Journal of Agrometeorology

(A publication of Association of Agrometeorologists)

ISSN : 0972-1665 (print), 2583-2980 (online)

Vol. No. 27 (3) : 395-398 (September - 2025)

https://doi.org/10.54386/jam.v27i3.3038

https://journal.agrimetassociation.org/index.php/jam

Short communication

Application of artificial intelligence and statistical recurrent models in predicting rainfall: A case study of Ludhiana, Punjab

SUBHRAJYOTI BHATTACHARJEE[1], NILESH BIWALKAR[1], KOYEL SUR[2*], SAMANPREET KAUR[1] and SOM PAL SINGH[3]

[1]Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, PAU, Ludhiana,
[2]Punjab Remote Sensing Centre, Ludhiana,
[3]Department of Climate Change and Agriculture Meteorology, PAU, Ludhiana,
Corresponding Author: koyelsur3@gmail.com

Sustainable management of agriculture is intricately linked to water resource management and climatic phenomenon. Monitoring precipitation is one of the major requirements for hydrological functioning of semi-arid and arid environments where surface water bodies are comparatively scarce and rainfall is scanty. Irregular or insufficient rainfall disrupts agriculture, causing crop failure, reduced yields, and increased dependency on groundwater irrigation, which in turn depletes water tables (Priyan *et al.,* 2021). Accurate forecasting of rainfall is quite difficult due to the complex and dynamic nature of the atmosphere in regional scale (Baig *et al.,* 2024). Due to stochastic nature of meteorological time-series datasets with trend and seasonal components over time, autoregressive integrated moving average (ARIMA) model generally can address some issues, by transforming non-stationary time-series data into stationary time-series data (Kothiyal *et al.,* 2025). Seasonal auto regressive integrated moving average (SARIMA) was introduced to incorporating seasonal components making it more suitable for data with repeating seasonal patterns (Amshi and Prasad, 2023). However, even SARIMA assumes linearity in data, which often limits its ability to capture complex and nonlinear climate interactions. Nonlinear Autoregressive (NAR) Neural Networks have gained prominence in recent times (Sarkar *et al.,* 2023). This NAR network is a type of artificial neural network (ANN) specifically designed for time-series forecasting. This particular study aims to compare the performance in Nonlinear Autoregressive (NAR) and Seasonal Autoregressive Integrated Moving Average (SARIMA) model results to predict the rainfall of Ludhiana, Punjab.

The monthly rainfall data of Ludhiana (30.89°N, 75.80°E) from 1970-2022 were collected from the Department of Climate Change and Agriculture Meteorology, PAU, Ludhiana. The rainfall data undergoes statistical standardization before Fourier transformation and thereafter the autocorrelation function (ACF) was computed using the correlogram formula:

$$r_\tau = \frac{n}{n-\tau}\sum_{j=1}^{n-\tau}\frac{(z_j-\bar{z})(z_{j+\tau}-\bar{z})}{\sum_{j=1}^{N}(z_j-\bar{z})^2}$$

where r_τ represents the correlation at lag τ, $\bar{Z}$ is the mean of the standardized series and n is the number of observations. The correlogram values range from -1 to +1, indicating negative or positive correlations, respectively.

SARIMA model

The general formulation of the SARIMA model can be expressed as SARIMA (p, d, q) × (P, D, Q) _s, where p, d, q represents the non-seasonal autoregression (AR), differencing and moving average (MA) orders where P, D, Q correspond to the seasonal components with seasonality period s. The model follows the equation:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D Y_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t$$

where Y_t is the actual value at time t,B is the backshift operator, represents random error, and Φ_P, Φ_p, Θ_Q and Θ_q are the parameters for seasonal and non-seasonal autoregressive and moving average terms. During model identification, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots help to determine the appropriate values of p, d, q, P, D, Q and s.

