Available online at http://www.journalcra.com

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

International Journal of Current Research
Vol. 6, Issue, 02, pp.4986-4990, February, 2014

ISSN: 0975-833X

RESEARCH ARTICLE

STATISTICAL MODELING OF TAMIL NADU ANNUAL RAINFALL

*[1]Mahalakshmi, N. [2]Umarani, P. R. and [3]Samuelselvaraj, R.

[1]Anna Adarsh College for Women, Chennai- 600 040
[2]Kunthavai Naacchiaar Govt. Arts College for Women, Thanjavur- 7
[3]Department of Physics, Presidency College, Chennai 600 005

ARTICLE INFO

Article History:
Received 05th November, 2013
Received in revised form
08th December, 2013
Accepted 26th January, 2014
Published online 21st February, 2014

Key words:
SARIMA,
Mann Kendall,
Dickey Fuller Test,
Annual Rainfall,
Tamil Nadu.

ABSTRACT

The annual rainfall data for a period of 63 years (1950-2012) of Tamil Nadu have been analyzed for climatologically pattern and trend. Climate change for a region can be analyzed only on a long–term average. Detecting climate change with seasonal variability is difficult, thus we have considered the annual rainfall to analyze any change in the climatic trend of Tamil Nadu and forecast the rainfall for the year 2013. The non parametric Mann-Kendall test is used for trend analysis. This resulted with h=0 and p-value = 0.4609 for an alpha value of 0.05. These values do not give us sufficient evidence to reject the null hypothesis (no trend) for the Tamil Nadu annual rainfall series. This result has led us to estimate the stationarity of the rainfall data. Dickey fuller unit root test was done to study the stationarity in the rainfall data. The result h=1 resulted in rejection the presence of stationatiry in the data series. Thus the rainfall data is a non-stationary series. With these preliminary statistical analysis attempt was made to predict the annual rainfall for the year 2013 for Tamil Nadu on a monthly basis. The Box-Jenkins time series seasonal ARIMA (Auto Regression Integrated Moving Average) approach was used for the forecast of annual rainfall on monthly scales. Past years data was used to formulate the seasonal ARIMA model and in determination of model parameters.

INTRODUCTION

The forecast of Tamil Nadu annual rainfall on a monthly scale is essential in planning and management of the state's resources like agriculture practices, flood management, drought management and also in balancing the economy of the state. The mitigation of the water resources, its effective planning and management thus becomes vital. The major source for the water resource is the precipitation. In short the forecast of the precipitation of a region enables the people of the state to manage their resource. A good forecast needs to project the realistic state of the current scenario and its future variability and change. The forecast of rainfall using mathematical methods can be found in many literatures. Researchers have studied the state-time models which include complex formulation to forecast rainfall (Kourosh *et al*., 2006). A simple mathematical model with less complex formulation is used in this paper. Considerable amount of work has been dedicated in forecasting of the rainfall of a region using time series models. The forecast of rainfall of a region using its meteorological parameters is desirable. These parameters are not always available, thus we use the past and present values of the rainfall data and study its future. A time series data may contain the components of trend and seasonality. The trend is a component that will not show repeatability for a considered time period. The seasonality is the component which may repeat itself over a period of time. In the present paper we are using a most simplified yet a strong forecasting model called the Auto Regressive Integrated Moving Average (ARIMA). Inderjeet and Sabita, 2008 has provided a reliable prediction of the rainfall and temperature on a monthly scale for the state of Uttar Pradesh. The annual rainfall in Tamil Nadu was studied using suitable Box-Jenkin's method by M.Nirmala and S.M Sundaram, 2010. Langu, 1993 used a time series analysis to detect a change in rainfall to search for significant changes in the component of the rainfall time series. Janhabi and Ramakar, 2013, studied the rainfall at the Mahanadhi river basin using the ARIMA model.

***Corresponding author: Mahalakshmi, N.**
Anna Adarsh College for Women, Chennai- 600 040.

DATA

The data for the Tamil Nadu rainfall for a period of 1950 -2012 was procured from IITM'S official website http://www. tropmet.res.in. The data points were arranged on a monthly scale with a total of 744 readings. This data was used to design the best possible model of ARIMA.

