Scientiarum: A Multidisciplinary Journal
Volume 1, Issue 3, June 2025, pp. 24–29
DOI: 10.54646/SAPARS.2025.15

SAPAR SERVICES LLP
Scholarly Articles Publication And Repository Services

RESEARCH

Predicting rainfall patterns in Palakkad district of South India based on time series forecasting approaches

C. Sreerag*

Department of Statistics, Government Victoria College, Palakkad, India

***Correspondence:**
C. Sreerag,
sreeragc737@gmail.com

Received: 04 June 2025; **Accepted:** 14 June 2025; **Published:** 27 June 2025

Rainfall prediction is inevitable for managing agricultural activities, water resources, and mitigating the risks of droughts and floods, particularly in regions like the Palakkad district, Kerala, which is known for its agricultural significance. Time series forecasting plays a vital role in identifying patterns, trends, and seasonal variations in rainfall data. In this study the Holt-Winters exponential smoothing model and the seasonal autoregressive integrated moving average (SARIMA) model are considered for predicting the rainfall in Palakkad district, Kerala. The forecast accuracy metrics root mean squared error (RMSE) and mean absolute percent error (MAPE) are utilised to evaluate the performance of the above models. While the Holt-Winters exponential smoothing model effectively captured simple seasonal patterns, the SARIMA model excelled in handling complex seasonal structures and trends. The SARIMA model was identified as the most accurate, with minimal forecast errors, adherence to assumptions, and low residual correlations.

Keywords: time series forecasting, rainfall prediction, seasonal autoregressive integrated moving average (SARIMA), Holt-Winters exponential smoothing model, root mean squared error (RMSE), mean absolute percent error (MAPE)

Introduction

Effective management of water resources, agriculture, and disaster preparedness all depend on the ability to predict rainfall patterns, especially in areas with highly seasonal climates like South India's Palakkad district. Different monsoon-driven rainfall cycles are experienced in Palakkad, with the Southwest and Northeast monsoon seasons seeing the most precipitation. Precise prediction of these trends can enhance agricultural planning, promote sustainable development, and lessen the negative effects of droughts and floods. With the goal of offering trustworthy insights that might support local planning and decision making, this study investigates time series forecasting techniques to examine past rainfall data and project future trends in Palakkad.

Rainfall is a crucial component of the hydrological cycle and has a significant influence on agriculture, water resource management, and disaster preparedness, particularly in countries like India that rely heavily on the monsoon. Predicting rainfall accurately is essential to reducing the impact of extreme weather events like droughts and floods, which nevertheless affect several states in the country. Researchers have emphasised the need for reliable prediction models to improve agricultural planning, resource management, and disaster response (1). To provide more accurate forecasts, these models can make use of cutting-edge methods like machine learning and satellite data. Policymakers and farmers alike can better prepare for the difficulties presented by erratic weather patterns by incorporating these cutting-edge strategies. Implementing more robust farming techniques, improving irrigation systems, and creating early warning systems for severe weather are a few examples of these measures. In the end, communities can lessen the negative consequences of climate change and guarantee food security for future generations by utilising technology and data. Because of geographic differences, regional projections are often inaccurate at the local level. Rainfall forecasting has long been an important area of research due to its importance in agricultural planning, water resource management, and disaster relief. Accurately forecasting rainfall patterns is crucial for preserving crop yields and reducing the risks of drought and flooding in monsoon-dependent regions such as South India and particularly in Kerala's Palakkad district, as well as in regions like the Western Ghats. Mass and Kuo (2) emphasises the importance of regional weather forecasting systems. Numerous studies have emphasised the need for precise and localised forecasting models due to the geographical variability of rainfall, especially in regions affected by complex geography such as the Western Ghats. Mass and Kuo (2) emphasises the importance of regional weather forecasting systems. Numerical Weather Prediction (NWP) models use current atmospheric data to predict future weather conditions, albeit their effectiveness varies based on the topography (3). Although rainfall is necessary for ecosystems and human activity, it can also result in floods and landslides (4). Climate change further complicates rainfall patterns; research indicates that human activity and global warming have led to more variability in rainfall frequency, intensity, and dispersion (5). Analysing these trends is essential for sustainable agriculture practices, urban planning, and water management strategies.

