International Journal of Statistics and Applied Mathematics 2020; 5(6): 104-110

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2020; 5(6): 104-110

www.mathsjournal.com
Received: 07-09-2020
Accepted: 12-10-2020

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

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

K Raja Sekhar
Research Scholar, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh, India

Dr. M Bhupathi Naidu
Professor, DDE, S.V. University, Tirupati, Andhra Pradesh, India

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

Modelling and prediction of coastal Andhra rainfall using ARIMA and ANN models

P Hema Sekhar, Dr. Kesavulu Poola, K Raja Sekhar and Dr. M Bhupathi Naidu

Abstract
Precipitation and climate forecasting in the Coastal Andhra are extremely needed for the agriculture and water resource management system. Most challenging factor in this study is nature of rainfall data. Due to drastic climatic conditions, the rainfall becomes nonlinear and dynamic in nature. Conventional statistical models are not enough to predict the actual rainfall trend, which require contemporary computer modelling and simulation techniques for accurate prediction. In this paper, we are using both traditional statistical technique autoregressive integrated moving average (ARIMA) and contemporary AI model Artificial Neural Network (ANN) for prediction of rainfall. In order to evaluate the forecasting efficacy, we used of 117 years of mean annual rainfall data from year 1901 to 2017 of Coastal Andhra (India). The models were trained with 100 years of annual rainfall data. The ARIMA and the ANN methods are used to the data to draw the accuracy. The accuracy of the model was assessed by using remaining 17 years of data. The study explains that ANN model can be used as a significant prediction tool to forecast the rainfall when compare with ARIMA model.

Keywords: Rainfall, forecasting, ARIMA, ANN

1. Introduction
Prediction of the Precipitation is the most significant and challenging subject in the present era. Agriculture is one of the major businesses in country like India, the relevance of rainfall forecast is necessary for to meet the future course of demands. Across the globe, several contributions have been developed to predict rainfall behaviour using numerous traditional and modern methods. In the present study, we have used two models to predict the rainfall pattern like: traditional statistical method autoregressive integrated moving average ARIMA and the contemporary artificial intelligence model artificial neural network ANN. Data appropriateness is the core content for any model fitting. Conventional statistical methods are performing well when the data is balanced as well as linear. But it is not possible in all cases. The actual rainfall trend for previous decades is abnormal and data is nonlinear in nature. In this case, the solicitation of ANN for prediction of time series model is fairly good (Mirko & Christian 2000). ANN is powerful technique to deal estimated nonlinear functions. This method designed based on human neurological system, It contains of a sequence of fundamental computing elements, called as neurons interrelated together to form a network, [Rummelhart & McClelland 1996]. The parallel-distributed processing architecture of ANN has end up being an extremely ground-breaking computational techniques which is presently being utilized in a few fields to show the dynamic cycles effectively [Mirko and Christian 2000; Mary 2002] including the precipitation [Singh and Chowdhury 1986; Cigizoglu 2002] [28, 10]. This method can take in and sum up from guides to create important arrangements. Among every climate occurring, precipitation has the most basic influence in human life. Human development generally relies on its recurrence and sum to different scales. A few stochastic models have been endeavoured to gauge the models of precipitation, to examine its occasional inconstancy, to figure yearly/month to month precipitation over some topographical region Recent exploration exercises in artificial neural organization have indicated that ANNs have amazing design of classification with proper behavioural pattern identification capacity. Roused by organic frameworks, especially by investigation into the human mind, ANNs can gain from and sum up for a fact.

In present days, ANNs are being utilized for a wide assortment of undertakings in various fields of business, industry and science. The current work convincingly shows the upsides of utilizing ANN over that of ARIMA procedure to demonstrate the precipitation conduct.

1.1 Study area

The data for annual rainfall Coastal Andhra, which is lies between 12°41' and 19.07° N latitude and 77° and 84°40' E longitude, is being used for the present study. Coastal Andhra summer monsoon trend pattern is dynamic based on the climatic changes. In India, June to September rainfall is vital for the Agricultural needs disaster management, economy development, and hydrological strategies of the country. Still agriculture is one of the major sectors for Indian economic development. Floods and famine due to heavy rainfall and in some cases like droughts due to weak monsoons becomes crucial for the country. Hence, it is significant to observe meticulously the pattern of rainfall variation across the country on monthly, seasonal and annual time scale is necessary.

