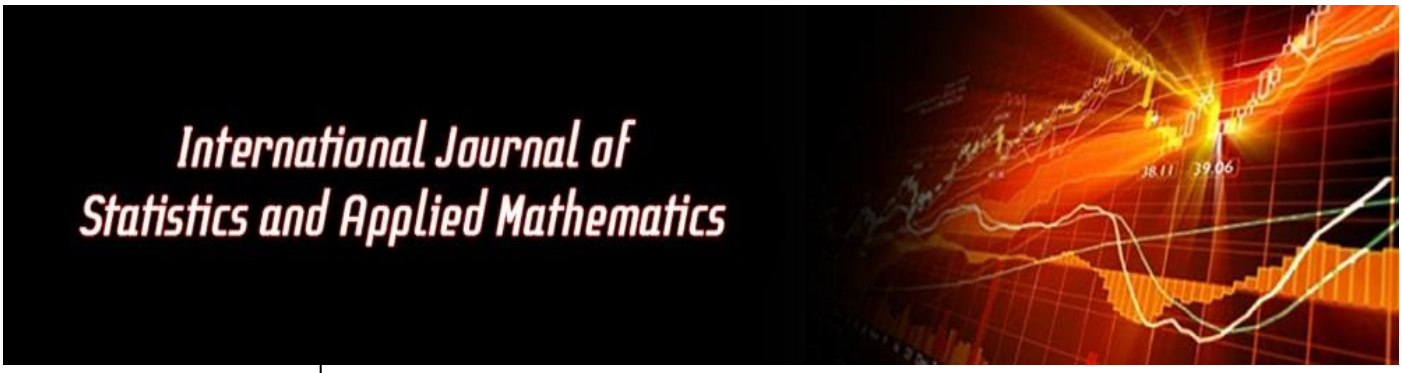
International Journal of Statistics and Applied Mathematics 2021; 6(3): 83-86



ISSN: 2456-1452
Maths 2021; 6(3): 83-86

www.mathsjournal.com
Received: 04-03-2021
Accepted: 06-04-2021

Dr. Kesavulu Poola
Assistant Professor,
Emeralds Advanced Institute of Management Studies, Tirupati, Andhra Pradesh, India

P Hema Sekhar
Research Scholar, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

Corresponding Author:
Dr. Kesavulu Poola
Assistant Professor,
Emeralds Advanced Institute of Management Studies, Tirupati, Andhra Pradesh, India

Prediction of rainfall by using extreme gradient boost (XG boost) in Vishakapattanam area, Andhra Pradesh

Dr. Kesavulu Poola and P Hema Sekhar

Abstract
Now-a-days the prophecy of rainfall in a particular time periods is scientifically challenging and tough jobs because, we found anomalous changes in climatic conditions. In the contemporary fast moving world, forecasting of Rainfall, it is very helpful for planning and satisfying Agricultural needs. There are many research institutes are rigorously trying to find the rainfall predictions by using different techniques. This paper describes, how XG Boost Analysis predicts the rainfall in Monthly Scales. In the paper we recommend that XG-Boost Model is best fit to forecast the rainfall up to 3 to 5 years with 95% accuracy by the evidence of Analyzing last 30 years of data (1987-2017) in Vishakapattanam region. The present article indicates that the XG-Boost Model provides Consistent and Satisfactory Prophecy for Rainfall on Monthly Scales in Vishakapattanam.

Keywords: prediction, rainfall, XG-boost algorithm

Introduction
The agricultural practices and crop yields of India are heavily dependent on the climatic factors like rainfall and water resources. Out of 142 million ha cultivated land in India, 92 million ha (i.e. about 65%) are under the influence of rain fed agriculture. Unlike irrigated agriculture, rain fed farming is usually diverse and risk prone. The monsoon season is the principal rain bearing season and in fact a substantial part of the annual rainfall over a large part of the country occurs in this season. Small variations in the timing and the quantity of monsoon rainfall have the potential to impact on agricultural output. Rainfall is the most important climatic element that influences agriculture. Monthly rainfall forecasting plays an important role in the planning and management of agricultural scheme and water resources systems. The main objective of the present study is to develop a valid stochastic model to simulate monthly rainfall in Vishakapattanam District. Rainfall is a seasonal phenomenon with twelve months period, but most probably depends on monsoon. Seasonal time series are often modeled by different techniques. In recent times, many researchers modeled monthly rainfall using SARIMA methods and Box-Jenkins ARIMA Methods. In present study, rainfall modeling and forecasting is tried for many traditional Algorithms like ARMA, ARIMA, SARIMA etc, but no model gives the good fit for this data. After the rigorous search researcher preferred machine earning technique called extreme gradient boosting algorithms for forecasting.

