

# A SEASONAL ARIMA MODEL FOR FORECASTING MONTHLY RAINFALL IN TAMILNADU

Nirmala M.<sup>1</sup>, Sundaram S. M.<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Mathematics, Sathyabama University, Chennai, India

<sup>2</sup>Department of Mathematics, Sathyabama University, Chennai, India

Email: <sup>1</sup>monishram5002@gmail.com

## Abstract

Rainfall is natural climatic phenomena whose prediction is challenging and demanding. Worldwide numerous attempts have been made to predict its behavioral pattern using various techniques. The present study is to find a suitable seasonal forecasting model for forecasting monthly rainfall in Tamilnadu. For this purpose, a dataset containing a total of 136 years (1871 - 2006) monthly rainfall totals of Tamilnadu is obtained from Indian Institute of Tropical Meteorology (IITM), Pune, India and the suitable forecasting model is identified through Box – Jenkins SARIMA modeling.

**Key words:** Model, Forecasting, seasonality, stationarity, Box-Jenkins seasonal ARIMA model

## I. INTRODUCTION

The effects of climate change on various environmental variables have been widely observed in many regions around the world. Among these variables, rainfall is the most concerned climate-change-affected variable due to its nonhomogeneous distributions in time and in space.

The average rainfall series of Tamilnadu for the northeast monsoon months of October to December and the season as a whole were analysed for trends, periodicities and variability using standard statistical methods by O.N. Dhar, P.R. Rakhecha and A.K. Kulkarni. Also they analysed the trends, periodicities and variability in the seasonal and annual rainfall series of Tamilnadu. The association between the southwest and northeast monsoon rainfall over Tamilnadu have been examined for the 100 year period from 1877 – 1976 through a correlation analysis by O.N. Dhar and P.R. Rakhecha. The prediction of annual rainfall over Tamilnadu has been made using Auto Regressive Integrated Moving Average method by M. Nirmala and S.M. Sundaram. The local and teleconnective association between Northeast Monsoon Rainfall (NEMR) over Tamilnadu and global Surface Temperature Anomalies (STA) is examined using the monthly grid-ded STA data for the period 1901-2004 by Balachandran S., Asokan R. and Sridharan S. The modeling and the prediction of rainfall is done through the statistical method based on autoregressive

integrated moving average (ARIMA) and the emerging computationally powerful techniques based on ANN. In this paper, modeling and forecasting of monthly rainfall in Tamilnadu is made through the conventional method called Box – Jenkins Seasonal ARIMA model.

## II. STUDY AREA

Tamilnadu stretches between 8° 5' N and 13°35' N latitudes and between 78°18' E and 80°20' E longitudes (fig.1). It is bounded on the north by Andhra Pradesh and Karnataka, on the east by the Bay of Bengal, on the south by the Indian Ocean and on the west by the state of Kerala. The state of Tamilnadu receives rainfall in both the southwest and northeast monsoons. Agriculture is more dependant on the northeast monsoon. The entire Cauvery delta zone, which an important agro-climatic zone, depends primarily on the southwest monsoon. Hence, the fate of the agricultural economy of the state is decided by the monsoons. Tamilnadu, located in southeast Peninsular India receives the major part of its annual rainfall during the northeast monsoon (the three-month period from October to December). While coastal Tamilnadu receives about 60% of its annual rainfall, interior Tamil Nadu receives about 40–50% of annual rainfall during Northeast monsoon season. Approximately 33% of annual rainfall in Tamilnadu is from the southwest monsoon and 48% is from the northeast monsoon.

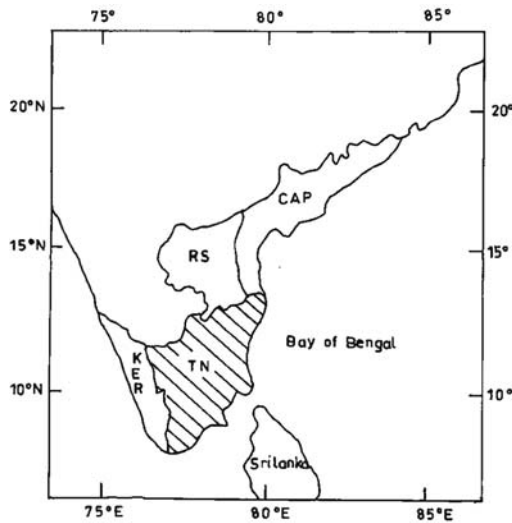


Fig 1: Geographical location of area of study, Tamilnadu

### III. DATA AND METHOD

A dataset containing a total of 136 years (1871 - 2006) monthly rainfall totals of Tamilnadu was obtained from Indian Institute of Tropical Meteorology (IITM), Pune, India.