This long-term rainfall time series data is collected from Indian Meteorological Department (http://www.data.gov.in). This consists of the mean annual rainfall from year 1901 to 2017 (117 years) as shown in Fig.2. The pattern of rainfall is nonlinear over the periods. The first 100 years of monthly rainfall time series data is used for training model and the remaining 17 years considered for prediction by using traditional statistical model ARIMA and contemporary machine learning technique ANN models.

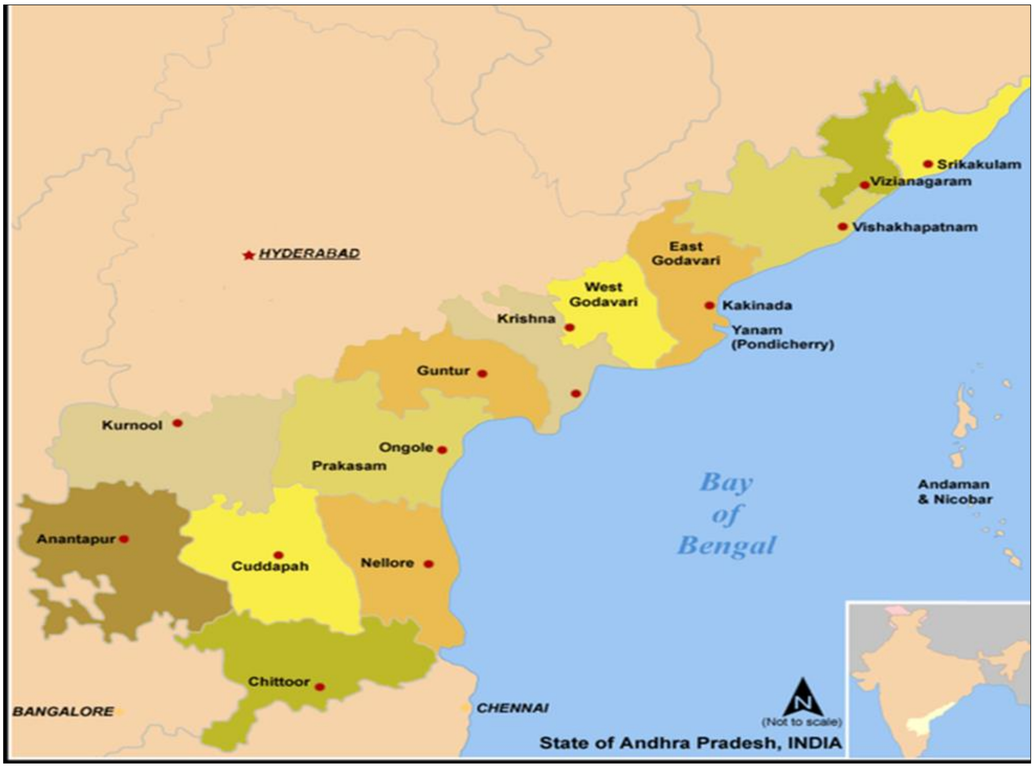

Fig 1: Coastal Andhra

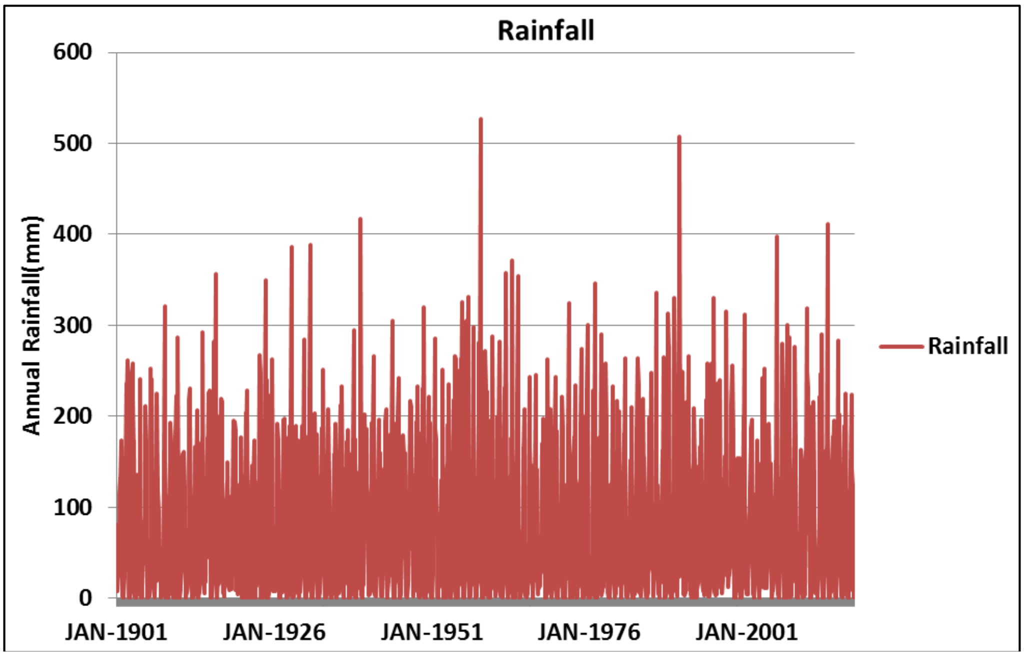

Fig 2: Graphical representation of annual rainfall data bounded by latitude 12°41' and 19.07°N latitude and 77° and 84°40'E longitude

2. 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.

3. Forecasting methods and parameters

3.1 Arima model

ARIMA model was introduced by Box & Jenkins (1970) as predicting technique which is still prevalent in many areas. Especially for weather and rainfall forecasting this is one of the best conventional models.

The autoregressive integrated moving average ARIMA model of the time series $\{r_1, r_2, r_3 \ldots\}$ is defined as

$$\varphi(B)\,\Delta^d r_t = \theta(B)\, e_t \qquad \text{-----(1)}$$

Where r_t and e_t are Average annual rainfall and random error of the time series data at time t. B is the lag variable explicated by $Br_y = r_{y-1}$ and $\Delta = 1 - B$ where d is the order of difference.

The φ (B) and θ (B) of order p and q are defined as

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots\cdots\cdots \phi_p B^p \qquad \text{---- (2)}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots\cdots\cdots \theta_q B^q \qquad \text{---- (3)}$$

Where $\varphi_1, \varphi_2 \ldots \varphi_p$ are the autoregressive coefficients and $\theta_1, \theta_2 \ldots \theta q$ are the moving averages coefficients