MATERIALS AND METHODS

Before the modeling of the rainfall forecast preliminary statistical tests were carried out on the time series to test the presence of trend and stationarity in it. A simple linear regression analysis may provide a primary indication of the presence of trend in the time-series data. Other methods, such as the non-parametric Mann- Kendall (MK) test, which is commonly used for hydrologic data analysis, can be used to detect trends that are monotonic but not necessarily linear. In the present paper we follow the Mann Kendall test for the trend analysis. The MK test does not require the assumption of normality, and only indicates the direction but not the magnitude of significant trends (Yue, Wang, 2004). The trend in the data if any has been quantified using Mann-Kendall's S-statistic (Kendall, 1962). The MK method assumes that the time series under research are stable, independent and random with equal probability distribution. The MK test is applied to uncorrelated data because it has been reported that the presence of serial correlation might lead to an erroneous rejection of the null hypothesis (Kulkarni and Storch, 1995; Yue and Pilon, 2003). The Box-jenkins method is univariate time series analysis. It is thus essential to analyze the presence of the unit root in the time series. The stationarity of the time series can be tested with Augmented Dickey fuller test (Adenomon, *et al*., 2013, Abdul-Aziz 2013, Dominick and Derrick, 2002). This test is popularly used to check the presence of unit root in the Time series data. Most time series consist of elements that are serially dependent in the sense that one can estimate a coefficient or a set of coefficients that describe consecutive elements of the series from specific, time-lagged (previous) elements. Each observation of the time series is made up of random error components (random shock, ε) and a linear combination of prior observations.

ARIMA model

The ARIMA model is an extension of the ARMA model in the sense that by including auto-regression and moving average it has an extra function for differencing the time series. If a dataset exhibits long-term variations such as trends, seasonality and cyclic components, differencing a dataset in ARIMA allows the model to deal with them. A large amount of literature can be sited for the procedure of the ARIMA. In this paper a brief procedure of the Box and Jenkins method is discussed (Box and Jenkins, 1970). The use of ARIMA process and differencing are the basis for this method (Gurjarati *et al.*, 2007)

The procedure for the Box-Jenkins method involves three steps

1.Identification of the model structure
2.Parameter estimation.
3.Model testing and validity.

The time series is checked for the presence of stationarity, if non stationary it is removed by way of differencing. The order of the AR and the MA component is identified. The PAC determines the order of the AR process and the ACF method determines the order of the MA process. The next step is to identify the parameter of the identified AR and MA components. Model selection is important in the time series analysis. In general a number of ARIMA models for a time series is possible. One may use the Maximum likelihood rule or the Mean square error method to identify the best ARIMA model. The final step is to validate the model. A forecast equation is formed and the residue series from the model is used to test its validation. The residue series must have a zero mean with no periodicities and is should be uncorrelated. The significance of the residual mean, significance of periodicities and the white noise test are the possible model validation tests for the ARIMA. The general form of the seasonal ARIMA model is given.

The seasonal ARIMA(p,q,d)(P,Q,D)

$$\varphi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D y_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t$$

where

$$\varphi_p(B) = 1 - \varphi_1 B - \cdots - \varphi_p(B^P),$$
$$\Theta_q(B) = 1 - \theta_1(B) - \cdots \theta_q(B^q)$$
$$\Phi_P(B^s) = 1 - \Phi_1(B^s) - \cdots - \Phi_P(B^{sP})$$
$$\Theta_Q(B^s) = 1 - \Theta_1(B^s) - \cdots - \Theta_Q(B^{sQ})$$

ε_t denotes the error term, φ's and Φ's are the non seasonal and the seasonal autoregressive parameters and ϴ's and θ's are the non seasonal and the seasonal moving average parameters. At the forecasting stage, the estimated parameters were tested for their validity using error statistics such as coefficient of determination (R^2), mean square error (MSE), and mean absolute error (MAE) criteria.

