International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2025; 10(5): 99-103

www.mathsjournal.com
Received: 10-03-2025
Accepted: 12-04-2025

Addanki Himaja
Department of Statistics and Computer Applications, SV Agricultural College, Acharya NG Ranga Agricultural University, Tirupati, Andhra Pradesh, India

Dr. Shaik Nafeez Umar
Department of Statistics and Computer Applications, SV Agricultural College, Acharya NG Ranga Agricultural University, Tirupati, Andhra Pradesh, India

Shaik Shameem
Department of Statistics and Computer Applications, SV Agricultural College, Acharya NG Ranga Agricultural University, Tirupati, Andhra Pradesh, India

G Mohan Naidu
Professor and Head, Department of Statistics and Computer Applications, SV Agricultural College, Tirupati, Andhra Pradesh, India

S Rajeswari
Assistant Professor, Advanced Post Graduate Centre, Lam, Guntur, Andhra Pradesh, India

Corresponding Author:
Addanki Himaja
Department of Statistics and Computer Applications, SV Agricultural College, Acharya NG Ranga Agricultural University, Tirupati, Andhra Pradesh, India

Rainfall prediction using machine learning models in Anantapuramu District, Andhra Pradesh

Addanki Himaja, Shaik Nafeez Umar, Shaik Shameem, G Mohan Naidu and S Rajeswari

DOI: https://www.doi.org/10.22271/maths.2025.v10.i5b.2038

Abstract
Rainfall is a critical determinant of agricultural productivity and water resource management in India, where approximately 60% of cultivated land relies on monsoon rains. In 2024, India recorded an average annual rainfall of 1190 mm. However, Andhra Pradesh experienced slightly below-average precipitation, receiving only 920 mm. The situation was particularly concerning in the Anantapuramu District a semi-arid and drought-prone region where the annual rainfall was limited to just 470 mm. In such vulnerable areas, timely and accurate rainfall forecasting is essential for strategic planning and sustainable agricultural practices.
This study aims to forecast monthly rainfall in the Anantapuramu district using historical data from 1988 to 2024. Three models were employed for comparison: the ARIMA (Autoregressive Integrated Moving Average) model, suitable for linear time series forecasting; the Artificial Neural Network (ANN), capable of modeling non-linear dependencies; and the Wavelet-ANN model, which combines wavelet decomposition with ANN to capture both time-frequency and non-linear patterns in the data. Model performance was evaluated using Root Mean Squared Error (RMSE) as the key metric. Among the three, the ANN model exhibited the lowest RMSE value, indicating superior prediction accuracy. These results highlight the effectiveness of ANN in improving rainfall forecasting, which is crucial for enhancing agricultural resilience and mitigating climate-related risks in drought-prone regions like Anantapuramu.

Keywords: ARIMA, ANN, Wavelet-ANN, RMSE, Rainfall prediction

Introduction
Rainfall is vital in India's agriculture and water resource management, where nearly 60% of the farmland depends on monsoon rains. In 2024, India received an average annual rainfall of 1190mm. Andhra Pradesh experienced slightly below-average rainfall, recording 920 mm. The situation was more concerning in Anantapuramu district, one of the most drought-prone and semi-arid regions in the state, which received only about 470 mm of rainfall. In such drought-prone regions, timely and accurate rainfall prediction becomes essential. For forecasting rainfall, there are several models, but ARIMA plays a crucial role in identifying linear trends in time series data and might miss out some hidden patterns in the data. For finding the complex patterns that are hidden in the data usage of ANN was done in this study and Wavelet-ANN (Artificial Neural Network) was also used for rainfall prediction as it is a powerful modeling technique that combines the time-frequency analysis power of wavelets with the nonlinear learning ability of ANN.
This research explored the use of ARIMA, ANN and wavelet modeling for advanced rainfall prediction in Anantapuramu, emphasizing its potential to transform agricultural practices and improve the livelihoods of farmers. Through a detailed analysis of recent climatic data and modeling techniques, we aim to demonstrate the efficacy of wavelet models in addressing the unique challenges faced by this drought-prone region.

Materials and Methods

The monthly rainfall data of the Anantapuramu district of Andhra Pradesh is collected from two sources: from January 1988 to December 2020, the data was obtained from ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), and from January 2021 to December 2024, it was sourced from official government websites such as APWRIMS (https://apwrims.ap.gov.in/). The Statistical Tools that are used are Descriptive Statistics, ARIMA, ANN and Wavelet ANN. Box Pierce Test and BDS Test was conducted to find autocorrelation and non-linearity in the data respectively.

