Proceedings of the Sixth International Conference on Inventive Computation Technologies [ICICT 2021]
IEEE Xplore Part Number: CFP21F70-ART; ISBN: 978-1-7281-8501-9

2021 6th International Conference on Inventive Computation Technologies (ICICT) | 978-1-7281-8501-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICT50816.2021.9358615

Time Series Analysis based Tamilnadu Monsoon Rainfall Prediction using Seasonal ARIMA

U.Ashwini, K.Kalaivani*, K.Ulagapriya, A.Saritha
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai
*kalai.se@velsuniv.ac.in

ABSTRACT:

Amount of Rainfall prediction is a major issue for the weather department as it is associated with the human's life and the economy. Excess rainfall is the major cause of natural disasters such as drought and flood which are encountered by the people every year across the world. The time series machine learning model is used for forecasting rainfall at Tamilnadu. Forecasting data required for the analysis is available in the Indian meteorological department. To model the monthly rainfall in Tamilnadu for the period from January 1990 to December 2017, the seasonal ARIMA (Auto Regressive Integrated Moving Average) technique is applied. Using the SARIMA (Seasonal Auto Regressive Integrated Moving Average), the stationarity of the time series flow was demonstrated by the rainfall prediction model and the seasonal correlogram assessed. In relation to the Mean Squared Error (MSE) and Root Mean Squared Error, the output of this model is assessed (RMSE). Therefore, it reveals that the ARIMA model accurately forecasts the Rainfall with less error and the derived model could be used to forecast Monsoon rainfall for the upcoming years.

KEYWORDS: Natural-disaster, Tamilnadu, Time-series analysis, stationarity, correlogram.

I. INTRODUCTION

In the year 2015, Tamil Nadu has suffered intense weather events frequently, south Chennai was flooded due to heavy rains and consequent flooding on the Adyar river. Since then, cyclones Gaja and Vardah, floods in Nilgiris and Cuddalore have ravaged various parts of the state during the Monsoon. Early forecast of heavy rainfall could minimize the disaster such as floods and droughts by evaluating the intermingling of risk and the weakness of a given region [1].

Rainfall prediction helps farmers in prior to manage crops better and also the country's economy. It helps to avoid flood which saves lives of humans. In the last few years, the pattern of rainfall has been changed and unpredictably, this fluctuating rainfall timing and quantity makes prediction difficult and challenging task for meteorological department. The prediction of accurate rainfall will help in the growth of the agricultural sector by the farmers taking decision accordingly for which techniques are available such as ML etc., [2]. Forecasting helps us to visualize future rainfall with use of traditional rainfall data. Time series model is one of many techniques used for forecasting [3] [4].

SEASONAL ARIMA

Seasonality of Time Series is a normal pattern of changes that occurs over periods of S time, where S is the number of time periods before the pattern repeats again. The workflow of SARMIA is shown in fig.1.

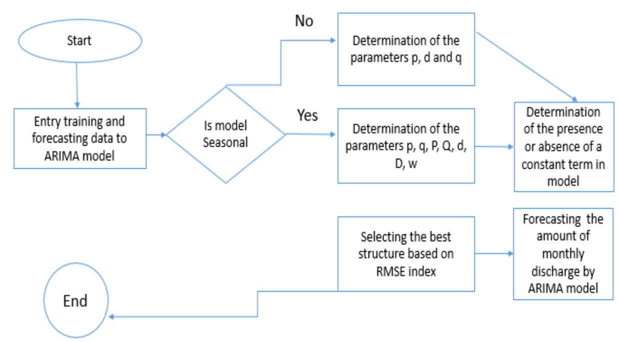

Fig.1 Workflow of SARIMA

Seasonal AR and MA words forecast x_t in a seasonal ARIMA model by means of values and errors at intervals with lags that are multiples of S (span of the seasonality).