***Article info - DOI**: https://doi.org/10.54386/jam.v27i3.3038*

Received: 5 May 2025; Accepted: 8 July 2025; Published online : 1 September 2025

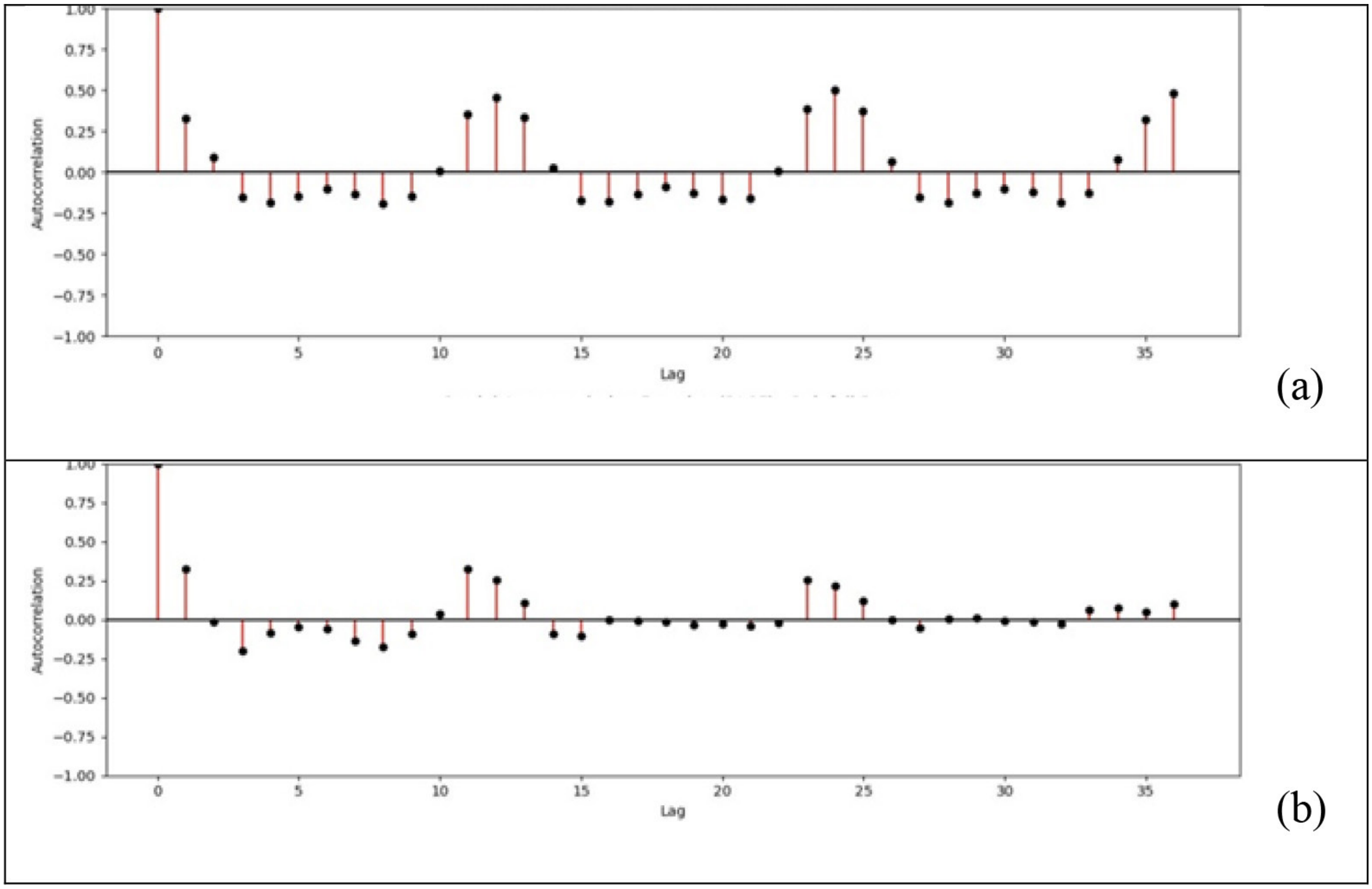

Fig. 1: (a) Autocorrelation function (ACF) and (b) partial autocorrelation function (PACF) analyses for rainfall of Ludhiana

NAR model

The Nonlinear Autoregressive (NAR) model is a type of time series model used to predict future values based on solely past observations of the same variable, incorporating nonlinear relationships. The fundamental structure of the NAR model is expressed as:

$$yt = f(yt-1, yt-2, \ldots, yt-p) + \varepsilon t$$

where y_t represents the predicted value at time t, f is a nonlinear function determined by the neural network, p is the number of past observations (lag), and is the random error term. The function f is learned through the training process of the ANN, where weights and biases are adjusted to minimize prediction errors. The model's effectiveness was evaluated using the fundamental performance metrics: the coefficient of determination (R^2) and the mean absolute error (MAE).