The primary task in the ARIMA (p,d,q) modelling is to test whether the time series data is stationary or not. If the data is non-stationary then it should be converted into a stationary time series by applying appropriate degree of differentiation by selecting proper value of d. The values of p is the number of autoregressive terms, and q is the number of lagged forecast errors in the prediction equation that are chosen by examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series data.

3.2 ANN Model

An ANN is a tremendous parallel distributed processor that has a neural affinity for keeping away the experimental information and making it accessible for additional utilization. It looks like the human cerebrum whose speed and effectiveness has been continually intriguing to scientists for a serious long time. The mission to comprehend these cycles and to take care of the related issues has prompted the improvement of ANN strategy. Neural networks basically include a nonlinear displaying approach that gives a genuinely precise general estimate to any capacity. Its control originates from the parallel processing of the data from information. No previous presumption of the model structure is needed in the model structure measure. All things considered, the organization model is generally controlled by the qualities of the information. Single shrouded layer feed forward network is the most generally utilized model structure for time series modelling and predicting. As forecasting is accomplished via forecasting of future behavioural pattern from examples of previous behaviour, it is a perfect modelling area for neural networks, at least in principle [Koizumi k 1999]. In the forward process the inputs are functioning through input layer and its impact is spread through network, layer by layer. The net impact is processed as the weighted amount of the yield of the neurons of the past layer. The amount of squared deviation of the yield from the objective incentive at the hubs of the yield layer characterizes the blunder signal that will be spread back to past layers with the end goal that the boundaries are acclimated to limit the mistake in additional calculations

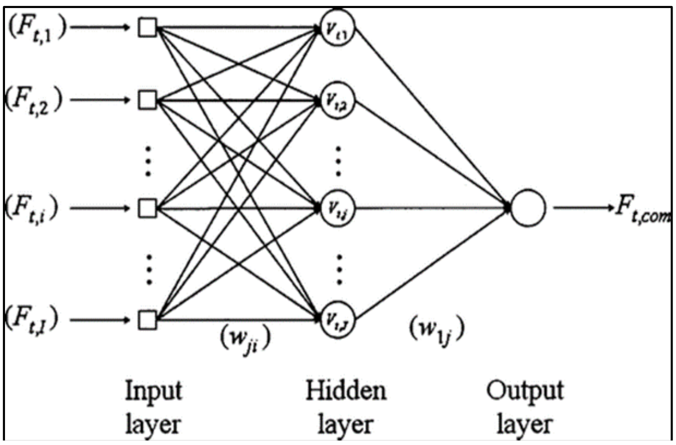

Fig 3: ANN design with four input and one output for mean annual rainfall data.