A. A seasonal first order autoregressive model will use xt−12 to forecast x_t with monthly data (and S = 12).
B. An autoregressive model of a seasonal second order will use xt−12 and xt−24 to forecast xt. Here we are forecasting the values of this August from the past two Augustans.
C. As a predictor, a seasonal first order MA(1) model (with S= 12) will use wt−12. Wt−12 and wt−24 will be used by a seasonal second order MA(2) model.

(p,q,d) = ARIMA (Non-Seasonal)

(P,Q,D)m = Seasonal , m= total number of observations / years.

The model's seasonal section comprises of elements that are analogous to the non-seasonal elements of the model, but contain reversals of seasonal cycles.

For instance, for quarterly data (m=4 m=4), an ARIMA (1,1,1) (1,1,1) 44 model (without a constant) is written as, (1−ϕ1B) (1−Φ1B4) (1−B)(1−B4) yt =(1+θ1B) (1+Θ1B4) εt.

(1−ϕ1B) (1−Φ1B4) (1−B)(1−B4) yt = (1+θ1B) (1+Θ1B4) εt.

The other seasonal terms are just compounded by the non-seasonal factors.

Auto Correlation Function and Partial Auto Correlation Function

A partial auto correlation can be defined as the sum of the correlation between a variable and a lag of itself which is not clarified by correlations in all lower-order-lags. The auto correlation at lag 1 of the time series Y is the correlation coefficient between Yt and Yt-1, which is also probably the correlation between Yt and Yt-2. But it can also be expected that if Yt is Yt-1-correlated and Yt-2-correlated, there will be a correlation between Yt and Yt-.

In fact, the square of the lag-1 correlation is precisely the amount of correlation we would expect at lag 2. Therefore, the relation to lag 1 disseminates to lag 2 and possibly to lags of greater order. The partial auto correlation at lag 2 is thus the difference among the real correlation at lag 2 and the predicted correlation because of the correlation propagation at lag 1.

II. LITERATURE REVIEW

Many researchers have carried out the forecasting and analysis of the time series based on different approaches.

Graham and Mishra [6] have reported seasonal Auto Regression Integrated Moving Average time series approach for forecasting rainfall at Allahabad. The performance of the model is analyzed and evaluated by correlation coefficients. For monthly water temperature series and stream flow, Moeeni et al [7] used seasonal differencing in ARIMA and model forecasting. The periodic effect elimination proves that this model performance is better and has been the best stationarizing method over spectral analysis and seasonal standardization. Valipour et al [8] proved the suitability of SEASONAL ARIMA model for long term forecasting in the US over the ARIMA models. SEASONAL ARIMA was the appropriate model for simulating water quality under extreme discharges and water temperature conditions in the Danube River (Europe) hybrid forecasting model that incorporates different non-linear scheme-switching time series to forecast periodic inflows into the Liptovská Mara reservoir in Slovakia.

Sinha and Gautham [9] used the SEASONAL ARIMA model to predict monthly (mean) reference crop evapo-transpiration of Bokaro. A prediction strategy for the climatic instability of the Batticaloa district of Sri Lanka has been implemented by Partheepan et al [10]. To research the climate variability and pattern using mathematical models, they have analyzed rainfall time series. Parmar et al have compared and reported the review of various models and approaches for rainfall forecasting. This study revealed that ANN is the most common solution because it deals with non-linear data relationships and has the ability to learn from past data from the Artificial Neural Network. Box-Jenkins method was used to build the ARIMA model for monthly rainfall data and time series analysis could be used to anticipate the peak values of the effects of rainfall. Seasonal Arima model is widely used by the researchers for accurate prediction in various applications [5].

III. MATERIAL AND METHODS

A. Methodology

The framework of this time-series modelling for forecasting Monsoon Rainfall in Tamilnadu consists of following modules.
1. Visualizing time-series
2. Plot ACF and Partial ACF Charts.
3. Building Seasonal Arima model
4. Predicting future values

B. Study Area

The primary aim of this proposed system is to design a model for machine learning to predict future predictions of rainfall during the Tamilnadu Monsoon Periods. We took the information from the Indian meteorological department; the dataset consists of periodic rainfall updates with a total of 117 readings from January 1901 to December 2017. This dataset's statistical summary is given in Table 1.