One of the most used methods for modelling rainfall data is time series forecasting. Numerous studies have effectively used conventional statistical techniques for short-term rainfall forecasting, such as Autoregressive Integrated Moving Average (ARIMA) (6). Research by Kumar and Jain (7) indicated that wavelet-ARIMA hybrid models were more accurate than traditional models when used to estimate monthly rainfall in India. Because deep learning methods like Long short-term memory (LSTM) can learn from temporal dependencies in huge datasets, they have lately been investigated for rainfall forecasting (8). The use of sophisticated time series models is essential for localised prediction in Kerala since studies have demonstrated that rainfall patterns are impacted by both local topography and wider climate variability (9). Rainfall predictability is made much more difficult by the effects of climate change, which add unpredictability to the monsoon's onset, duration, and intensity. Because of this, trend analysis and anomaly detection are becoming essential components of contemporary rainfall forecasting research. Furthermore, the spatial resolution and dependability of rainfall prediction models have been enhanced by the combination of satellite data, remote sensing, and reanalysis datasets (10). There is still a lack of region-specific research despite these developments, especially for areas like Palakkad, where extremely varied rainfall patterns are caused by orographic effects and monsoon dualism (Southwest and Northeast). Therefore, local farmers, water management authorities, and policymakers might benefit greatly from the application and evaluation of various time series forecasting models tailored for this region. Global climate change may affect long-term rainfall patterns, which could affect water supply and increase the probability of extreme weather events like droughts and floods. Since they have a direct impact on agricultural productivity and determine the environmental conditions of a place, temperature and rainfall are the two most significant climatic elements (11, 12). Agriculture, food security, and energy security are among the businesses that depend on a favourable climate and a timely and enough supply of water. The quantity of water available for household consumption, industry, agriculture, and the generation of hydroelectric power is greatly influenced by rainfall. The amount and distribution of rainfall largely determine agricultural output, which is crucial to India's economy and people's quality of life (11, 13). Numerous studies have analysed climate trends using various models and approaches. Kumar and Jain (14) studied long-term rainfall trends in Kashmir. Saha et al. (15) used seasonal autoregressive integrated moving average (SARIMA) models to analyse monthly temperature data in Giridih, India, and Sethi and Garg (16) studied data mining approaches for rainfall prediction. Sathish et al. (17) forecasted rainfall in West Bengal using SARIMA.

Theory and methods

Data collection is the process of methodically obtaining and assessing information on pertinent factors in order to answer research questions, test hypotheses, and assess results. With an emphasis on evaluating stationarity, spotting outliers, and ensuring a Gaussian distribution, the study used time plots to analyse the overall behaviour and patterns of the rainfall time series. Correlograms were used to test for stationarity. Monthly rainfall data from the Palakkad district from 2000 to 2023, including 23 years of observations, were used to create rainfall forecasting models for 1 and 2 months in advance. The Indian Meteorological Department in Pune provided this dataset; the location was chosen because of its extensive and reliable meteorological data record. Following analysis, some of the data was set aside for testing, while the rest was used to train artificial neural network (ANN) models.

The study's objective was to foresee and comprehend seasonal trends by analysing rainfall data using sophisticated statistical and machine learning approaches. In order to capture seasonal trends, the methodology involved gathering monthly rainfall data from a trustworthy source, preprocessing it, and then converting it into a time series format with a frequency of 12. In order to find notable lags and seasonal effects, exploratory data analysis entailed visualising the data using time series plots and analysing autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. Seasonal ARIMA models, which featured seasonal and non-seasonal differencing, were used for model building, as well as additive and multiplicative versions of the Holt-Winters approach. The Bayesian information criterion (BIC) and

Akaike information criterion (AIC) were used to optimise the model's parameters. Performance measures, including mean error (ME), mean percent error (MPE), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percent error (MAPE), were used for validation. Additionally, residual diagnostics such as the Ljung-Box test to check for autocorrelation were used. Lastly, future rainfall values were predicted using the models, and the uncertainty in the projections was measured by computing confidence intervals.

Seasonal ARIMA modelling

The identification of pertinent models and the incorporation of suitable seasonal elements are essential for seasonal modelling and its applications. The Integrated Moving Average with Seasonal Autoregression. If $Z_t + s = Z_t$ for $t = 0, \pm 1, \pm 2, \ldots$, then a time series Z_t is considered seasonal with periods.

The basic tenet of seasonal time series with period "s" is that observations spaced "s" apart are comparable. The period is $s = 12$ if the series is influenced by monthly seasonal factors.

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t \tag{1}$$

For a seasonal time series, this is the seasonal model of order $(p, d, q) \times (P, D, Q)$s.

where,

$$\begin{aligned}
\phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p \\
\Phi(B) &= 1 - \Phi_1 B - \cdots - \Phi_P B^P \\
\theta(B) &= 1 + \theta_1 B + \cdots + \theta_q B^q \\
\Theta(B) &= 1 + \Theta_1 B + \cdots + \Theta_Q B^Q
\end{aligned}$$

$\{a_t\}$ is a white noise process.

