Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

A Comparative Study of Time Series Rainfall Prediction System in Thiruvallur District

[1]Dr. S. Santha, [2]T. Subasini

[1]Assistant professor and Head, Department of Mathematics, Government Arts and Science College, Nagercoil- 629004.

[2]Reg.No: 18121172092018, Research Scholar, Rani Anna Government Arts and Science College for Women, Tirunelveli.

Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, TamilNadu, India.

Abstract: Rainfall prediction plays a vital role in human life,science and technology.Many researchers has been developed various techniques in rainfall prediction.In the present study aims to developed a new sarima model along to the comparison with other statistical approaches such as various type of regression model and simple exponential smoothing model.The rainfall data set of thiruvallur district preceding 11 years(2012-2022) used to predict the proposed method.The comparison has been evaluated by Root Mean Squared Error(RMSE) and Mean Absolute Percentage Error(MAPE).

Keywords*: Rainfall, forecasting, ACF, PACF.*

1.Introduction

Rainfall is a common natural occurrence brought on by the hydrological cycle, which includes evaporation, condensation, and precipitation. It is common knowledge that India has long been interested in learning more about rainfall variability. Every time series model must begin by assuming that some characteristics of the observed pattern will endure into the future. In many fields, including economics, business, engineering, the social sciences, medicine, and politics, time series analysis and its applications have gained popularity. There are countless possible uses for this technique, including descriptive analysis, spectrum analysis, forecasting, intervention analysis, and explanatory analysis. There are numerous models available for time series analysis and forecasting techniques. Moving averages are among the simplest one and widely used.

In general, linear time-series models are used in time-series research. Particularly, a predictive statistical technique for modeling is linear regression analysis. A scientist named Sir Francis Galton conducted the first comprehensive study of this kind of predictive model in the 19th century. A technique for predicting univariate time series data is exponential smoothing. Zhengzhou, Wang and Sheng (2010) developed a new forecast model by applying a generalized regression neural network (GRNN) model.

In this Researchwork,the focus is development of new SARIMA model for predicting the rainfall in Thiruvallur district.

Review of related works

Box and Jenkins proposed the autoregressive integrated moving average model (ARIMA) [2].In 1976,they proposed the model for forecasting in Time series.

M. Sidiq [9] seeks to investigate the predicted rainfall using the time series model. For calculations of the ARIMA (1,0,1), ARIMA (0,1,1), AR(1), and MA(1) for months of data from 2011 to 2014.

Kalekar [7] Arima and Holt-Winter's method are traditional methods of exponential smoothing.All the methods are verified to be acceptable.Hence, these methods are useful for decision makers to lunch schemes for the agriculturist, drainage schemes, and also water resource schemes so on.Later Akinbobola et al. [1]used SARIMA and ARIMA model to forecast rainfall in Nigeria's forest and Savannah eco-climatic areas. The seasonal Autoregressive Integrated Moving Average (ARIMA) model effectively modelled and predicted monthly rainfall in the study's selected stations.

Stochastic Time Series

Definition 1:Time series

A time series is a sequence of observations collected at regular time intervals and there is a correlation among successive observations.

Definition 2: Simple Exponential Smoothing

Simple exponential smoothing is the method of time series forecasting used with univariate data with no trend and no seasonal pattern .It needs a single parameter called (α) also known as the smoothing factor.Alpha controls the rate at which the influence the past observations decreases exponentially.The parameter value between 0 and 1.

$F_t = \alpha A_{t-1} + (1-\alpha)F_{t-1}$ Where,F_t is theforecast rainfall for time t, α is the smoothing constant lies $0 < \alpha < 1$,A_{t-1} is the previous period of actual demand,F_{t-1} is the previous period of forecast demand

Definition 3: Seasonal ARIMA

ARIMA(p,d,q) $(P,D,Q)_m$ models are defined by seven parameters

ARIMA(p,d,q)$(P,D,Q)_m$

$\left(1-\varphi_1 A-\varphi_2 A^2-\cdots-\varphi_p A^p\right)(1-\beta_1 A^s-\beta_2 A^{2s}-\cdots-\beta_P A^{Ps})\,(1-A)^d(1-A^s)^D\,Y_t = \left(1-\psi_1 A-\psi_2 A^2-\cdots-\psi_q A^q\right)\left(1-\theta_1 A^s-\theta_2 A^{2s}-\cdots-\theta_Q A^{Qs}\right)\varepsilon_t$Where

$\left(1-\varphi_1 A-\varphi_2 A^2-\cdots-\varphi_p A^p\right)$ is the autoregressive part of order (p),$(1-\beta_1 A^s-\beta_2 A^{2s}-\cdots-\beta_P A^{Ps})$is the seasonal autogressive part of order (P),$(1-A)^d$ is the differencing order d,$(1-A^s)^D$ is seasonal differencing of order D,$\left(1-\theta_1 A^s-\theta_2 A^{2s}-\cdots-\theta_Q A^{Qs}\right)$ seasonal moving average part of order Q, $\left(1-\psi_1 A-\psi_2 A^2-\cdots-\psi_q A^q\right)$ is moving average part of order q, s =12 is the period of seasonal pattern.