RESULTS AND DISCUSSION

The Mann Kendall test for detecting the trend was carried out with Mat Lab soft ware. The output of the test was h=0, the p-value was 0.46, for a significance level of 0.05. Thus the value of h implies that the null hypothesis of no trend of time series data cannot be rejected. But the p-values is greater than the significant level which is in favor for us, as there is insufficient evidence to completely accept the null hypothesis of no trend in the time series. A time series is said to be stationary if its autocorrelation function (ACF) is essentially constant in time. Figure 1 shows the time series plot of the rainfall of Tamil Nadu.

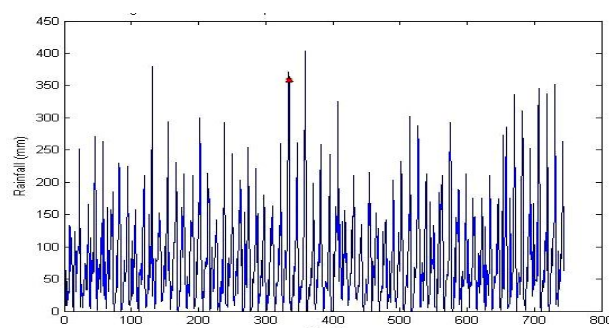

Figure 1. Time series plot of Tamil Nadu Rainfall data

It is evident from the plot that the rainfall of Tamil Nadu is not stationary and a very meager trend is present in the series. The non-stationarity of the time series can also be verified from the auto correlation function of the data. The result of the Dickey Fuller test suggest that there no stationarity in the time series data. The value h=1 resulted in rejecting the null hypothesis of presence of unit root in the series. Thus the series is stationarised by differencing or by standardizing. We have used the method of differencing to make the time series stationary. Figure 2 shows the ACF plot for the time series considered. The correlogram for the stationary time series will decay soon, if the time series is not stationary it will decay slowly. Figure 3 shows the ACF correlogram for a differenced series of the Tamil Nadu annual Rainfall. From Figure 3 the decay of the correlogram can be noticed and the pattern is repeated every 12 months. Thus we also employ seasonality in the model. The order of the MA and the AR parameters are then decided based on the ACF and PACF plot of the differenced rainfall data. Figure 4 shows the ACF and PACF plot of the data. From the correlogram the best model for ARIMA was chosen and the model with the decided AR and MA parameters were fed in to SPSS software. The forecast values are shown in Table 1 and the output of the SPSS for ARIMA model is given in Table 2.

Table 1. Forecast values for Tamil Nadu Annual Rainfall for the year 2013

	Model $(5,1,2)(0,1,1)^{12}$
January	20.07
February	21.40
March	23.38
April	44.42
May	72.37
June	48.39
July	68.32
August	89.98
September	114.89
October	194
November	198.95
December	93.47

Table 2. Model statistics for SARIMA (5, 1, 2) (0, 1, 1)

Model	Model Fit statistics			
	Stationary R-squared	R^2	RMSE	MAPE
(5,1,2)(0,1,1)	0.731	0.556	49.469	686.875