#### A. Box-Jenkins Seasonal ARIMA Model:

Univariate time series analysis using Box-Jenkins ARIMA model is a major tool in hydrology and has been used extensively, mainly for the prediction of such surface water processes as precipitation and streamflow events. It is basically a linear statistical technique and most powerful for modeling the time series and rainfall forecasting due to ease in its development and implementation. The ARIMA models are a combination of autoregressive models and moving average models. The autoregressive models AR ( $p$ ) base their predictions of the values of a variable  $x_t$ , on a number  $p$  of past values of the same variable number of autoregressive delays  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$  and include a random disturbance  $e_t$ . The moving average models MA ( $q$ ) generate predictions of a variable  $x_t$  based on a number  $q$  of past disturbances of the same variable prediction errors of past values  $e_{t-1}, e_{t-2}, \dots, e_{t-q}$ . The combination of the autoregressive and moving average models AR ( $p$ ) and MA ( $q$ ) generates more flexible models called ARMA ( $p, q$ ) models. The stationarity of the time series is required for the implementation of all these models. In

1976, Box and Jenkins proposed the mathematical transformation of the non stationary time series into stationary time series by a difference process defined by an order of integration parameter  $d$ . This mathematical transformation transforms ARMA ( $p, q$ ) models for non stationary transformed time series as the ARIMA ( $p, d, q$ ) models, Autoregressive Integrated Moving Average models.

The ARIMA model building strategy includes iterative identification, estimation, diagnosis and forecasting stages. Identification of a model may be accomplished on the basis of the data pattern, time series plot and using their autocorrelation function and partial autocorrelation function. The parameters are estimated and tested for statistical significance after identifying the tentative model. If the parameter estimates does not meet the stationarity condition then a new model should be identified and its parameters are estimated and tested. After finding the correct model it should be diagnosed. In the diagnosis process, the autocorrelation of the residuals from the estimated model should be sufficiently small and should resemble white noise. If the residuals remain significantly correlated among themselves, a new model should be identified estimated and diagnosed. Once the model is selected it is used to forecast the monthly rainfall series. Time series analysis provides great opportunities for detecting, describing and modeling climatic variability and impacts. Ultimately, to understand the meteorological information and integrate it into planning and decision making process, it is important to study the temporal characteristic and predict lead times of the rainfall of a region. This can be done by identifying the best time series model using Box – Jenkins Seasonal ARIMA modeling techniques.

### IV. RESULTS AND DISCUSSIONS

A Time Series plot is used to display the time variation of one or more scalar. It is a graph showing a set of observations taken at different points in time and charted in a time series (fig. 2). These observations are usually successive and equally spaced in time intervals. At first the time series data is lag transformed. A good starting point for time series analysis is the graphical plot of the data.

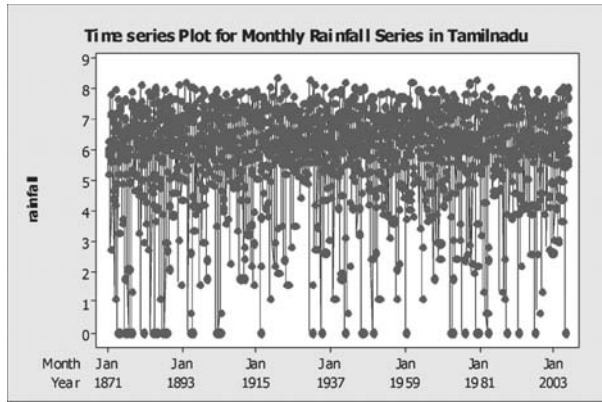


Fig. 2. Time series plot of Monthly Rainfall in Tamilnadu

Some of the statistical measures are tabulated in table 1 to understand the characteristics of the monthly rainfall series in Tamilnadu.