As shown in Fig.3, an ANN contains of different stratums of neurons. The model is characterized by a network of three stratums of simple processing units, which are associated to each other. The first stratum collects the input data and it is called an input layer. The last stratum produces the output information can be known as output layer. Between output and input layers there are many hidden layers and the

Information is transferred through the networks between nodes in different layers.

4. Data analysis and model selection

One of the greatest challenges in the contemporary data analysis is to extract the meaningful inferences about a complex data from a single dignified parameter. To measure the forecasting performance of models, the data can divide into two parts like training and testing. The training data is normally used for the development of model selection and the testing is used to evaluate the proposed model. In the present study the first 100 years of monthly rainfall time series data is used for training model and the remaining 17 years considered for prediction by using traditional statistical model ARIMA and contemporary machine learning technique ANN models.

4.1 Performance evaluation criteria

Model performance evaluation is crucial part when we compare with two different models. In conventional time series models and machine learning techniques error metrics are the real aspects in order to define the best model. The major error metrics are

Mean Absolute Error MAE

$$MAE = \frac{\sum_{t=1}^{n} |E_t|}{\eta} \quad \text{------- (4)}$$

Mean Absolute percentage Error (MAPE)

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \left|\frac{Et}{Yt}\right|}{n} \times 100 \quad \text{------- (5)}$$

Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n} Et^2} \quad \text{------- (6)}$$