SARIMA models are the most general and accurate forecasting models. Time series models were created using SARIMA to estimate monthly parameters for the Palakkad District, taking into account the significance and inevitable role of seasons in Kerala. There have been no prior attempts to use the SARIMA model to evaluate the district's temperature and rainfall trends. Therefore, the current paper attempts to use the SARIMA model to examine the rainfall patterns for Kerala's Palakkad District.

Holt-Winters exponential smoothing

Holt-Winters Exponential smoothing is a statistical forecasting technique that includes seasonality and trend support in time series data. It is frequently employed for time series data forecasting, particularly in cases where a regular seasonal trend is present.

A time series forecasting technique called the Holt-Winters Additive Model incorporates trend and seasonality to go beyond exponential smoothing. It works well with data that has additive seasonal effects and a linear trend, which means that seasonal variations don't change with time. The three primary parts of the model are seasonality (recurring patterns), trend (pace of change over time), and level (total value of the series). By extrapolating the found patterns, it estimates future values and iteratively updates these components using smoothing equations.

The Additive Holt-Winters model consists of three components:

Level (L_t): The smoothed value of the series at time t.

Trend (T_t): The change in the level over time.

Seasonality (S_t): The seasonal effect at time t, capturing the recurring patterns over a fixed period.

$$L_t = \alpha(Y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{2}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{3}$$

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-m} \tag{4}$$

where α, β, and γ are the smoothing parameters.

To forecast the equation of k periods ahead

$$\hat{Y}_{t+k} = L_t + kT_t + S_{t-m+k} \tag{5}$$

where,

L_t, T_t, and $S_{t-\ m+\ k}$ are the forecasted values for k periods ahead.

L_t, T_t, and $S_{t\ -\ m\ +\ k}$ are the latest values of level, trend, and seasonality.

For starting values, it seems sensible to set the level component L_0, equal to the average observation in the first year, i.e.,

$$L_0 = \sum_{t=1}^{s} \frac{y_t}{s} \tag{6}$$

where "s" represents how many seasons there are. The average difference between the first and second-year averages per time period can be used as the initial value for the slope component.

That is:

$$T_0 = \frac{\sum_{t=s+1}^{2s} \frac{y_t}{s} - \sum_{t=1}^{s} \frac{y_t}{s}}{s} \tag{7}$$

Finally, the seasonal index starting value can be calculated after allowing for a trend adjustment, as follows:

$$S_0 = y_k - [L_0 + (k-1)T_0/2] \tag{8}$$

where k is equal to 1, 2,..., s. Naturally, this will result in distinct values for S_0, which is necessary to obtain the original seasonal pattern.

The range of 0.02–0.2 is frequently used for choosing smoothing locations. Once more, they can be estimated by minimising the sum of the squared one-step-ahead mistakes; however, there isn't a single set of α, β, and γ that will minimise the square errors for every t.

Measures of forecast accuracy

The discrepancy between a time series' actual value and its expected or forecast value is known as a forecast error in statistics. Consider a situation where we have a system for predicting a variable's value at time t and some actual values for the same variable. Let $\hat{y}_t$ represent the forecast at time t and y_t represent the actual value at time t. Forecasts will always differ from the actual value; they are never totally correct. The goal of forecasting is to make it as small as feasible. There are numerous ways to gauge forecast accuracy, but the most widely used ones are

1. RMSE

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{9}$$

2. MAPE

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{10}$$

where:

1. y_t is the actual value at time t
2. $\hat{y}_t$ is the forecasted value at time t
3. n is the total number of observations

Results and discussion

In Kerala, India, the district of Palakkad has a unique seasonal rainfall pattern that is influenced by monsoon winds. The primary rainy season, the Southwest monsoon (June–September), brings 2000–2500 mm of rain, which is particularly heavy close to the Western Ghats because of the orographic impact. Although less powerful, the 500–700 mm of rainfall that the Northeast monsoon (October–December) brings is essential for agriculture. The area experiences a dry season from January to May with little rainfall—typically less than 100 mm per month—which results in a shortage of water and a greater reliance on irrigation. The local climate, agriculture, and water supplies all depend on these seasonal rains.

Time series plots

Time series analysis is used to fit a model for the monthly rainfall in Palakkad district from 2000 to 2023.

Strong seasonality is indicated by the time series plot of rainfall in **Figure 1**, which displays recurrent peaks. Extreme spikes that occur occasionally indicate infrequent yet heavy rainfall occurrences.