Methodology

ARIMA

ARIMA is a popular statistical model used for forecasting time series data. It combines three components:

- **AR (Autoregressive):** Uses the relationship between an observation and a number of lagged observations (previous time steps).
- **I (Integrated):** Differencing the raw observations to allow for the time series to become stationary.
- **MA (Moving Average):** Uses the relationship between an observation and a residual error from a moving average model applied to lagged observations.

ARIMA Model Specification:

The ARIMA model is denoted as ARIMA (p, d, q), where:
P: Number of lag observations (AR term)
D: Number of times that the raw observations are differenced (I term)
Q: Size of the moving average window
(MA term)
Formula

The ARIMA model can be expressed as:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Where,
Y_t is the value of the time series at time t.
ϕ terms are the coefficients of the AR part.
ϕ terms are the coefficients of the MA part.
ε_t is the white noise error term

ANN (Artificial Neural Networks)

Artificial Neural Networks (ANNs) are sophisticated computational models that draw inspiration from the neural networks of the human brain. These networks are made up of interconnected nodes, commonly referred to as neurons, which are arranged in layers: An input layer, one or more hidden layers, and an output layer. ANNs excel at identifying and modeling complex nonlinear relationships within data, making them valuable tools in various fields such as machine learning, data analysis, and artificial intelligence.

Architecture

- **Input Layer:** Receives input features.
- **Hidden Layer(s):** Processes inputs with activation functions to learn patterns.
- **Output Layer:** Produces the forecasted output.

Training

- ANNs are trained using a dataset by adjusting the weights and biases through a process called back propagation, minimizing the difference between predicted and actual outputs.

Formula

For a single-layer feed forward ANN, the output Y can be computed as:

$$Y = f\left(\sum_{i=1}^{n} w_i X_i + b\right)$$

Where,
Y is the output of the neuron.
X_i are the input features.
w_i are the weights associated with the inputs.
b is the bias term.
f is the activation function (e.g., sigmoid, ReLU).

Wavelet Transform Approach

To conduct the wavelet transform, we utilized the Discrete Wavelet Transform (DWT) instead of the Maximal Overlap Discrete Wavelet Transform (MODWT). The DWT was chosen for its straightforward implementation and its ability to facilitate the analysis of the signal across discrete scales.

Implementation

The wavelet transform was performed using the `wt. Filter` function in R, with the following parameters set:

- **Filter:** Filter="la8"
- **MODWT:** modwt=FALSE
- **Level of Decomposition:** level=1

This setup enabled the initial decomposition of the time series data into two main components, the approximation coefficients, which capture low-frequency trends, and the detail coefficients, which reflect high-frequency variations.

Wavelet-ANN: This approach involves breaking down the time series into various frequency components using the wavelet transform, training separate artificial neural networks on each of these components, and then reassembling the forecasted series through an inverse wavelet transform. This method allows for capturing both linear and nonlinear patterns, thereby enhancing accuracy.

Performance Evaluation Metrics: To evaluate the models, the following performance metrics were used.

Root Mean Square Error (RMSE): Measures the square root of the average squared differences between actual and predicted values.

Formula

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}$$

Where,

y_i and $\hat{y}_i$ are the actual and predicted values of response variable, respectively.

Mean Absolute Error (MAE): Calculates the mean of absolute differences between predicted and actual values.

Formula

$\mathrm{MAE}=\frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t|$

Where,

N=the total number of observations,

Y_t=the actual value at time t,

$\hat{Y}_t$ =the predicted value at time t,

$|Y_t-\hat{Y}_t|$=the absolute error between the actual and predicted values.

Mean squared error (MSE)

The mean square error (MSE) is the average of the sum of squared error values and is given as,

$$MSE = \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{N}$$

Where, Y_i is the actual value, $\hat{Y}_i$ is the predicted value and N is the number of observations

Results and Discussions

Table 1: Descriptive statistics of monthly rainfall data of Anantapuramu district

Statistic	Value
Mean	49.988
Standard Error	3.22631
Median	18
Mode	0
Standard Deviation	67.9826
Sample Variance	4621.64
Kurtosis	3.90159
Skewness	1.87934
Range	395.3
Minimum	0
Maximum	395.3
Sum	22194.7
Count	444