Table 1. Statistical summary of dataset

	ANNUAL	JF	MAM	JJAS	OND
count	117.000000	117.000000	117.000000	117.000000	117.000000
mean	947.021368	36.958974	134.123077	330.905983	438.477778
std	168.151177	37.007509	45.601557	70.045558	131.885644
min	318.000000	0.100000	45.700000	94.200000	149.300000
25%	866.900000	9.600000	97.300000	283.000000	345.700000
50%	950.500000	27.100000	130.500000	326.100000	432.200000
75%	1045.900000	53.200000	160.500000	380.000000	523.200000
max	1365.300000	165.900000	287.300000	481.200000	782.300000

C. Time Series Exploration:

It is required to analyze the trends before we build any time series model. The rainfall data set is loaded and the data frame is converted into time series object as required for the process. The cycle of data set is considered as a yearly data set from the year 1990 to 2016 and it is set as the 1st month of every year. Therefore, these cycles are aggregated and a year-on-year trend of an Annual rainfall in Tamilnadu is shown in the Fig.3. The trend analysis of rainfall and result of Augmented Dickey-Fuller Test (ADF) is shown in the fig.4 and fig. 5 respectively, stating that the series is stationary enough to perform any kind of time-series modelling. The correlation plot such as Auto Correlation Function (ACF) and Partial AutoCorrelation Function (PACF) are plotted are shown in fig.6(a) and 6(b).

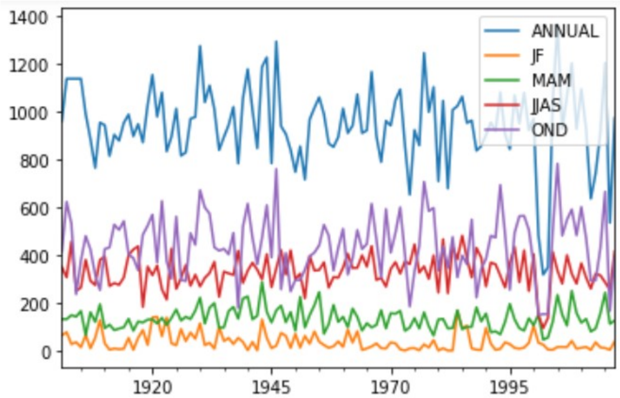

Fig.2. Monsoon Rainfall Series of Tamilnadu

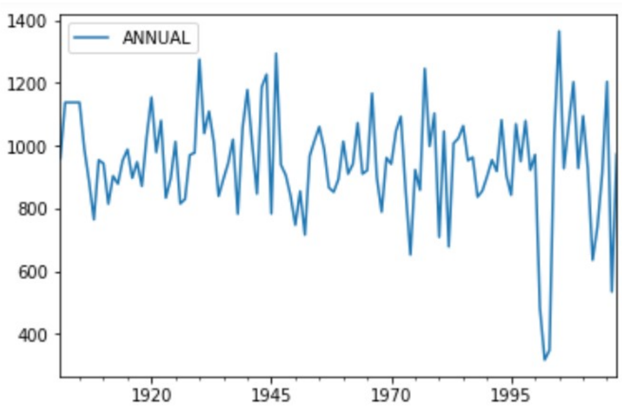

Fig.3 Annual Rainfall Series of Tamilnadu

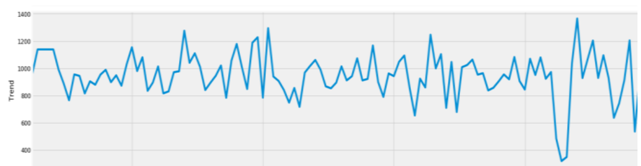

Fig.4 Trend Analysis of Rainfall

```
ADF Test Statistic : -8.447148826983852
p-value : 1.700664179531161e-13
#Lags Used : 0
Number of Observations Used : 116
strong evidence against the null hypothesis(Ho), reject the null hypothesis. Data has no unit root and is stationar
```