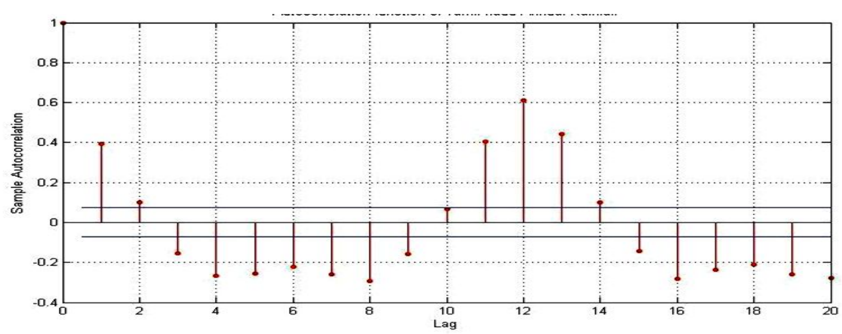

Figure 2. Autocorrelation function of original time series data

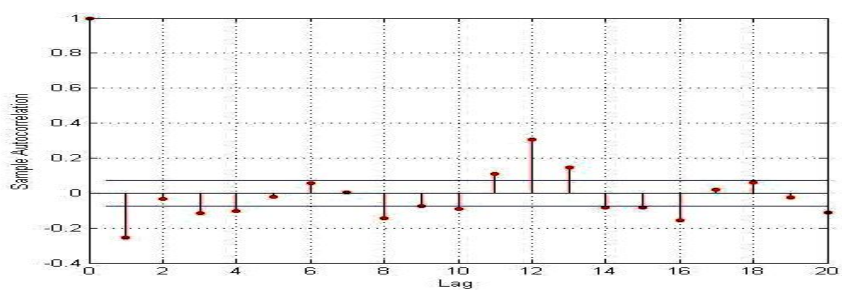

Figure 3. Autocorrelation function for differenced time series data

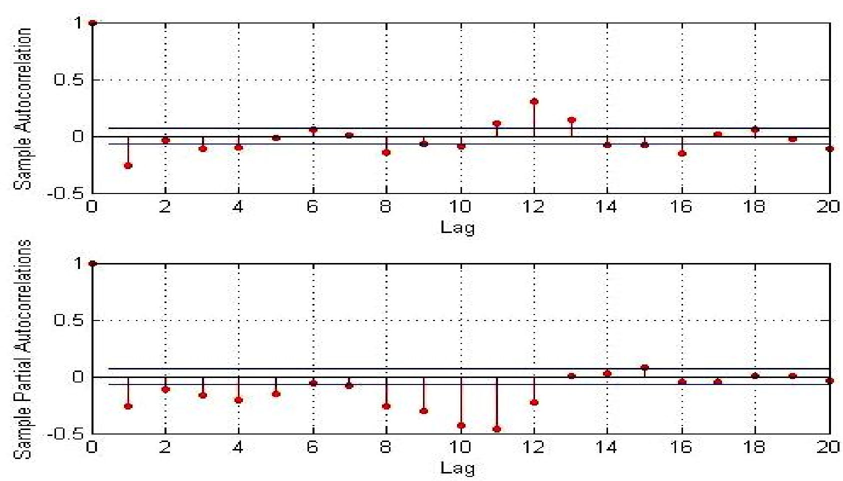

Figure 4. ACF and PACF of differenced Annual Rainfall

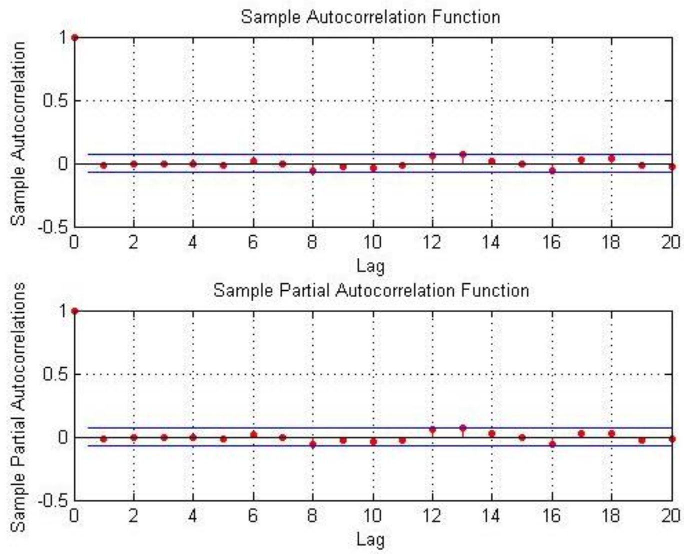

Figure 5. Residual ACF and PACF plot for the selected ARIMA

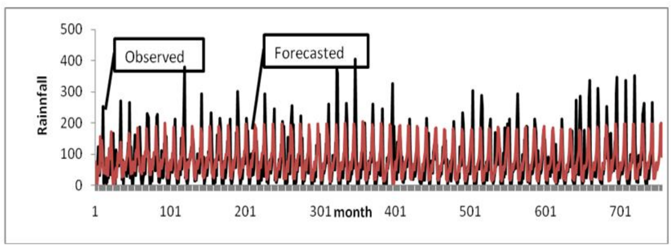

Figure 6. Observed and forecasted values for Tamil Nadu

Model Validation

After fitting the appropriate ARIMA model, the goodness of fit can be examined by plotting the ACF of residuals of the fitted model. If most of the autocorrelation coefficients of the residuals are within the confidence intervals then the model is a good fit. Since the coefficients of the residual plots of ACF and PACF are lying within the confidence limits, the fit is a good fit. The graph showing the observed and fitted values is shown in Figure 5. Figure 6 shows the graph plotted for the observed values and the fitter values.

Conclusion

The Tamil Nadu rainfall time series data was modeled using the seasonal ARIMA model; the best fit model was selected using the SPSS software. The value of statistics 0.556 also shows that the fit model is good. The residual ACF and PACF plots also ensure the goodness of the model selected. The ARIMA model though very simple it is one of the most popular methods in forecasting the rainfall of region.

REFERENCES

Abdul Aziz, A.R., M. Anokye, A. Kwame, L. Munyakazi and N.N.N. Nsowah Nuamah,Modeling and Forecasting Rainfall Pattern in Ghana as a Seasonal Arima Process-The Case of Ashanti Region, *International Journal of Humanities and Social Science* Vol. 3 No. 3; February 2013.

Adenomon, M. O, Ojehomon, V.E.T and Oyejola, B. A. 2013. Modelling The Dynamic Relationship Between Rainfall and Temperature Time Series Data In Niger State, Nigeria. *International Journal of Humanities and Social Science Vol. 3 No. 3.*