The descriptive statistics for monthly rainfall in Anantapuramu district were summarized, revealing a mean rainfall of 49.99 mm. However, the median was merely 18 mm, and the mode was 0 mm, indicating that many months experienced little to no rainfall. This disparity, coupled with a positive skewness of 1.8793, suggested a right-skewed distribution, where a few months of high rainfall affected the average significantly. The kurtosis value of 3.90 indicated a leptokurtic distribution, signifying that extreme rainfall events were more frequent than expected. The standard deviation was calculated at 67.98 mm, and the variance was determined to be 4621.64 mm², reflecting a high variability in rainfall. The range between the minimum of 0 mm and the maximum of 395.3 mm was found to be 395.3 mm. During the study period, the total monthly rainfall amounted to 22194.7 mm, derived from 444 months of data. Overall, these statistics suggested that rainfall in Anantapuramu district was highly variable and sporadic, with frequent extreme events. To forecast the monthly rainfall in Anantapuramu district, the ARIMA (Autoregressive Integrated Moving Average) model was first applied. By utilizing the auto.arima() function from the forecast package in R, the analysis determined the most appropriate model to be ARIMA (5,0,1).

The fitted equation for the ARIMA (5,0,1) model is given below:

$$Y_t = 49.749 + 0.724 \cdot Y_{t-1} - 0.013 \cdot Y_{t-2} - 0.119 \cdot Y_{t-3} + 0.019 \cdot Y_{t-4} - 0.195 \cdot Y_{t-5} + \varepsilon_t + 0.494 \cdot \varepsilon_{t-1}$$

Further, the ANN model was applied which used past rainfall observations as input with a lag of 12 periods (p=12) to capture temporal dependencies. It features a hidden layer with 10 neurons to model nonlinear relationships and is trained over 22 repetitions to improve performance and minimize over fitting. The architecture is structured as 12-10-1, adjusting 141 weights for optimal forecast accuracy. Additionally, a Wavelet-Artificial Neural Network (Wavelet-ANN) model was employed to enhance forecast accuracy by capturing both short-term fluctuations and long-term trends in the rainfall data.

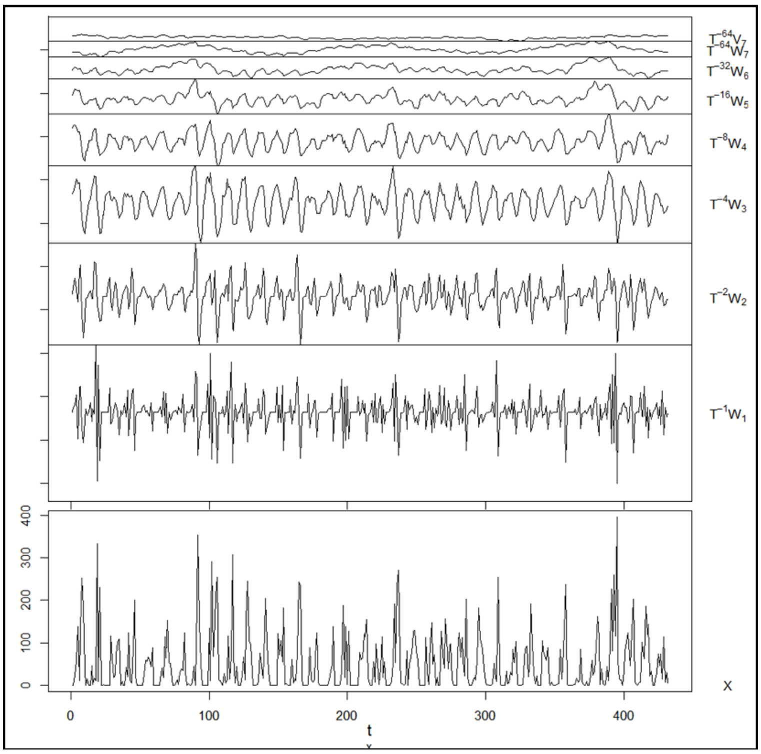


Fig 1: Wavelet decomposition of original data

This plot showed a wavelet decomposition of the time series data using the Haar filter. The bottom part (marked as X) displayed the original data, while the panels above (W1 to W7 and V7) broke it down into layers. W1 captured fast changes or noise, while W5 to W7 revealed slower trends. The top panel, V7, represented the smoothest version of the data, highlighting the overall trend without minor fluctuations. As you move up the plot, the lines became smoother, allowing for a better understanding of both quick changes and long-term behaviour. The Haar filter effectively simplified the data, enabling analysis at different time scales.

Table 2: Performance comparison metrics of models

Model	RMSE (Train)	RMSE (Test)	MAE (Train)	MAE (Test)	MSE (Train)	MSE (Test)
ARIMA (5,0,1)	61.0111	69.9689	44.0961	55.5706	3722.353	4895.966
ANN (12,10)	28.7019	54.87233	16.1054	44.63861	823.796	3010.973
Wavelet ANN	29.11727	52.98941	15.86718	43.60125	847.8155	2807.878

The evaluation of three forecasting models presented in Table 2 demonstrates the performance metrics of ARIMA (5, 0, 1), ANN (12, 10), and Wavelet-ANN using RMSE, MAE, and MSE. The results clearly indicate that the ANN model outperformed the other models, particularly on the training data, achieving the lowest RMSE of 28.70, MAE of 16.11, and MSE of 823.80.