**Table 1: Descriptive Statistics of log transformed monthly rainfall series**

Mean	Std Deviation	Coeffi of Variation	Median
5.9562	1.6065	26.97	8.4109

An important guide to the properties of a time series is provided by a series of quantities called sample autocorrelation coefficients or serial correlation coefficient, which measure the correlation between observations at different distances apart.

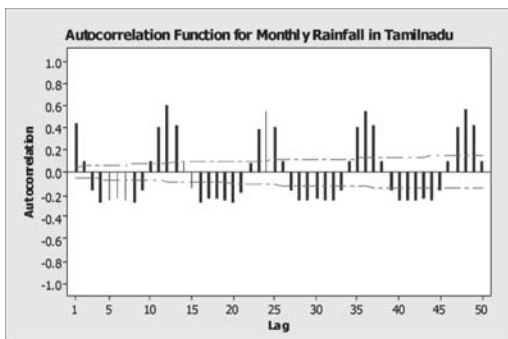


Fig. 3. The Autocorrelation function plot of Monthly Rainfall in Tamilnadu

These coefficients often provide insight into the probability model which generated the data. The sample autocorrelation coefficient is similar to the ordinary correlation coefficient between two variables

( $x$ ) and ( $y$ ) except that it is applied to a single time series to see if successive observations are correlated.

Given ( $N$ ) observations, on discrete time series we can form ( $N - 1$ ) pairs of observations. Regarding the first observation in each pair as one variable and the second observation as a second variable, the correlation coefficient between and is called autocorrelation coefficient of order one. The correlogram is a commonly-used tool for checking randomness in a dataset. This randomness is ascertained by computing autocorrelations for data values at varying time lags. The autocorrelation coefficient is a valuable tool for investigating properties of an empirical time series. Autocorrelations can be used to determine whether there is any pattern like AR, MA, ARMA and ARIMA in the time series. The absence of such patterns shows that the time series is random. The autocorrelation plot shows how values of the time series are correlated with the past values of the series.

The partial autocorrelation measures the degree of association between  $y_t$  and  $y_{t+k}$  when the effect of other time lags 1,2,3 ...  $k - 1$  is somehow removed.

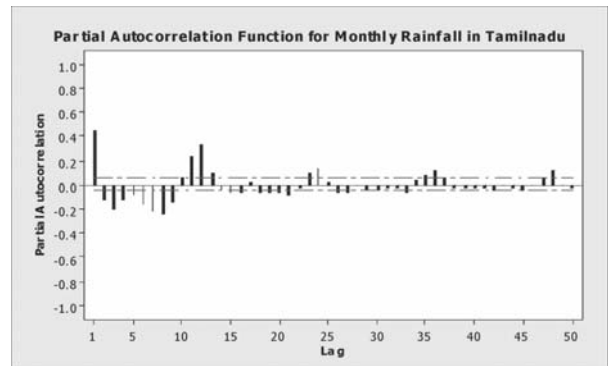


Fig. 4. The Partial Autocorrelation function plot of Monthly Rainfall in Tamilnadu

A stationary time series is one whose statistical properties such as mean, variance, autocorrelation are all constant over time. Most statistical forecasting methods are based on the assumption that the time series is approximately stationary through the use of mathematical transformations. A stationarized series is relatively easy to predict. Another reason for trying to stationarize a time series is to be able to obtain meaningful sample statistics such as means, variances, and correlations with other variables. Such statistics are useful as descriptors of future behavior only if the

series is stationary. The graph of autocorrelation function of monthly rainfall series in Tamilnadu (fig. 3) shows that the series is not stationary.