Study Area
Visakhapatnam District is one of the North Eastern Coastal districts of Andhra Pradesh and it lies between 17o - 15' and 18o-32' Northern latitude and 18o - 54' and 83o - 30' in Eastern longitude. It is bounded on the North partly by the Orissa State and partly by Vizianagaram District, on the South by East Godavari District, on the West by Orissa State and on the East by Bay of Bengal. The District receives annual normal rainfall of 1202 MM, of which south-west monsoon accounts for 72.0% of the normal while North-East monsoon contributes 13.9% of the normal rainfall during 2006-2007. The rest is shared by summer showers and winter rains. Agency and inland Mandals receive larger rainfall from the South West Monsoon, while Coastal Mandals get similarly larger rainfall from North-East monsoon.

The total geographical area of the district is 11.16 lakh hectares of this 36.45% alone is arable area while 39.53% is forest area. The rest is distributed among "Barren and uncultivable land" about 11.7% and "Land put to non agricultural uses" about 9.0%. Out of the arable area, the net area sowed form 27.2% while cultivable waste and fallow (current and old) lands constitute about 9.2% during 2006-2007. Agriculture is the main stray of nearly 70% of the households. Though Visakhapantam city is industrially developing, the rural areas continued to be backward. Rice is a staple food of the people and Paddy is therefore the principal food crop of the district followed by Ragi, Bajra and Jowar and Cash Crops such as Sugarcane, Groundnut, Sesamum Niger and Chillies are important. Since there is no Major Irrigation system, only about 36% of the cropped area is irrigated under the Ayacut of the Medium Irrigation System and Mimnor Irrigation Tanks. The rest of the cultivated area is covered under dry crops depending upon the vagaries of the monsoon. The productivity of the crops is low.

Review of Literature

Kalogirou *et al*. (1997) applied ANN to renovate the rainfall time series data over Cyprus. Wong *et al*. (1999) fabricated fuzzy rule bases with the help of Self Organizing Map (SOM) and back propagation neural networks and then with the help of the rule base developed predictive model for rainfall over Switzerland. Op pandey *et al*. (2006) [23] Reveals that Traditional ARIMA is most preferable model to predict the rainfall, but when compare with ANN, ANN is one most useful and appropriate technique for better prediction of rainfall behaviour. Fadhilah Yusof *et al*. (2013) [15] Initiated to implement hybrid ARIMA-GARCH for the serial dependency and volatility in the rainfall data. Also suggested that Seasonal ARIMA model predicts well for dynamic behaviour of the rainfall. Rahman *et al*. (2013) [21] proved that the conventional statistical models ARIMA and SARIMA are the appropriate techniques with more efficacy than other modelling techniques like Adaptive neuro-fuzzy inference system ANFIS. Pinky Saikia Dutta (2014) established the ARIMA model by considering the parameters like: mean sea level, temperature and wind speed to predict the rainfall behavioural pattern. H. Yasin *et al*., (2016). Empirical approach determines the number of hidden nodes in which ANN is reinstructed with fluctuating number of hidden layers, and the output error is depending on the function of the number of hidden layers. Swapnil S. Potdar *et al*. (2019) [29] used regression analysis, MK-test and Sen's analysis to analyse the long-term time series data for prediction of rainfall. This explains that the momentum of spatial variability shows much impact on the behaviour of the rainfall pattern. Lata, K *et al*. (2020) [18] constructed SARIMA model for prediction of rainfall behaviour. This model explains, stochastic modeming is also best one, to predict the rainfall. Shuni Qian *et al*. (2020) build the conventional statistical models based on GCMs and SST dipoles with bias and without bias correction to prediction of monsoon rainfall in the Yangtze River basin. These models are the most appropriate dipoles for the forecaster and it creates a functional relationship between the SST dipoles and monsoon rainfall.