SARIMA models

The identified model for rainfall is SARIMA (0,0,1)(1,1,2)[12]. The model equation is

$$\Phi(B)y_t = \Theta(B)\varepsilon_t \tag{11}$$

where:

$$y_t = y_{t-12} + 0.8247y_{t-12} - 0.8247y_{t-24} + \\ \varepsilon_t + 0.2112\varepsilon_{t-1} - 0.0103\varepsilon_{t-12} - 0.6583\varepsilon_{t-24}$$

Plot of forecast

From **Figure 2**, we have used the fitted model for forecasting the rainfall for upcoming months. There is a growing seasonality in the data. That is, the district of Palakkad will have more rainfall in the months to come.

Holt-Winters exponential smoothing model

A model for the monthly rainfall in the Palakkad district from 2000 to 2023 is fitted using time series analysis.

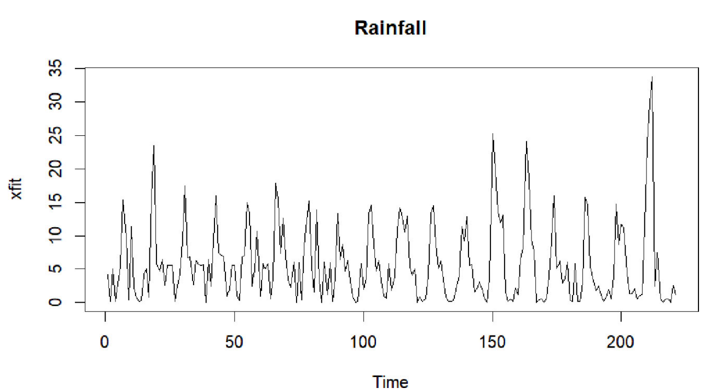

FIGURE 1 | Time series plot of rainfall during the years 2000–2023.

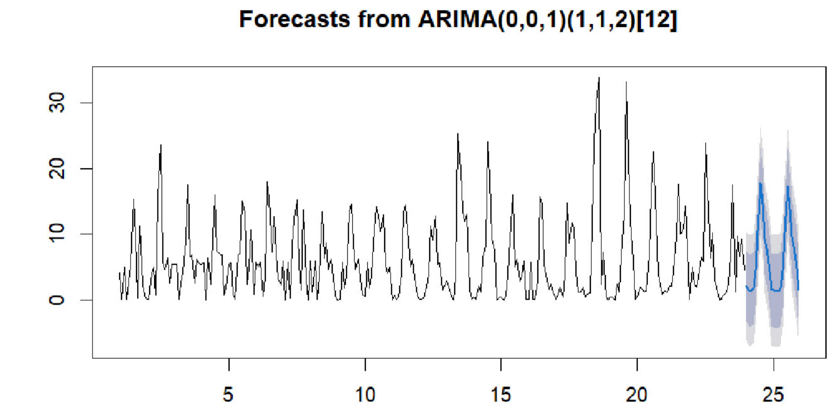

FIGURE 2 | Plot of forecast in rainfall.

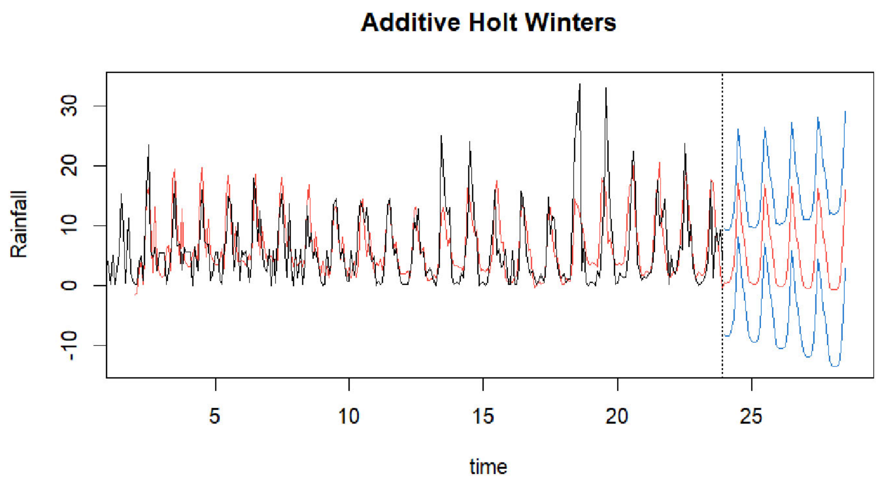

FIGURE 3 | Plot of forecast Holt-Winters using rainfall.

A time series forecasting technique called the Holt-Winters Additive Model expands on Holt's linear trend model to take seasonality into consideration. It is especially helpful for data that exhibits additive seasonality, or a consistent seasonal trend across time.