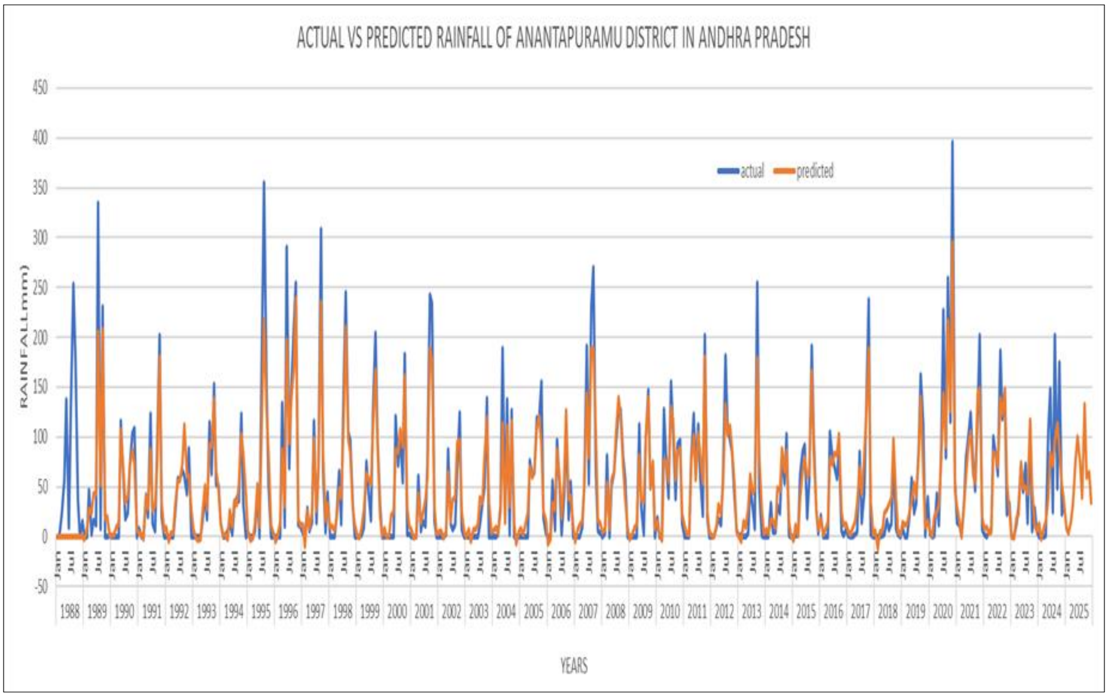
On the test set, the Wavelet-ANN model performed slightly better, recording a test MSE of 2807.88. However, the ANN model closely followed, with a test MSE of 3010.97 and a test RMSE of 54.87. These findings reinforce the conclusion that the ANN (12, 10) model is the best choice for forecasting rainfall in Anantapuramu, given its superior error metrics across both training and test datasets. As such, it is recommended for further applications in rainfall prediction.

Further analysis using the Box-Pierce test on the residuals of the ANN model indicated no significant autocorrelation present. The low test statistic (X^2=0.8933) and a corresponding p-value of 0.3446, which is greater than the common significance level of 0.05, led to the acceptance of the null hypothesis. This suggests that the residuals are approximately white noise, indicating that the model has adequately captured the dependencies within the data without any remaining autocorrelation.

On the other hand, the results from the BDS (Brock-Dechert-Scheinkman) test on the residuals of the ANN model pointed to the presence of significant nonlinear dependencies. For both embedding dimensions of 2 and 3, the standard normal test statistics were consistently high across all epsilon values, ranging from 2.4621 to 8.2409. The corresponding p-values were either 0 or very small, leading to the rejection of the null hypothesis of independent and identically distributed (I.I.D.) residuals. This indicated that while the ANN model captures much of the data's structure, some nonlinear patterns within the residuals remain unexplained. Therefore, additional refinement of the model or the incorporation of nonlinear components may enhance forecasting accuracy even further.