Methodology

Data Collection

Monthly rainfall data for the past 30 years from 1987 to 2017 was collected from India Meteorological Department (IMD), and other Government departments of Vishakapattanam, Andhra Pradesh.

Software Used: R program

XGBoost is an algorithm were selected using R LANGUAGE to find the best fit of a time series to past values of this time series in order to make forecasts. XGBoost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

Data Preprocessing Techniques

Time series is defined as a set of observations arranged chronologically i.e. a sequence of observations usually ordered in time. The principal objective of a time series analysis is to describe the history of movements in time of some variable at a particular period. The objective is to generate data having properties of the observed historical record. To compute properties of a historical record, the historical record or time series is broken into separate components and analyzed individually to understand the casual mechanism of different components. Once properties of these components are understood, these can be generated with similar properties and combined together to give a generated future time series. Analysis of a continuously recorded rainfall data time series is performed by transforming the continuous series into a discrete time series of finite time interval.

Mathematical modeling of rainfall data is a stochastic process. Several mathematical models based on the probability concept are available. These models help in knowing the probable weekly, monthly or annually rainfall. Over the past decade or so, a number of models have been developed to generate rainfall and runoff. Monthly rainfall and temperatures were analyzed using time series analysis. Time series models have been extensively studied by Box and Jenkins (1976) and as their names have frequently been used with synonymously with general ARIMA process applied to time series analysis and forecasting. In this study, the data of rainfall most influenced by the monsoon season every year. Because Rainfall is a seasonal phenomenon with twelve months period, but most probably depends on monsoon. In order to getting good fit XG Boost algorithm is used for this data.

XGBoost (eXtreme Gradient Boosting) is one of the most loved machine learning algorithms. It can be used for supervised learning tasks such as Regression, Classification, and Ranking. It is built on the principles of gradient boosting framework and designed to "push the extreme of the computation limits of machines to provide a scalable, portable and accurate library." Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Extream Gradient Boosting (XGBoost) is a machine learning technique. XGBoost utilizes boosting. It uses, different trees are made consecutively such that every last one of the following trees attempts to lessen the mistakes from the past tree. XGBoost was first delivered in 2014 and had been carried out in Python, R program and so forth XGBoost is exceptionally mainstream in machine learning. Right now,

XGBoost has been utilized for different purposes, for example, forecast of raw petroleum costs, conclusion of ongoing kidney infection, forecast of mishaps, expectation of workers changing positions, forecast of material particulates in the climate, and interruption discovery. Notwithstanding, as of recently there has been no research utilizing XGBoost for precipitation expectation. So this examination is the primary exploration to utilize XGBoost for precipitation expectation. XGBoost itself can deal with both regression trees and classifications. In spite of the fact that XGBoost has a decent presentation, it has a disadvantage that is the chance of overfitting. This can be taken care of by trying different things with displaying boundaries.

Results and Discussion

The model that seems to represent the behaviour of the series is searched, by the means of autocorrelation function (ACF) and partial auto correlation function (PACF), for further investigation and parameter estimation. The behaviour of ACF and PACF is to see whether the series is stationary or not. For modelling by ACF and PACF methods, examination of values relative to auto regression and moving average were made. An appropriate model for estimation of monthly rainfall for Vishakapattanam District was finally found. Many models for ishakapattanam District, according to the ACF and PACF of the data, were examined to determine the best model.