Definition 4: Regression

A regression analysis is a form of predictive modeling technique which investigates the relationship between a dependent and independent (explanatory) variables.

Definition 5: Polynomial regression

The general equation for a polynomial regression is $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$where n=2,3,4,5,6.

In the equation y is the dependent variable,x is the independent variable and $\beta_0 \ldots \beta_n$ are the parameters you can optimize.

Definition 6: 4 Mean Square Error

The formula for MSE is $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ where n= number of data points,y_i is actual value ,$\hat{y}_i$ is the predicted value.

2.Materials and methods

2.1Area of study

Thiruvallur district is the one of the 38 district in the indian state of tamilnadu.Thiruvallur district located in the north eastern part of tamilnadu.

2.2 Data collection

For the current study monthly rainfall data was collected from the statistical department of thiruvallur from the year 1981 to 2022.The data was chosen from the ponneri , pallipattu, gummidipoondi and poonamallee which lies in the eastern, western, northern, and Southern Zones, of thiruvallur district.

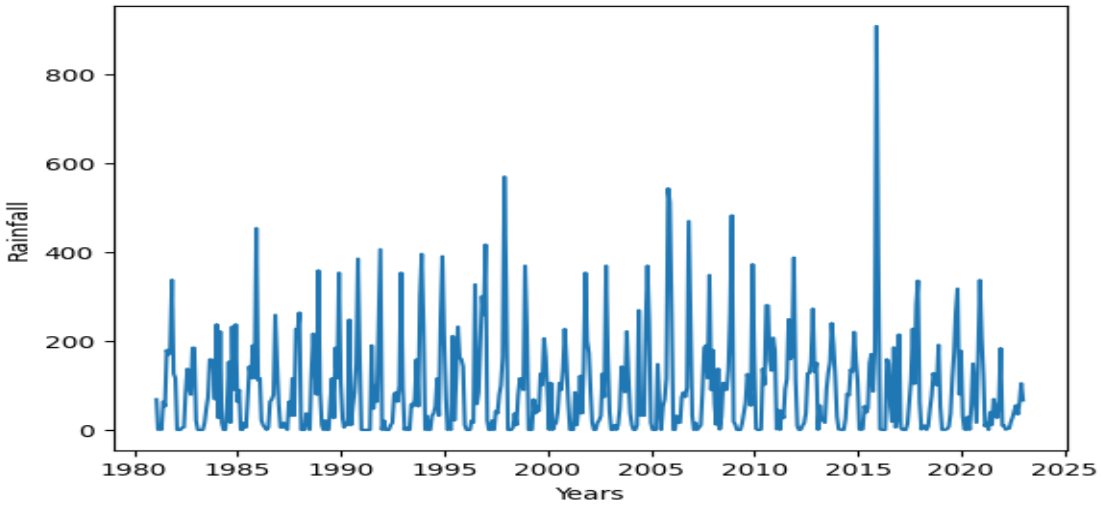

Figure1: monthly rainfall in thiruvallur from 1981 to 2022

2.3 Time series Forecasting Model

2.3.1 SARIMA (seasonal auto-regressive integrated moving average)

BOX & JENKINS [1976] proposed the forecasting rainfall in time series.The following steps are given below:

Step 1: Create the graph.

Step 2: Verify the data's stationary.

Step 3: To find suitable AR and MA models by using ACF and PACF plot for the SARIMA

Step 4: The SARIMA model's best fit is determined .

Step 5: Plot ACF and PACF for the SARIMA model residual in to make sure there is no more data left to be extracted.

Step 6: Predict the series using the SARIMA that fits you the best.

2.3.2 **Simple Exponential smoothing method**

To forecast the rainfall for the following equation

$F_t = 0.8(A_{t-1} - F_{t-1}) + F_{t-1}$

2.3.3 Polynomial Regression model

To forecast the rainfall in equations

Polynomial Regression model: $y = \beta_0 + \beta_1 x + \cdots + \beta_n x^n$ where n=2,3,4,5,6

3. Result and Discussion

3.1 Rainfall forecasting with SARIMA model

In the present study to forecast the rainfall in SARIMA model has been implemented in the python language.To achieve the best result, the basic parameters of this model p,d and q and P,D,Q and m has been determined ,which are dependent on the nature of the time series data.

The optimum order of the best model is (1,0,0)(1,1,0,12) for the parameters of p,d,q and P,D,Q and m respectively,which has the lowest AIC and BIC among the other orders,and to get the best result of the model.The data has been divided into training data from 1981 to 2011 and the test data from 2012 to 2022.The following algorithms has been used to forecast the next 11 years rainfall from 2012 to 2022.