Table 3: 2025 Predicted Rainfall (mm) in Anantapuramu District by using the ANN

Months of 2025	Predicted Rainfall (mm)
January	3.56053
Febuary	0
March	8.373589
April	43.33985
May	60.07714
June	74.27584
July	59.33712
August	18.69847
September	85.18256
October	85.35163
November	71.09254
December	24.05259



Graph 1: Actual vs Predicted Rainfall of Anantapuramu district in Andhra Pradesh

Conclusion

This study evaluated the performance of various models like ARIMA, ANN, and Wavelet ANN for predicting rainfall in Anantapuram district in Andhra Pradesh. Among the models analyzed, the Artificial Neural Network (ANN) model demonstrated superior accuracy, outperforming the other models based on the performance metrics such as RMSE, MAE and MSE. Diagnostic checks revealed no significant autocorrelation in the residuals of ANN, indicating a good fit. However, the presence of significant nonlinear dependencies in the residuals suggests that while ANN effectively captures much of the underlined structure in the data, sum complex nonlinear patterns may still remain unexplained. Despite this the ANN model remains the most robust and reliable approach for rainfall prediction in Anantapuram District, Andhra Pradesh.

References

1. Ajmera TK, Rastogi AK. Artificial neural network application on estimation of aquifer transmissivity. Journal of Spatial Hydrology. 2008;8(2):15-31.
2. Bhattacharya B, Price RK, Solomatine DP. Machine learning approach to modeling sediment transport. Journal of Hydrologic Engineering. 2007;133(4):440-450.
3. Cigizoglu HK. Forecasting of meteorological data by artificial neural networks. In: Advances in Soft Computing: Proceedings of the Sixth International Conference on Neural Networks and Soft Computing, 2002.
4. Dabral PP, Murry MZ. Modelling and forecasting of rainfall time series using SARIMA. Environmental Processes. 2017;4:399-419.
5. Farajzadeh J, Alizadeh F. A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. Journal of Hydroinformatics. 2018;20(1):246-262.
6. Fausett L. Fundamentals of neural networks. Englewood Cliffs, NJ: Prentice Hall, 1994.
7. Ghamariadyan M, Imteaz M, Mekanik F. Application of wavelet transform and artificial neural networks for predicting rainfall time series with various variable inputs.
8. Gomes E, Progenio M, Holanda P. Modeling with artificial neural networks to estimate daily precipitation in the Legal Amazon, 2023. Available from: 10.21203/rs.3.rs-3822672/v1.
9. Gürsoy Ö, Engin SN. A wavelet neural network approach to predict daily river discharge using meteorological data. Measurement and Control. 2019;52(5-6):599-607.
10. Hema Sekhar P, Poola K, Raja Sekhar K, Bhupathi Naidu M. Modelling and prediction of coastal Andhra rainfall using ARIMA and ANN models. International Journal of Statistics and Applications. 2020;5(6):104-110.
11. Hussein E, Ghaziasgar M, Thron C, Vaccari M, Jafta Y. Rainfall prediction using machine learning models: Literature survey. 2022. Available from: DOI 10.1007/978-3-030-92245-0_4.
12. Mishra CS, P, S, M R, N, K V. Probability analysis and rainfall forecasting using ARIMA model. MAUSAM. 2023;74(4):1081-1092.
13. Narasimhareddy K, James A, Gautam S. Rainfall variability in Rayalaseema region of Andhra Pradesh. International Journal of Current Microbiology and Applied Sciences. 2020;9(12):124-131.
14. Pandey OP, Somvanshi VK, Agrawal PK, Kalanker NV, Prakash MR. Modelling and prediction of rainfall using artificial neural network and ARIMA techniques. Journal of Indian Geophysical Union. 2006;10(2):141-151.
15. Poulton MM. Neural network as an intelligence amplification tool: A review of applications. Geophysics. 2002;67(3):979-993.
16. Paul RK, Paul AK, Bhar LM. Wavelet-based combination approach for modeling sub-divisional rainfall in India. Theoretical and Applied Climatology. 2020;139(3):949-963.
17. Ramana R, Krishna B, Kumar S, Pandey N. Monthly rainfall prediction using wavelet neural network analysis.

Water Resources Management. 2013;27:10.1007/s11269-013-0374-4.

18. Shoaib M, Shamseldin AY, Melville BW. Comparative study of different wavelet-based neural network models for rainfall-runoff modeling. Journal of Hydrology. 2014;515:47-58.
19. Singh B, Rajpurohit D, Vasishth A, Singh J. Probability analysis for estimation of annual one day maximum rainfall of Jhalarapatan area of Rajasthan. Plant Archives. 2012;12:1093-1100.
20. Singh VP, Chowdhury PK. Comparing some methods of estimating mean area rainfall. Water Resources Bulletin. 1986;22(2):275-282.