Fig.5 Augmented Dickey Fuller Test Result

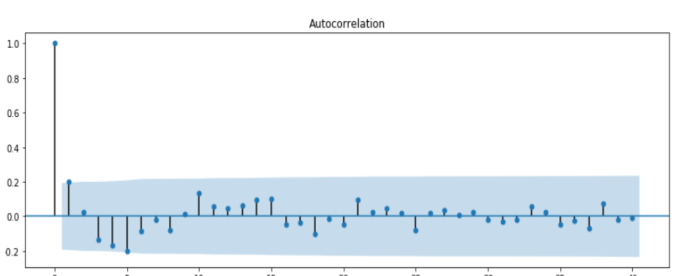

Fig. 6(a) Auto Correlation Function plot of rainfall data

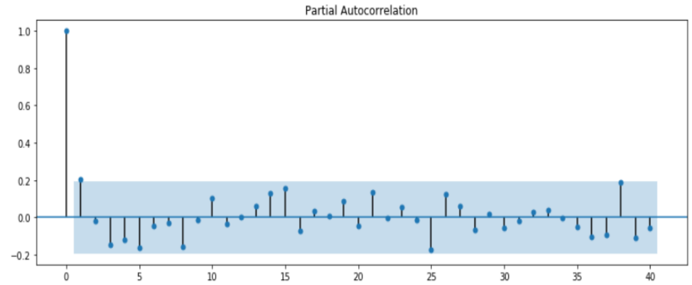

Fig 6(b) Partial Auto Correlation Function plot of rainfall data

IV. RESULTS AND DISCUSSION

Rainfall forecasting in Tamilnadu is carried out by constructing the Arima model for the time series. Developed the Arima model in python for forecasting using different packages such as pandas, seaborn etc., [11] [12]. Prediction and modeling on the basis of the Seasonal Auto Regressive Moving Average statistical process. By 4incorporating seasonal components, the Arima model is acceptable and is designed to forecast rainfall for the next twenty-five years.
The prediction is visualized by plotting as shown in the Fig.7 where the orange line indicates the predicted values and the model summary is shown in the Fig.8. Likewise, the Prediction for Monsoon periods such as Jan to Feb (JF), March to May (MAM), June to Sept (JJAS) and Oct to Dec (OND) is plotted as shown in the following fig.9-12.

Statespace Model Results

Dep. Variable:	ANNUAL	No. Observations:	117
Model:	SARIMAX(0, 1, 0)x(1, 1, 0, 12)	Log Likelihood	-724.337
Date:	Sun, 28 Jun 2020	AIC	1452.674
Time:	19:40:24	BIC	1457.963
Sample:	01-01-1901	HQIC	1454.817
	- 01-01-2017		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.S.L12	-0.5493	0.067	-8.257	0.000	-0.680	-0.419
sigma2	6.34e+04	5982.501	10.598	0.000	5.17e+04	7.51e+04

Ljung-Box (Q):	58.58	Jarque-Bera (JB):	27.37
Prob(Q):	0.03	Prob(JB):	0.00
Heteroskedasticity (H):	2.17	Skew:	-0.15
Prob(H) (two-sided):	0.02	Kurtosis:	5.49