Autocorrelation function (ACF) and partial autocorrelation function (PACF)

The time series analysis of Ludhiana's monthly rainfall data reveals dominant seasonal structure, as evidenced by the autocorrelation function (ACF) with partial autocorrelation function (PACF) plots (Fig. 1). The ACF plot, measures the correlation of the time series with its own lagged values, which supports the presence of strong seasonality. Noticeable peaks at lags 12, 24, and 36 months suggest that the rainfall in a given month is significantly correlated with rainfall in the same month of previous years. This repeating pattern of correlations at multiples of 12 months reinforces the conclusion that the data is strongly seasonal. The PACF plot, shows direct influence of a particular lag on the current value after removing indirect effects from shorter lags, reveals significant spikes at lag 1 and lag 12, among others. The spike at lag 1 indicates a short-term autoregressive (AR) component, meaning that rainfall in one month is partially influenced by rainfall in the previous month. The strong spike at lag 12 again confirms a seasonal autoregressive relationship, where the rainfall in a given month is directly influenced by the same month in the previous year.

Rainfall modelling

Based on the analysis of the observed verses predicted plots (Fig. 2) the SARIMA model and the NAR model forecasts monthly rainfall showing distinctive differences. In the observed vs. predicted graph for the NAR model, there are visible discrepancies, particularly underestimating peak rainfall months, with predicted values often deviating by 30–60 mm from observed peaks during high rainfall years. In contrast, the SARIMA forecast closely follows the observed trends, with most deviations appearing to be within ±20 mm but in peaks the SARIMA cannot outperform to pick the pattern. The spread of NAR model suggests improved model accuracy and reduced bias in case of non-identical rainfall condition. Overall, while the NAR model has potential especially in

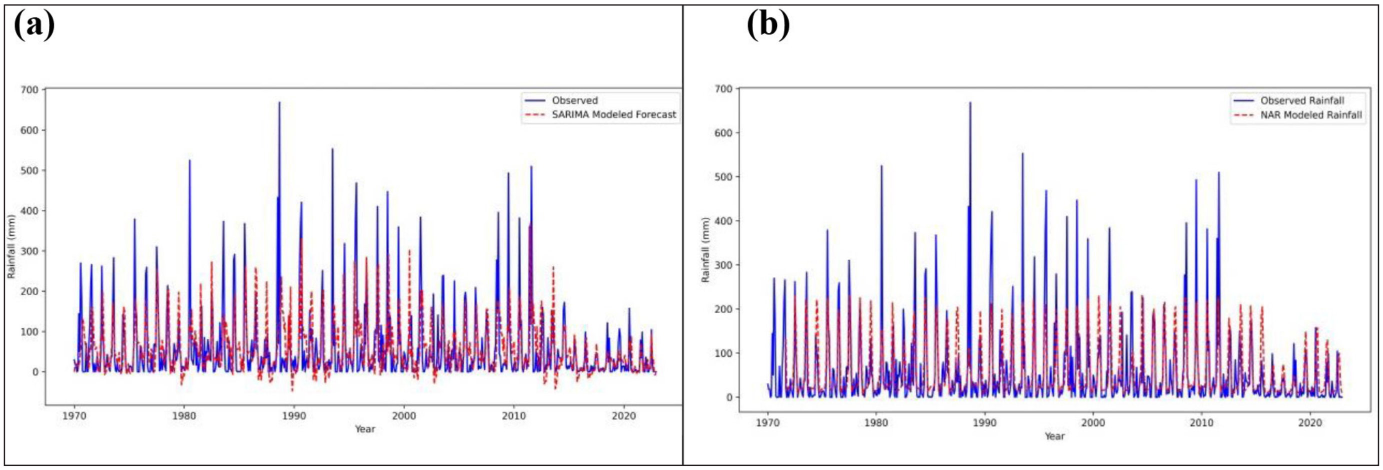

Fig. 2: Observed and predicted annual rainfall by (a) SARIMA and (b) NAR models for Ludhiana during 1970-2022