Forecasting rainfall in the different model

Model	Equations
Simple Exponential model	$F_t = 0.8(A_{t-1} - F_{t-1}) + F_{t-1}$
Second degree polynomial	$Y=-0.781x^2+3128x+0.6$
Third degree polynomial	$Y=-0.063x^3-379.6x^2-75831x+08$

Fourth degree polynomial	$Y = -0.003x^4 + 24.70x^3 - 73993x^2 + 1E+08x - 5E+10$
Fifth degree Polynomial	$y = -0.000x^5 + 2.935x^4 - 11737x^3 + 2E+07x^2 - 2E+10x + 9E+12$
Seasonal ARIMA model$(1,0,0)(1,1,0)_{12}$	$(1 - 0.102254A)(1 + (.455564)A^{12})\ (1 - A^{12})Y_t = 9455.907226$

3.2 Comparison of different forecasting method

Forecasting values are compared to simple exponential smoothing, various degree for polynomial regression equation and seasonal ARIMA .The results are shown in the table2

Forecasting monthly rainfall using SARIMA(1,0,0)(1,1,0,12) from 2012 to 2022 depitch the figure 2.

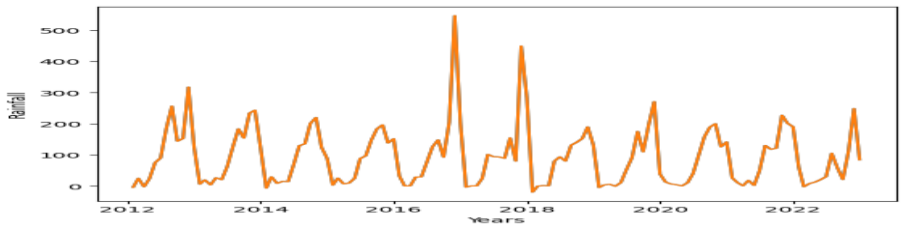

Figure 2: Forecast monthly Rainfall from 2012 to 2022

Model	**MSE**	**RMSE**
SARIMA$(1,0,0)(1,1,0)_{12}$	232.099014123473	15.234796162846207

year	Actual Rainfall	Simple Exponential Smoothing Model	Second degree polynomial	Third degree Polynomial	Fourth degree Polynomial	Fifth degree Polynomial
2012	1032.7	1418.991253	1079.133	1130.3623	1130.61	1130.61
2013	1067.2	1109.958251	1087.299	1135.3925	1135.67	1135.67
2014	938	1075.75165	1012.834	1055	1055	1055
2015	2019.3	1185.95297	1542.836	1577	1578.2011	1578.201134
2016	814.2	978.465	928.857	953.666	953.93176	953.931761
2017	1202.9	1109.877	1110.996	1122.78	1122.9978	1122.997811
2018	733.4	1035.4586	863.2527	859.07	859.20455	859.204553
2019	1174.37	1277.10548	1069.963	1046.68	1046.692	1046.691977
2020	1078.4	1194.917096	1007.421	961.71	961.55979	961.559785
2021	591.97	642	748.8671	677.21	676.85	676.85
2022	427.67	497	650.5964	549.288	548.66	548.66
MSE		469.80	460.7603	414.28682	414.099	413.91292

RMSE		21.67498	21.46533	20.354037	20.3495	20.34485
MAPE		0.133803	0.133803	0.098556	0.0985	0.0985

Table 2: Comparison of different forecasting model

4. Conclusion

In this paper,we calculated and compared the predicted values of the rainfall data of Thiruvallur district using the forecasting models namely SARIMA(1,0,0)(1,1,0,12) various degrees of regression model and simple exponential model.From the table 2 we see that Seasonal ARIMA is the lowest RMSE value other than simple exponential and various degress of polynomial regression models.Hence SARIMA(1,0,0)(1,1,0,12) is the best model.

References

[1] Akinbobola A,Okogbue EC,Ayansola Ak(2018) statistical modeling of monthly rainfall in selected stations in forest and savannah eco-climatic regions of Nigeria.
[2] BOX G,Jenkins G(1970) Time series analysis:forecasting and control.
[3] Box,G.E.P and D.A.Pierce, ”Distribution of the Residual Autocorrelations in Autoregressive- integrated moving Average time series models”,Journal of the American Statistics,Vol.65,(1970)pp.1509-26
[4] Box,G.E.P.,and G.M.Jenkins,”some Recent Advances in forecasting and control”,Applied Statistics,Vol17,(1968) pp.91-109.
[5] Box,G.E.P.,and Jenkins.G.M.,”Time series Analysis with Applications to Economic and environmental problems”.Journal of the American Statistical Association,Vol.70,No.349,(1976)pp.70-79.
[6] Box,G.E.P.,and Jenkins.G.M., and Reinsed,”Time series Analysis”, third edition,prentice Hall 1994.
[7] Kalekar,P.S.Time series forecasting using Holt Winters Exponential Smoothing(2004).
[8] Charles C.Holt., ‘Forecasting seasonal and Trends by Exponentially weighted Moving Averages’,International Journal of Forecasting,Vol.20(2004)pp.5-10
[9] M.Sidiq,”Forecasting rainfall with time series model”,IOP Conf.Series:Materials Science and Engineering,pp.1-7,2018.
[10] James W.Taylor,”Exponentially weighted methods for forecasting intraday time with multiple seasonal cycles”,International Journal of Forecasting.Vol.26,(2010)pp.627-646.