Fig.7 Sarima Model Summary

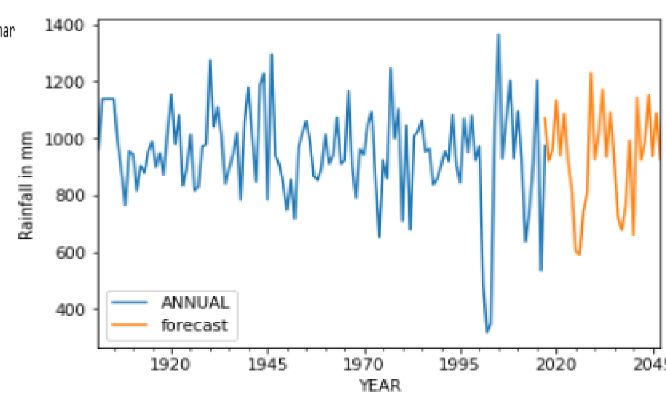

Fig.8 Annual Rainfall prediction in Tamilnadu

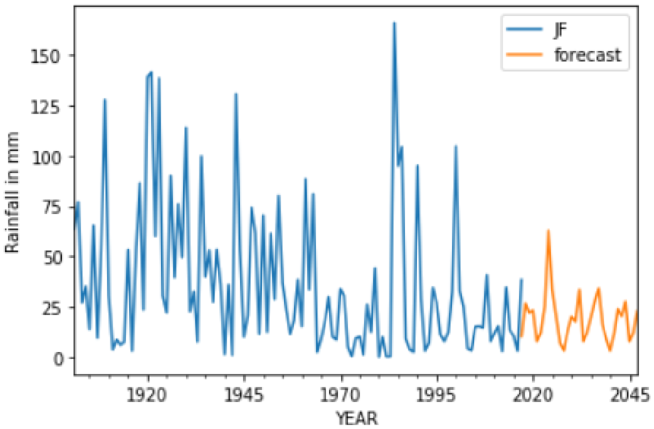

Fig.9. Jan - Feb prediction

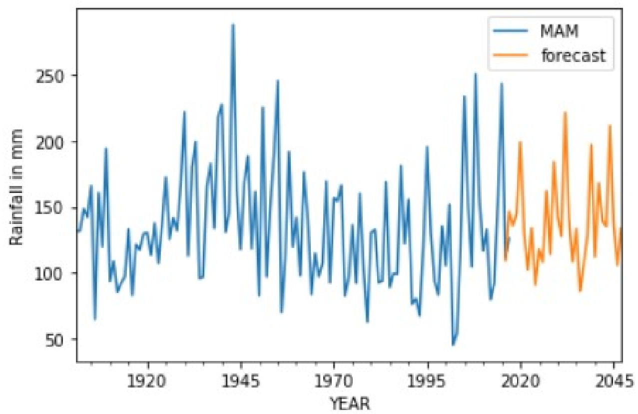

Fig.10. Mar - May prediction

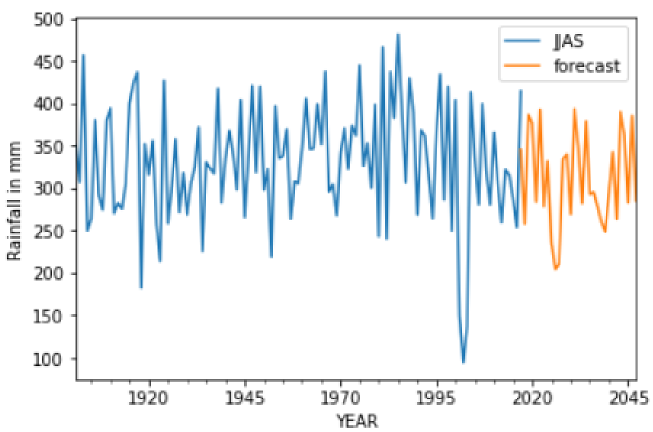

Fig.11. Jun - Sept Prediction

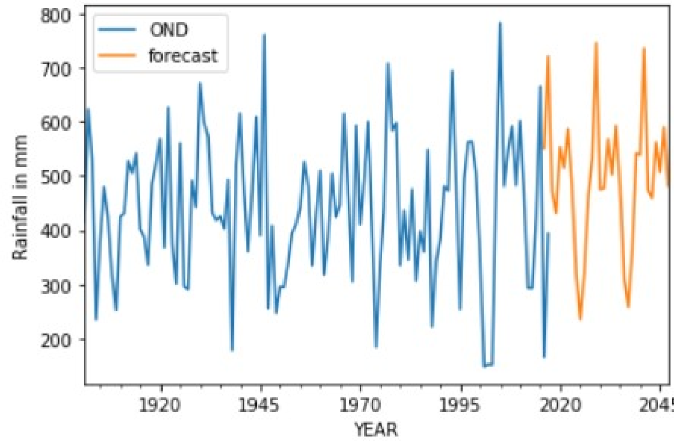

Fig. 12. Oct-Dec Prediction

V. CONCLUSION

The time series analysis is recorded to solve problems such as drought, flood and other agricultural problems and used a machine learning model to forecast the future values of Tamilnadu Monsoon rainfall. For this prediction, the rainfall dataset is retrieved from Data.gov, a free to use open web-based data platform. The data consists of periodical and annual rainfall updates from January 1901 to December 2016. The time-series data is visualized by plotting correlation plots and time-series plot. The time-series forecasting of rainfall in Tamilnadu is carried out by statistical methods such as the Seasonal Arima model.

The performance of this model is evaluated based on the metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The results clearly show that the ARIMA model accurately forecasts the Rainfall with less error and the derived model could be used to forecast Monsoon rainfall for the upcoming years.

VI. REFERENCES

[1]. Venkatesh, R., C. Balasubramanian, and M. Kaliappan. "Rainfall prediction using generative adversarial networks with convolution neural network." *Soft Computing*: 1-14.

[2]. Naidu, D., Majhi, B., & Chandniha, S. K. (2021). Development of Rainfall Prediction Models Using Machine Learning Approaches for Different Agro-Climatic Zones. In *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science* (pp. 72-94). IGI Global.

[3]. Coban, Veysel, Ezgi Guler, Taner Kilic, and Suheyla Yerel Kandemir. "Precipitation forecasting in Marmara region of Turkey." *Arabian Journal of Geosciences* 14, no. 2 (2021): 1-10.