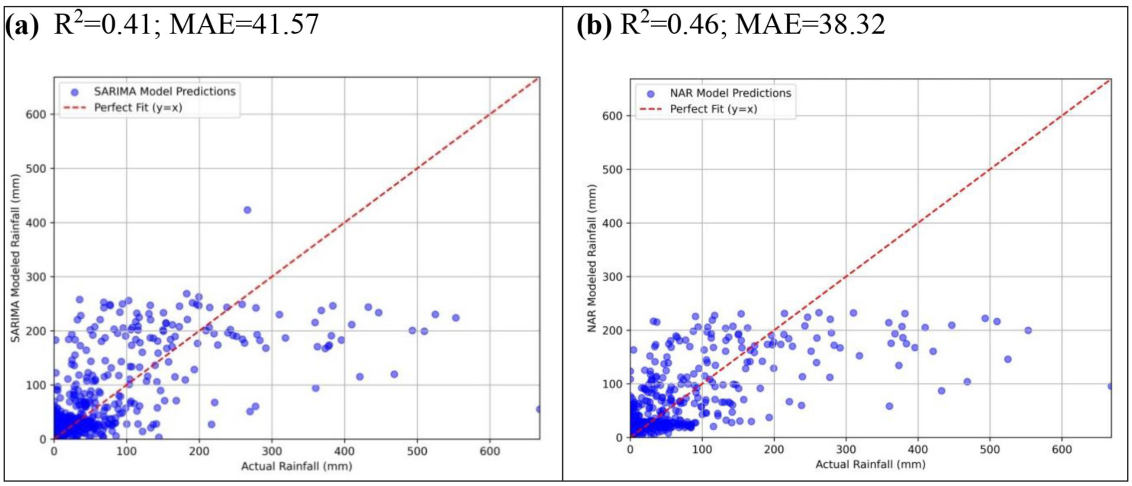

Fig. 3: Observed and predicted rainfall by (a) SARIMA and (b) NAR models

capturing nonlinear dependencies.

The validation of the SARIMA and NAR models was conducted by comparing the predicted rainfall values against the actual observed rainfall (Fig. 3). The scatter shows that while the NAR model generally captures the trend of rainfall, there is a significant dispersion of data points, at higher rainfall values. The model tends to predict extreme rainfall events well, as indicated by many points falling below the 1:1 line. Quantitatively, the correlation between actual and predicted values appears moderate, and the model shows a tendency to cluster predictions in the lower range (below 200 mm). In contrast, the SARIMA model validation plot displays clustering of points along the 1:1 line, suggesting better predictive accuracy apart from the general trends. The SARIMA model exhibits less dispersion and a higher degree of linearity in its predictions, indicating a stronger correlation with actual rainfall values. NAR model slightly outperforms SARIMA in terms of both R^2 (0.56, 0.41) and MAE (38.32, 41.57) but both the models demonstrate limitations in capturing high-magnitude rainfall events. The NAR model's strength lies in its non-linear learning ability, which provides better adaptability to complex rainfall dynamics. Visual analysis of residuals and scatter plots reveals both models consistently under predict extreme rainfall events (>250 mm), limiting their applicability in peak monsoon months.

ACKNOWLEDGEMENT

The authors are thankful to Department of Climate Change and Agriculture Meteorology, PAU, Ludhiana for providing the long-term rainfall data.

***Funding*: No funding was received for this work**

Conflict of interest: The authors declare that there is no conflict of interest related to this article

Data availability: Data can be provided on request

Authors contribution: **S. BHATTACHARJEE:** Writing original draft, Visualization, Validation, Methodology, Investigation, Formal analysis; **N. BIWALKAR:** Conceptualization, reviewing original draft, Supervision; **K. SUR:** Conceptualization, reviewing original draft, Visualization, Methodology, Supervision; **S. KAUR:** Reviewing original draft, Supervision; **S. P. SINGH:** Data, Reviewing original draft, Supervision.

Disclaimer: The contents, opinions and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Publisher's Note: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Amshi, A. H. and Prasad, R. (2023). Time series analysis and forecasting of cholera disease using discrete wavelet transform and seasonal autoregressive integrated moving average model. *Sci. African,* 20: e01652.

Baig, F., Ali, L., Faiz, M. A., Chen, H. and Sherif, M. (2024). How accurate are the machine learning models in improving monthly rainfall prediction in hyper arid environment? *J. Hydrol.,* 633: 131040.

Kothiyal, S., Goel, S., Singh, R. K., Chilwal, A. and Ranjan, R. (2025). SARIMA-based time series analysis of rainfall and temperature for the Tarai region of Uttarakhand. *J. Agrometeorol.,* 27(1): 73-76. https://doi.org/10.54386/jam.v27i1.2838

Priyan, K. (2021). Issues and challenges of groundwater and surface water management in semi-arid regions. Groundwater resources development and planning in the semi-arid region, 1-17.

Sarkar, P., Basak, P., Panda, C. S., Gupta, D. S., Ray, M. and Mitra, S. (2023). Prediction of major pest incidence in Jute crop based on weather variables using statistical and machine learning models: A case study from West Bengal. *J. Agrometeorol.,* 25(2): 305-311. https://doi.org/10.54386/jam.v25i2.1915