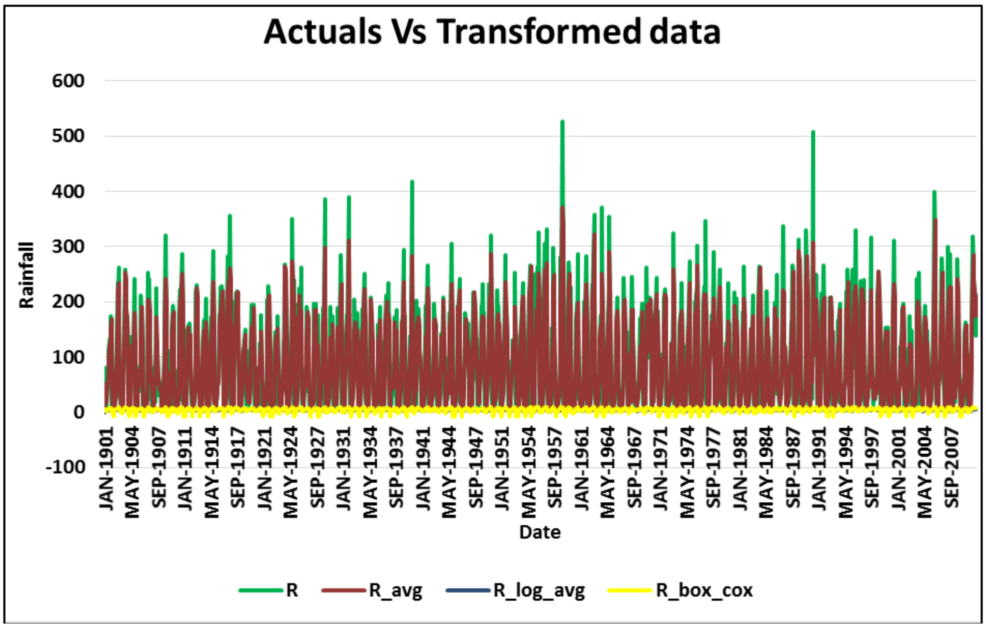

Fig 4: Actual and transformed mean annual rainfall for training data set

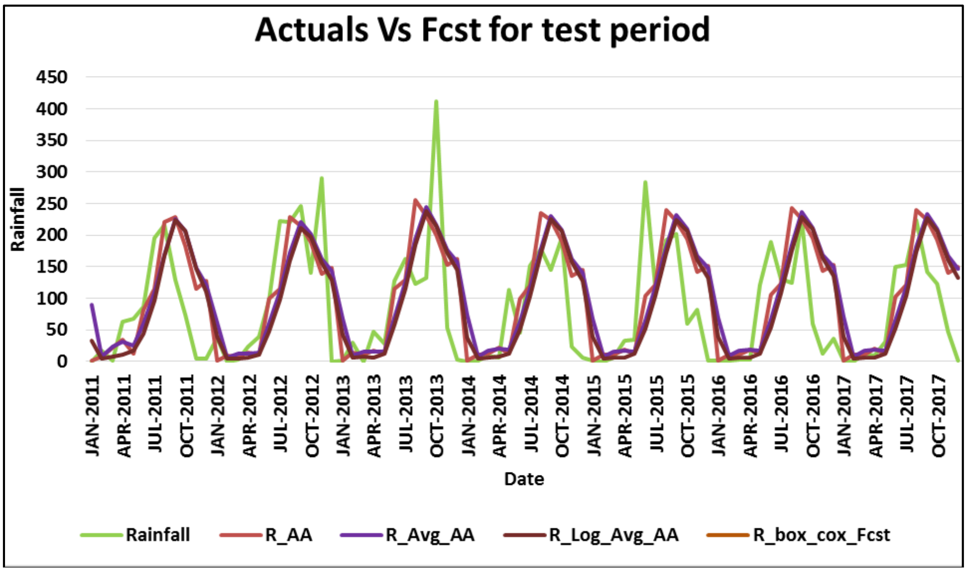

Fig 5: Predicted and observed mean annual rainfall for testing data set using ARIMA model

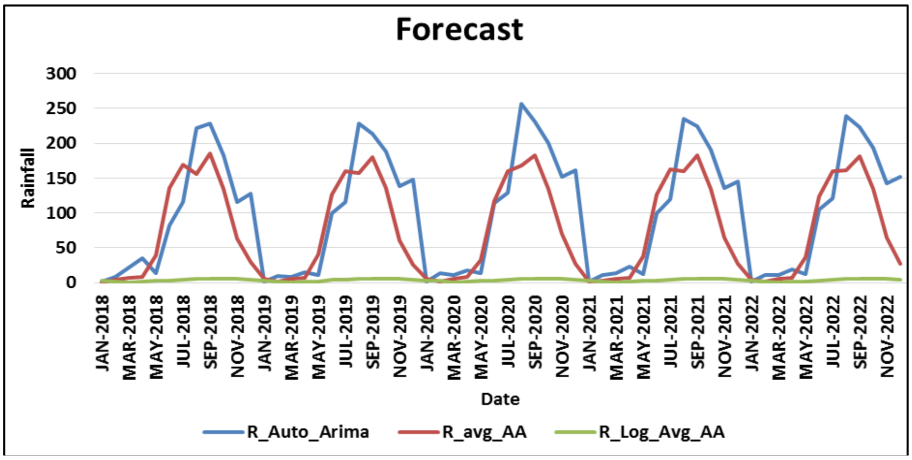

Fig 6: Predicted mean annual rainfall for testing data set using ARIMA model

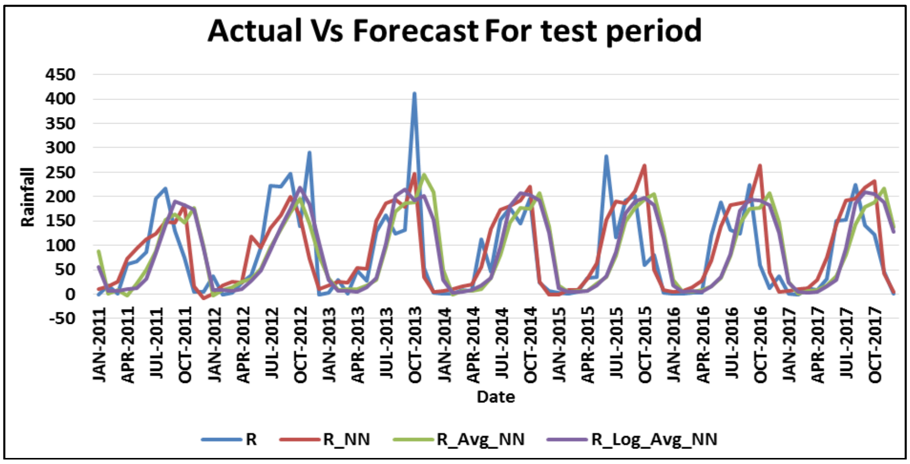

Fig 7: Predicted and observed mean annual rainfall for testing data set using ANN model