[4]. Kumar, Vikram, Manvendra Singh Chauhan, and Shanu Khan. "Application of Machine Learning Techniques for Clustering of Rainfall Time Series Over Ganges River Basin." In *The Ganga*

River Basin: A Hydrometeorological Approach, pp. 211-218. Springer, Cham, 2021.

[5]. Wu, Xianghua, Jieqin Zhou, Huaying Yu, Duanyang Liu, Kang Xie, Yiqi Chen, Jingbiao Hu, Haiyan Sun, and Fengjuan Xing. "The Development of a Hybrid Wavelet-ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis." *Atmosphere* 12, no. 1 (2021): 74.

[6]. Feng, Zhong-kai, and Wen-jing Niu. "Hybrid artificial neural network and cooperation search algorithm for nonlinear river flow time series forecasting in humid and semi-humid regions." *Knowledge-Based Systems* 211 (2021): 106580.

[7]. Abdullah, A. S., B. N. Ruchjana, and I. G. N. M. Jaya. "Comparison of SARIMA and SVM model for rainfall forecasting in Bogor city, Indonesia." In *Journal of Physics: Conference Series*, vol. 1722, no. 1, p. 012061. IOP Publishing, 2021.

[8]. Liu, Tianyang, Zunkai Huang, Li Tian, Yongxin Zhu, Hui Wang, and Songlin Feng. "Enhancing Wind Turbine Power Forecast via Convolutional Neural Network." *Electronics* 10, no. 3 (2021): 261.

[9]. Alhumoud, Jasem M., and Nourah Almeshaan. "Time Series Analyses and Modeling of Environmental Systems: Case Studies in Kuwait."

[10]. Harper, Corey D., Sean Qian, and Constantine Samaras. "Improving Short-Term Travel Speed Prediction with High-Resolution Spatial and Temporal Rainfall Data." *Journal of Transportation Engineering, Part A: Systems* 147, no. 3 (2021): 04021004.

[11]. Shuofeng, Liu, Lei Puwen, and Koji Koyamada. "LSTM Based Hybrid Method for Basin Water Level Prediction by Using Precipitation Data." *Journal of Advanced Simulation in Science and Engineering* 8, no. 1 (2021): 40-52.