Box, G. E. P., and G. M. Jenkins. 1970. *Time-series analysis: Forecasting and control*. Rev. ed. San Francisco: Holden-Day.

Dominick, S. and Derrick, R. 2002. *Schaum's Outline of the Theory and Problems of Statistics and Econometrics (2nd ed).* New York: McGraw-Hill Company.

Engle, R. F., and Granger, C. W. J. 1987. Cointegration and Error Correction: Representation, Estimation and Testing. Econometrica. 55(2):251-276.

Inderjeet Kaushik and Sabita Madhvi Singh, 2008,Seasonal Arima Model For Forecasting Of Monthly Rainfall And Temperature, *Journal Of Environmental Research And Development* Vol. 3 No. 2. 506-514.

Janhabi Meher, Ramakar Jha, 2013,Time-series analysis of monthly rainfall data for the Mahanadi River Basin, India Sciences in Cold and Arid Regions, 5(1): 0073–0084. DOI: 10.3724/SP.J.1226.2013.00073.

Kendall MG. 1962. *Rank Correlation Methods*. Hafner Publishing Company: New York.

Kourosh Mohammadi, H.R. Eslami, Rene Kahawita, 2006, Parameter estimation of an ARMA model for river flow forecasting using goal programming, *Journal of Hydrology*, 331, 293– 299.

Kulkarni A, von Storch H 1995. Monte Carlo experiments on the effect of serial correlation on the Mann–Kendall test of trend. Meteorol Z 4(2):82–85.

Langu, E.M., 1993. Detection of changes in rainfall and runoff patterns. *J. Hydrol.*, 147: 153-167. http://cat.inist.fr/?aModele=afficheN&cpsidt=4788801.

Nirmala M., S. M. Sundaram, 2010, Modeling And Predicting The Monthly Rainfall In Tamilnadu As A Seasonal Multivariate Arima Process, *International Journal Of Computer Engineering And Technology* (Iissn 0976 – 6375(Online) Volume 1, Number 1, May - June 2010.

Yue S., Wang, C., 2004. The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. *Water Resources Management* 18, 201–218.

Yue, S., Pilon, P., and Caradias, G., 2002. Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series, *J. Hydrol.*, 259, 254-271.

****