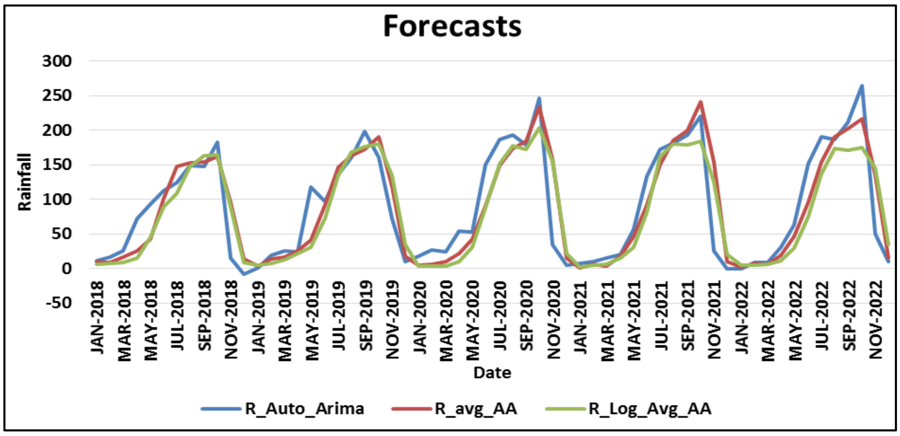

Fig 8: Predicted mean annual rainfall for testing data set using ANN model

4.2 Results and Discussion

Figure 4 explains that the plot of Transformed data and Actual (observed) rainfall data from 1901 to 2017. Figure 5 and figure 7 show the plot of predicted and observed rainfall for testing data set using ARIMA and ANN respectively. Similarly figure 6 and figure 8 clearly shows that predicted mean rainfall using ARIMA and ANN from Jan 2018 to Dec 2022.Based on this plots we observe that ANN model forecasting is performing well than the ARIMA.

4.3 Error metrics

Table 1: The error metrics of ARIMA and ANN models

Model	Metric	ME	RMSE	MAE	MPE	MAPE	ACF1	Theils U
Arima	R	18.48	76.78	54.47	-52.93	124.96	0.13	1.10
	R_ Avg	20.46	82.79	63.18	-5.77	88.94	0.34	2.19
	R_ Log_ Avg	10.06	80.66	60.51	-50.40	122.91	0.34	3.64
ANN	R	7.69	59.03	37.07	-188.90	322.90	-0.05	0.98
	R_ Avg	4.77	88.30	63.95	-44.56	127.39	0.33	3.01
	R_ Log_ Avg	3.50	84.50	61.90	-68.64	136.87	0.33	5.02

The error metrics of ARIMA and ANN models in are shown in Table 1. The table shows that the ANN model performs well in compare with the ARIMA model. The RMSE error for ARIMA model is 76.78 and for ANN model is 59.03 respectively. And the other performance metric such as Theils U represent that the ANN forecast is better to ARIMA forecast. The RMSE error metric is progressively increases when calculate average and log average values Therefore, this research founds that ANN method is a prediction tool to model and forecast annual rainfall than the ARIMA model.

5. Conclusions

Nonlinear behavioural pattern of annual rainfall data has been studied using ARIMA and ANN techniques. Monthly rainfall data during the period of 1901- 2017 of Coastal Andhra region was utilized to train and test the models. Root mean square error and Theils U values for the rainfall data was used to find out the best model fit for forecasting. This research discloses that ANN model will be the significant prediction tool to predict the rainfall, which performs better when compare with the ARIMA model.