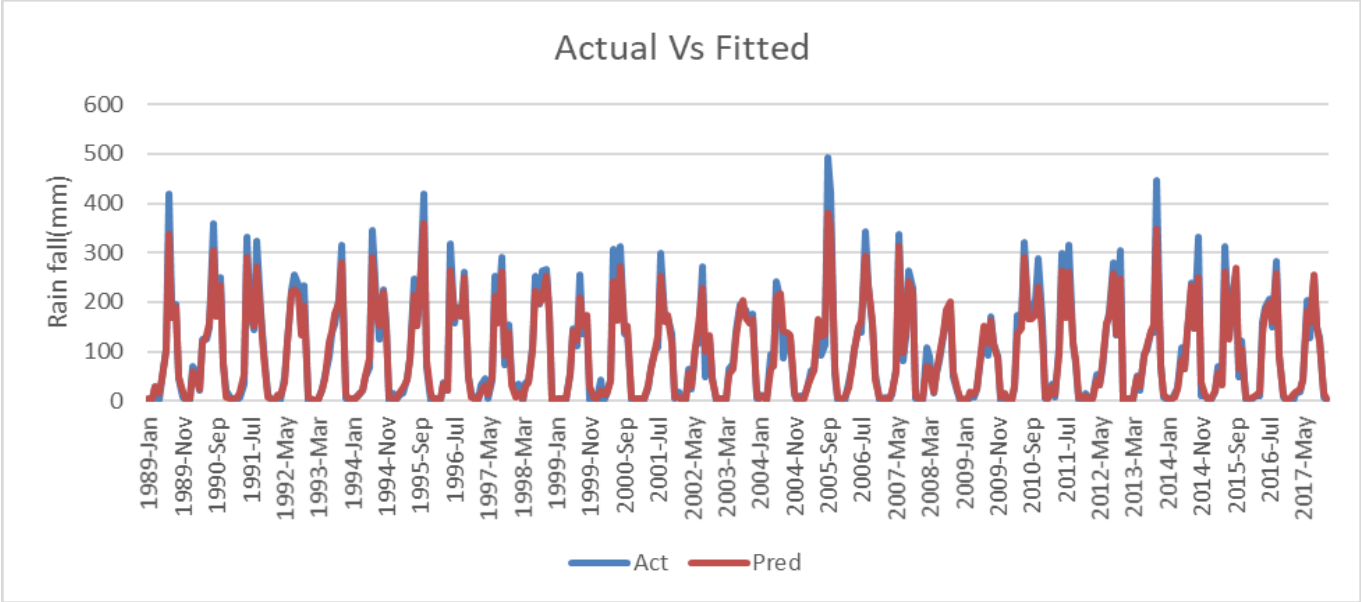


Fig 1: Actual and predicted rainfall in Vishakapattanam district (Jan. 1987-Dec. 2017)