Plot of forecast Holt-Winters using rainfall

From **Figure 3**, we see that the Additive Holt-Winters model effectively captures the seasonal rainfall pattern, with forecasted values showing periodicity and increasing uncertainty over time.

Comparison of SARIMA and Holt-Winters models

According to MAPE and RMSE for rainfall, the SARIMA model outperforms the Holt-Winters model in general. SARIMA's lower RMSE and MAPE values suggest higher accuracy and better prediction performance.

TABLE 1 | Accuracy comparison of the SARIMA model and the Holt-Winters model

	RMSE	MAPE
SARIMA model	0.40718	2.95645
Holt-Winters model	1.0163	4.095531

Comparison of the SARIMA model and the Holt-Winters model for rainfall

From **Table 1**, the SARIMA model is the best option when comparing two models since it performs better predictively and has lower RMSE and MAPE values.

Conclusion

Using the SARIMA and Holt-Winters models, an effort has been made to create new statistical tools for weather forecasting in the Palakkad District. The SARIMA model fits rainfall better and has lower RMSE and MAPE values than the Holt-Winters model. The SARIMA model demonstrated higher accuracy for a number of meteorological indicators due to its lower error metrics. The fitted model has been used to forecast all of the rainfall for the upcoming months. According to the statistics, which indicate an increasing tendency, these indicators are anticipated to rise in Palakkad District in the months ahead.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Author contributions

SC: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Code availability

Code is available upon a reasonable request.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Hewage P, Behera A, Trovati M, Pereira E, Ghahremani M, Palmieri F, et al. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Comput.* (2020) 24:16453–82.
2. Mass CF, Kuo YH. Regional real-time numerical weather prediction: current status and future potential. *Bull Am Meteorol Soc.* (1998) 79: 253–64.
3. Benjamin SG, Brown JM, Brunet G, Lynch P, Saito K, Schlatter TW. 100 years of progress in forecasting and NWP applications. *Meteorol Monogr.* (2019) 59:13.1–67.
4. Ratnayake U, Herath S. Changing rainfall and its impact on landslides in Sri Lanka. *J Mt Sci.* (2005) 2:218–24.
5. Obot N, Chendo M, Udo S, Ewona PO. Evaluation of rainfall trends in Nigeria for 30 years (1978-2007). *Int J Phys Sci.* (2010) 5(14): 2217–22.
6. Box G, Jenkins G. *Analysis: Forecasting and Control*. San Francisco: Holden-Day Publisher (1976).
7. Kumar U, Jain VK. Time series models (grey-Markov, grey model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy.* (2010) 35:1709–16.
8. Sharma A, Kumar V, Shahzad B, Tanveer M, Sidhu GPS, Handa N, et al. Worldwide pesticide usage and its impacts on ecosystem. *SN Appl Sci.* (2019) 1:1–16.
9. Krishnakumar K, Rao GP, Gopakumar C. Rainfall trends in twentieth century over Kerala, India. *Atm Environ.* (2009) 43:1940–4.
10. Prakash A, Hasan SA, Lee K, Datla V, Qadir A, Liu J, et al. *Neural Paraphrase Generation with Stacked Residual LSTM Networks*. (2016). arXiv preprint arXiv:1610.03098.
11. Kumar R, Gautam HR. Climate change and its impact on agricultural productivity in India. *J Climatol Weather Forecast.* (2014) 2:1–3.
12. Modarres R, da Silva V.d.P.R. Rainfall trends in arid and semi-arid regions of Iran. *J Arid Environ.* (2007) 70:344–55.
13. Gajbhiye S, Meshram C, Singh SK, Srivastava PK, Islam T. Precipitation trend analysis of Sindh river basin, India, from 102-year record (1901–2002). *Atm Sci Lett.* (2016) 17:71–7.
14. Kumar V, Jain SK. Trends in seasonal and annual rainfall and rainy days in Kashmir valley in the last century. *Quat Int.* (2010) 212:64–9.
15. Saha E, Hazra A, Banik P. Sarima modeling of the monthly average maximum and minimum temperatures in the eastern plateau region of India. *Mausam.* (2016) 67:841–8.
16. Sethi N, Garg K. Exploiting data mining technique for rainfall prediction. *Int J Comput Sci Inform Technol.* (2014) 5:3982–4.
17. Sathish G, Narasinhaiah L, Babu PM, Laha S, Kumar NB. Time series analysis of monthly rainfall for gangetic west Bengal using box Jenkins Sarima modeling. *Int J Curr Microbiol App Sci.* (2017) 6:2603–10.