6. References

1. Ahaneku IE, Otache YM. Stochastic characteristics and modeling of monthly rainfall time series of Ilorin, Nigeria. Open J. Mod. Hydrol 2014;4:67-79.
2. Aitken AP. Assessing systematic errors in rainfall runoff models, J of Hydrology 1973;20:121-136.
3. Akaike H. Information theory and an extension of the maximum likelihood principle. In Proceedings of the Second International Symposium on Information Theory, Tsahkadsor, Armenia 1973,267–281.
4. Bang S, Bishnoi R, Chauhan AS, Dixit AK, Chawla I. Fuzzy logic based crop yield prediction using temperature and rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. In Proceedings of the 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India 2019,1-6.
5. Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis Forecasting and Control; Prentice-Hall: Upper Saddle River, NJ, USA, 1994. ISBN 0253-9624.
6. Breinl K, Turkington T, Stowasser M. Simulating daily precipitation and temperature: Aweather generation framework for assessing hydro meteorological hazards. Meteorol 2015;22:334–347.
7. Burlando P, Rosso R, Cadavid LG, Salas JD. Forecasting of short-term rainfall using ARMA models. J Hydrol. 1993;144:193-211.
8. Chang TJ, Kavvas ML, Delleur JW. Daily precipitation modelling by discrete autoregressive moving average processes. Water Resour. Res 1984;20:565-580.
9. Chatfield C. Time-Series Forecasting, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2000. ISBN 978-1-4200-3620-6.
10. Cigizoglu HK. Forecasting of meteorological data by Artificial Neural Networks, In: Advances in Soft Computing (Proc. Sixth Int. Conf. On Neural Networks and Soft Computing, Zakopane, Poland 2002.
11. Cujia A, Agudelo-Castañeda D, Pacheco-Bustos C, Teixeira EC. Forecast of PM10 time-series data: A study case in Caribbean cities. Atmos. Pollut. Res 2019;10:2053-2062.
12. Sutiksno DU, Gio PU, Caraka RE, Ahmar AS. “Brief Overview of STATCAL Statistical Application Program,” J Phys. Conf. Ser 2018,1028.
13. Dabral PP, Murry MZ. Modelling and forecasting of rainfall time series using SARIMA. Environ. Process 2017;4:399-419.
14. Etuk EH, Moat IU, Chims BE. Modelling monthly rainfall data of Port Harcourt, Nigeria by Seasonal Box-jenkins methods. Int. J Sci 2013;2:1-8.
15. Fadhilah Yusof, Kane IL, Yusop Z, Structural break or long memory: an empirical survey on daily rainfall data sets across Malaysia. Hydrol. Earth Syst. Sci 2013;17:1311-1318,
16. Hazards 2016 81, 1811–18272. Venkata, R.R.; Krishna, B.; Kumar, S.R.; Pandey, N.G. Monthly rainfall prediction using wavelet neural network analysis. Water Resour. Manag 2013;27:3697-3711.
17. Wu J, Jin L. “Daily rainfall prediction with SVR using a novel hybrid PSO-SA algorithms,” in Communications in Computer and Information Science 2011;163CCIS:508-515.
18. Lata K, Misra AK. The influence of forestry resources on rainfall: A deterministic and stochastic model. Appl. Math. Model 2020;81:673-689.
19. Lee T. Stochastic simulation of precipitation data for preserving key statistics in their original domain and application to climate change analysis. Theor. Appl. Climatol 2015;124:91-102.
20. 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.
21. 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.
22. 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.
23. Pandey OP, Somvanshi VK, Agrawal PK, Kalanker NV, Prakash MR. Modelling and prediction of rainfall using artificial neural network and ARIMA techniques J Ind. Geophys. Union 2006;10(2):141-15.
24. 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.
25. 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.
26. Caraka RE *et al.*, “Ecological Show Cave and Wild Cave: Negative Binomial Gllvm’s Arthropod Community Modelling,” Procedia Comput. Sci 2018;135:377-384.
27. 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)
28. Singh VP, Chowdhury PK, Comparing some methods of estimating mean area rainfall. Water Resources Bulletin 1986;22(2):275-282.
29. 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.

30. Valbuena DL. Geomorfología y Condiciones Hidráulicas del Sistema Fluvial del RÍO SINÚ. Integración Multiescalar. 1945–1999–2016. Ph.D. Thesis, Universidad Nacional de Colombia, Bogotá, Colombia, 2017.
31. Sanusi W, Jemain AA, Zin WZW, Zahari M. "The Drought Characteristics Using the First-Order Homogeneous Markov Chain of Monthly Rainfall Data in Peninsular Malaysia," Water Resour. Manag 295,1523–1539.
32. Wang S, Feng J, Liu G. Application of seasonal time series model in the precipitation forecast. Math. Comput. Model 2013;58:677–683.