Seasonality can be defined as a pattern of a time series, which repeats at regular intervals every year. Seasonal fluctuations in a time series data make it difficult to analyse whether changes in data for a given period reflect important increases or decreases in the level of the data, or are due to regularly occurring variation. Seasonal patterns of time series can be examined through correlograms. Seasonality can be found by identifying those autocorrelation coefficients of more than two or three time lags that are significantly different from zero. The autocorrelation function coefficient and the partial autocorrelation coefficient for the first twelve lags are given in table 2. Since the decay of seasonal ACF is very gradual, the series remains seasonally nonstationary and is in need of seasonal differencing.

**Table 2. Autocorrelation and partial autocorrelation functions coefficients for first twelve lags for the monthly rainfall in Tamilnadu**

LAG	ACF	PACF
1	0.376	0.376
2	0.104	-0.044
3	-0.139	-0.190
4	-0.268	-0.177
5	-0.270	-0.114
6	-0.241	-0.133
7	-0.273	-0.248
8	-0.266	-0.265
9	-0.133	-0.155
10	0.120	0.023
11	0.373	0.159
12	0.571	0.334

**The Box – Jenkins SARIMA model:**

A powerful model for describing stationary and non - stationary seasonal time series is Seasonal Autoregressive Integrated Moving Average process (SARIMA) of order  $(p, d, q) \times (P, D, Q)$  s. Model

identification can be made through the autorrelation function and partial autocorrelation function plots (fig. 3 & 4). The tentative SARIMA model for the monthly rainfall series of Tamilnadu is identified as  $(0,1,1) \times (0, 1, 1)_{12}$ .

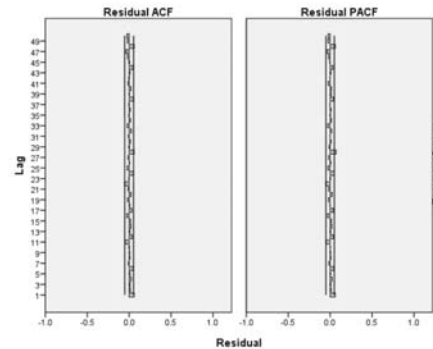


Fig. 5. The residual plots of ACF and PACF of Monthly Rainfall Series

**Performance Evaluation Criteria:**

The error measure used in this research work is the Mean Absolute Percentage Error (MAPE) measure, which is given by

$$MAPE = 100 \times \frac{\sum_{t=1}^N \frac{|E_t|}{Y_t}}{N} \quad \dots(1)$$

where  $Y_t$  and  $E_t$  represent desired inputs and corresponding errors at  $t = 1, 2, \dots, N$  respectively. For this model SARIMA  $(0,1,1) \times (0, 1, 1)_{12}$ , the error measure, Mean Absolute Percentage Error is 17.545. Also, the residual plots (fig. 5) show that the autocorrelation and partial autocorrelation coefficients all lie inside the confidence limits, which shows that the fit is a good fit. The graph (fig. 6) shows the observed and fitted values obtained through SARIMA Model for the monthly rainfall in Tamilnadu.

**V. CONCLUSION**

Present study concludes that the time series forecasting model SARIMA can be of great use in forecasting monthly rainfall over Tamilnadu.

As climate and rainfall prediction involves tremendous amount of imprecision and uncertainty, if more input parameters like. El-Nino Southern Oscillations (ENSO) resulting from the pressure

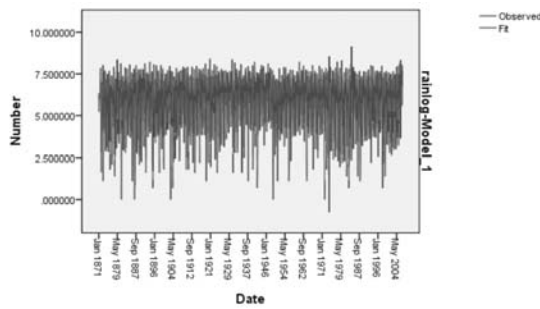


Fig. 6. Graph showing observed and fitted values through the model SARIMA  $(0,1,1) \times (0, 1, 1)_{12}$

oscillations between the tropical Indian Ocean and the tropical Pacific Ocean and their quasi periodic oscillations, land surface temperatures are available, then a prediction of higher accuracy would be possible. As the MAPE value of the series is comparatively less, the forecasting model is reliable.