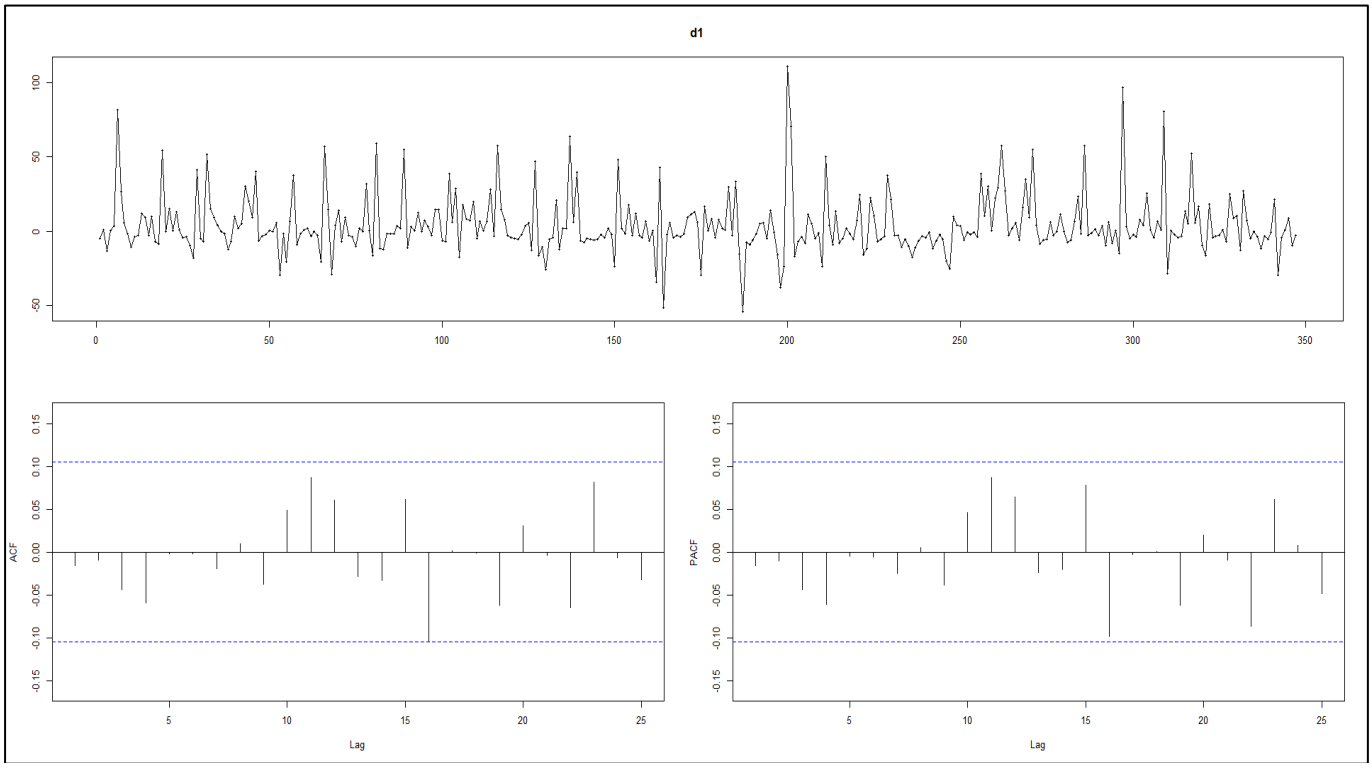


Fig 2: Autocorrelation and Partial Autocorrelation function of rainfall.

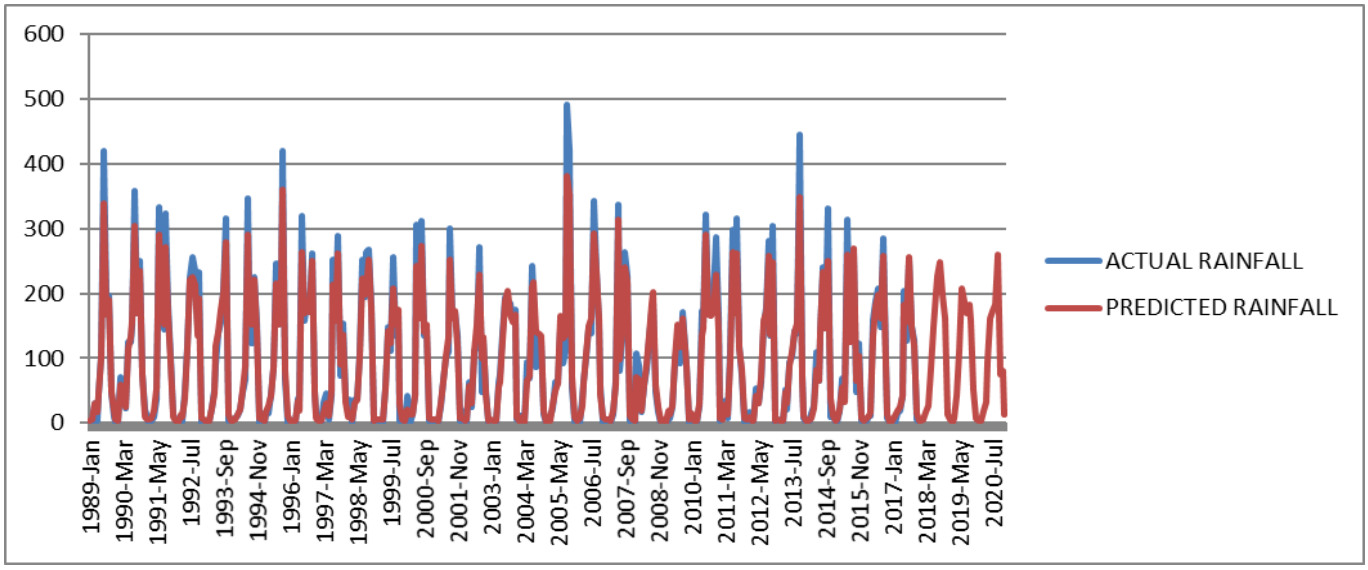


Fig 3: Actual and predicted rainfall in Vishakapattanam district (Jan. 1989-Dec. 2020)

Conclusion

The XG-Boost algorithm methodology was used to develop monthly rainfall of Vishakapattanam. The monthly rainfall is panning over the period of 1987-2017 at vishakapattanam. The performance of resulting XG-Boost algorithm was evaluated by using the data from the year 1987-2017 through graphical comparison between the forecasted and observed data. In XG-Boost algorithm the forecasted and observed data of rainfall showed good results. The study reveals that XG-Boost algorithm can be used as an appropriate tool to forecast rainfall in Vishakapattanam for upcoming years. The predictions made for rainfall by XG-Boost algorithm are appropriate based on the data Accuracy.

