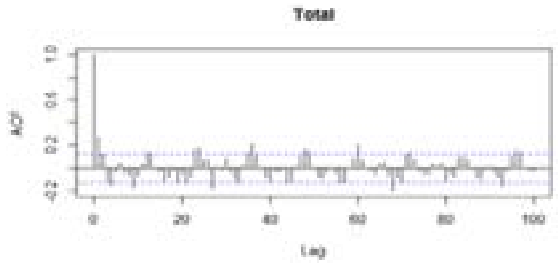

Figure 3-ACF plot of observations

Table 5: Hetroskedasticity test

Hetroskedasticity Test: ARCH

F-statistic	0.131918	Prob. F(1,212)	0.7168
Obs*R-squared	0.133080	Prob. Chi-Square(1)	0.7153

f. Model Selction

Table 3: model selection Results

Model	AIC
SARIMA(1,0,0)(1,0,1)[12]	11.67333
SARIMA(1,0,0)(1,0,2)[12]	11.66028
SARIMA(2,0,0)(0,0,1)[12]	12.066

Next step is the model selection. Thus we were selected most suitable model for the rainfall forecasting. Model selection was totally based on minimum AIC value. The Table 3 shows the AIC values of selected models and suggested that SARIMA (1,0,0) (1,0,2) (12) model is best for future predictions.

g. Diagnostic checking Recidual Autocorrolation test

After selecting a best model, checked the residual to conform the significant autocorrelation using a Breusch-Godfrey serial correlation LM test. The Table 4 Breusch-Godfrey Serial Correlation LM Test results conclude that residual has no autocorrelation.

Table 4: Breusch-Godfrey Serial Correlation LM Test results

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	2.115486	Prob. F(2,209)	0.1231
Obs*R-squared	4.262026	Prob. Chi-Square(2)	0.1187

Thus we were test whether residual contain constant variance or not. We were used hetroskedasticity test to test this assumption. Table 5 shows the results of Breusch-Godfrey Test.

Table 5 shows the homoscedasticity test results concluded that residual has constant variance. Usually we say that residual has homoscedasticity. Thus also this results implies that residual haven't ARCH effect.

VI. CONCLUSION

This current study explored the presenting time series analyses for average monthly rainfalls in Nuwara Eliya district. In presence of strong seasonality in every 12 months, we used the SARIMA model for forecasting the average monthly rainfalls in Nuwara eliya district from January 1996 to December 2015. The minimum AIC criterion results suggested SARIMA (1, 1, 0) (1, 0, 2) [12] is a best model. Model MAPE is 4.4%. It may be used for forecasting process and the forecasted values can be used for planning and managing future decisions of this region. This model is considered appropriate to predict the monthly rainfall for the upcoming years to assist decision makers establish priorities for water demand, storage, distribution and cultivations.

V. ACKNOWLEDGEMENT

without getting supports from a lot of individuals we cannot complete successful paper. Either directly or indirectly individuals who help in this research project, we will take this opportunity to thank all of them. First of all, we wish to express our sincere gratitude and due respect to our supervisor Dr. R.M. Kapila Tharanga Rathnayaka lecturer in department of Physical science and Technology, Sabaragamuwa University of Sri Lanka and Dr.D.D.M. Jayasundara, head of the department of statistics and computer science university of Kelaniya. We are extremely grateful to them for valuable guidance, continuous encouragements and positive supports which helped us a lot during the period of our work. Last but not the least we are also thankful to our faculty staff and staff of department of statistics and computer science for their unselfish help.

VI. REFERENCES

Babazadeh, H. and Shamsnia, S.A., 2014. Modeling climate variables using time series analysis in arid and semi arid regions. African Journal of Agricultural Research, 9(26), pp.2018-2027.

Brath, A., Montanari, A. and Toth, E., 1999, October. Short-term rainfall prediction with time series analysis techniques for real-time flash flood forecasting. Mediterranean storms, Proceedings EGS Plinius Conference, Maratea (PZ).

Ette Harrison Etuk1*, T. M. M., 2014. Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by Sarima Methods. International Journal of Scientific Research in Knowledge, pp. 320-327.

Mohamed, E. E.-T., 2014. Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by Sarima Methods. Issue International Journal of Scientific Research in Knowledge

S. Soltani, a. M. a. S. S. E., 2007. The use of time series modeling for the determination ofrainfall climates of Iran. International journal of climatology.