## REFERENCES

- [1] Abraham A, Philip NS and Joseph KB, (2001), Will we have wet summer? Soft Computing models for long – term rainfall Forecasting, Proceedings of 15th European Simulation Conference ESM2001, Prague.
- [2] Anderson J.A. (2003) An Introduction to Neural Networks, Prentice Hall.
- [3] Balachandran S., Asokan R. and Sridharan S (2006). Global surface temperature in relation to northeast monsoon rainfall over Tamilnadu, Journal of Earth System Science, vol. 115, no.3, pp. 349-362.
- [4] Damodar N. Gujarati and Sangeetha (2007). Basic Econometrics, 4th edition, the Tata Mcgraw – Hill Publishing Company Limited.
- [5] Fallah G-A, Mousavi M. Habiby M, (2008), Using Artificial neural network (ANN) for seasonal rainfall forecasting based on tele-connection patterns, Geophysical Research Abstracts, Vol. 10, EGU 2008 -A- 11664.
- [6] Gadgil, S., Rajeevan, M. & Nanjundiah, R., (2005) Monsoon prediction – Why yet another failure? Current Science 88: 1389-1499.
- [7] George E.P. Box, Gwilym M. Jenkins and Gregory C. Reinsel (1994). Time Series Analysis – Forecasting and Control, 3rd Edition, Pearson Education, Inc.
- [8] Hung N Q, Babel M S, Thailand, Weesakul S, and Tripathi N K, An artificial neural network model for rainfall forecasting in Bangkok, Hydrology and Earth System Sciences discussions, vol. 5, pp.183–218, 2008.
- [9] Khaing Win Mar and Thinn Thu Naing (2008), Optimum Neural Network architecture for Precipitation Prediction of Myanmar, Proceedings of World Academy of Science, Engineering and Technology, vol.36, December.
- [10] Meaza Demissie (2003), Characterizing trends and Identification of rainfall predicting model: An Application to Melksa observatory Nazreth rainfall series, Proceedings of Biometric society, pp. 132 – 134.
- [11] Moller AF(1993), A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Vol.6, pp.525 – 533.
- [12] Nirmala M and Sundaram S.M.(2009), Rainfall Predicting Model: An Application to Tamilnadu Rainfall Series, Proceedings of National Conference on Effect of Climatic Change and Sustainable Resource Management, SRM University and India Meteorological Department, Chennai.
- [13] Dhar O.N., Rakhecha P.R., Kulkarni A.K. (1982). Fluctuations in northeast monsoon rainfall of Tamilnadu, International Journal of Climatology Volume 2, Issue 4, Pages339 – 345.
- [14] Dhar O.N., Rakhecha P.R., Kulkarni A.K. (1982). Trends and fluctuations of seasonal and annual rainfall of Tamilnadu, Proceedings of Indian Academy of Sciences (Earth planetary Sciences), vol. 91, No.2, pp. 97 –104.
- [15] Dhar O.N. and Rakhecha P.R. (1983). Foreshadowing Northeast Monsoon Rainfall over Tamilnadu, Monthly Weather Review, vol.3, issue 1, pp. 109 - 112.
- [16] Philip NS and Joseph KB, On the Predictability of Rainfall in Kerala: An application of ABF Neural Network, Proceedings of Workshop on Intelligent Systems Design and Applications, International Conference on Computational Sciences, ICCS 2001, San Francisco, May 2001.
- [17] Somvanshi V.K., O.P. Pandey, P.K. Agrawal, N.V. Kalanker1, M. Ravi Prakash and Ramesh Chand, (2006), Modelling and prediction of rainfall using artificial neural network and ARIMA techniques, J. Ind. Geophys. Union Vol.10, No.2, pp.141-151.
- [18] Spyros Makridakis, Steven C. Wheelwright, Victor E. McGee. Forecasting – Methods and Applications, 2nd edition
- [19] Sunyoung Lee, Sungzoon Cho and Patrick M. Wong (1998) Rainfall Prediction Using Artificial Neural Networks, Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 233 – 242.
- [20] Yegnanarayana, B (2000) Artificial Neural Networks, Prentice-Hall of India Pvt Ltd, New Delhi, India.