References

1. Anosh Graham, Ekta Pathak Mishra Time series analysis model to forecast rainfall for Allahabad region, Journal of Pharmacognosy and Phytochemistry 2017;6(5):1418-1421
2. Box GEP, Jenkins GM. Time series analysis: forecasting and control, Prentince Hall, Inc, 1976, 575.
3. Caraka RE *et al*., "Ecological Show Cave and Wild Cave: Negative Binomial Gllvm's Arthropod Community Modelling," Procedia Comput. Sci 2018;135:377-384.
4. Dr. Kesavulu Poola, Prof. Bhupathi Naidu M. Importance of studentized and press residuals for nonlinear multivariate regression models in Journal of Emerging Technologies and Innovative Research 2008;5(7):353-356
5. Etuk EH, Mohamed TM. Time Series Analysis of Monthly Rainfall data for the Gadaref rainfall station, Sudan, by SARIMA Methods. International Journal of Scientific Research in Knowledge 2014;2(7):320-327.
6. International Journal of Agriculture Sciences 2008;5(8):112-25.
7. Li YP, Nie S, Huang CZ, McBean EA, Fan YR. Huang GH. An integrated risk analysis method for planning water resource systems to support sustainable development of an arid region. J. Environ. Inform.
8. Mahmud I, Bari SH, Ur Rahman MT. Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method. Environ. Eng. Res 2017;22:162-168.
9. Mohamed TM, Ibrahim A. Time Series Analysis of Nyala Rainfall Using ARIMA Method Journal of Engineering and Computer Science (JECS) 2016; 17(1):5-11
10. Nirmala M, Sundaram SM. A Seasonal Arima Model for Forecasting monthly rainfall in Tamil Nadu. National. Journal on Advances in Building Sciences and Mechanics 2010;1(2):43-47.
11. Nury AH, Hasan K, Alam MJB. Comparative study of wavelet-ARIMA and wavelet ANN models for temperature time series data in northeastern Bangladesh. J King Saud Univ. Sci 2017;29:47-61.
12. Hema Sekhar P, Dr. Kesavulu Poola, Dr. M Bhupathi Naidu. Combined multiple forecasting model using regression, International Journal of Statistics and Applied Mathematics 2020;5(6):147-150
13. Hema Sekhar P, Dr. Kesavulu Poola, Raja Sekhar K, Dr. Bhupathi Naidu M. Modelling and prediction of coastal Andhra rainfall using ARIMA and ANN models in International Journal of Statistics and Applied Mathematics 2020;5(6):104-110.
14. Kesavulu P, Vasu K, Bhupathi Naidu M, Abbaiah R, Balasiddamuni P. The effect of multicollinearity in nonlinear regression models in International Journal of Applied Research 2016;2(12):506-509
15. Papalaskaris T, Panagiotidis T, Pantrakis A. Stochastic monthly rainfall time series analysis, modeling and forecasting in Kavala City, Greece, North-Eastern Mediterranean Basin. Procedia Eng 2016;162:254-263.
16. Caraka RE, Bakar SA. "Evaluation Performance of Hybrid Localized Multi Kernel SVR (LMKSVR) In Electrical Load Data Using 4 Different Optimizations," J Eng. Appl. Sci 2018;13:17.
17. Rezzy Eko Caraka, Sakhinah Abu Bakar, Muhammad Tahmid. Rainfall forecasting multi kernel support vector regression seasonal autoregressive integrated moving average (MKSVR-SARIMA) AIP Conference Proceedings 2111, 020014, 2019.
18. Singh VP, Chowdhury PK, Comparing some methods of estimating mean area rainfall. Water Resources Bulletin 1986;22(2):275-282.
19. Swapnil Potdar S, Kulkarni S, Patil P, Pawar RP, Jakhalekar VV, Nade DP. The long-term trend analysis of rainfall data from 1901 to 2015 for Maharashtra and Goa region from India International Journal of Water 13(3), 293-309.