Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 2543-2548 | ISSN:2251-6727

Modelling and Prediction of Annual Rainfall Using ARIMA

Abhinaya A & Kuleena Das

Department of Mathematics, Sree Narayana College, Kollam

Corresponding author- Kuleena Das

(Received: 16 January 2026 Revised: 25 February 2026 Accepted: 30 March 2026)

KEYWORDS

Time series analysis;

Trend;

Stationary data;

ARIMA;

Rainfall Prediction

ABSTRACT: Rainfall is a key climatic parameter that regulates hydrological cycles and ecosystem balance. Analyzing annual rainfall variability aids in understanding long-term climatic trends and supports effective resource and risk management. This study employs the Autoregressive Integrated Moving Average (ARIMA) model to forecast annual rainfall in Kollam district, Kerala. Historical rainfall data were subjected to time series analysis to examine trends, patterns, and stationarity. The series was found to be non-stationary and was transformed through differencing to achieve stationarity. The optimal ARIMA model was selected based on the Bayesian Information Criterion (BIC) and validated using statistical performance indicators. The model achieved a predictive accuracy of 83.89% for the year 2024. Forecasts for the period 2024–2040 indicate realistic rainfall variations, demonstrating the potential of ARIMA modeling for reliable rainfall prediction and climate-related planning in monsoon-prone regions such as Kerala.

Introduction

Time series forecasting is about using past data to make predictions about the future. It's useful when we don't fully understand how and what are the factors that affect the data or when there is no substantial link between the data and the factors that influence it. By observing patterns over time, these models can reveal significant insights. Over the years, researchers have worked hard to refine these methods, making them more reliable and useful for various applications.

ARIMA is one of the most widely used methods in time series forecasting because of its accuracy and strong statistical base. It has proven effective across many fields like economics, hydrology, and environmental science that includes modelling and forecasting of carbon-dioxide emissions in South Africa [1], air quality of Delhi [2], unemployment rates in Nigeria [3].

This model relies solely on past values and errors which makes it especially useful in situations where external explanatory variables are unavailable or unreliable.

Rainfall Prediction is a difficult task in meteorology, which attempts to predict rain. Its importance is central to agriculture, water management, disaster prevention and preparedness, as well as social and economic development. Due to increasing climate variability and extreme weather, the demand for reliable precipitation prediction is increasing. Traditional meteorological methods are generally inadequate to predict the irregular and nonlinear character of precipitation, particularly in areas with highly variable spatial and temporal distribution. The ability to predict when and how much it will rain has always been one of the trickiest challenges in weather forecasting but few things are more important for our day-to-day lives and long-term planning. A misleading forecast is a form of real economic loss, wrecked lives and on some occasions even death. The high risks involved clearly marks the necessity for better and more accurate rainfall prediction methods.

In the context of rainfall forecasting, ARIMA has been applied successfully to model seasonal and inter-annual variability, offering valuable insights for water resource planning and agricultural decision-making. ARIMA was effectively used by Chandan Kumar Pandit, Anamol Kumar Lal (2023) to forecast rainfall in Ranchi, showing 87.35% accuracy and revealing key seasonal trends useful for water and climate planning [4].

The main goal of this paper was to build and evaluate a rainfall forecasting model using ARIMA, which to an

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 2543-2548 | ISSN:2251-6727

extent can accurately reflect long-term seasonal patterns in rainfall data.

The ARIMA Model

The ARIMA (Autoregressive Integrated Moving Average) model is a popular method for time-series forecasting in statistics. It has three components :

- Autoregressive (AR) indicates that the variable of interest is regressed on its own past values.
- I (Integrated) indicates that the data values have been replaced with the difference between each value and the previous value.
- The Moving Average, or MA, indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past [5].

The ARIMA model is given by ARIMA (p, d, q), where p,d,q are non-negative integers: p denotes order of the Autoregressive model, d denotes degree of differencing and q denotes order of moving average model [6]. In order to build an ARIMA model, we use test like Dickey-Fuller to check whether our time series is stationary [4]. If the series exhibits seasonality, we use differencing to make it stationary which helps in estimating d value. After applying differencing, we run ACF(Autocorrelation Function) and PACF(Partial Autocorrelation Function) plots so as to find the optimal lag values for our autoregressive(p) and moving average(q) terms. We choose the best model with minimal BIC and fit an ARIMA(p,d,q) model on the data.

Study Area

Kollam District, known earlier as Quilon, is one of the 14 districts in the state of Kerala, India. The district has a long coastline and an important Laccadive Sea seaport and an inland lake. Kollam has a tropical monsoon climate which varies very little from average temperatures. There is a dry season from December to March with less than 60 mm of rainfall per month. The wet season is from April to November, somewhat more than December to March, especially in June and July when the South West Monsoon blows.

Data and Methodology

The monthly rainfall data for Kollam district, spanning the period 1977–2023, were obtained from the India Meteorological Department (IMD). The Autoregressive Integrated Moving Average (ARIMA) model was employed to forecast monthly rainfall patterns. Prior to model development, the Augmented Dickey–Fuller (ADF) unit root test was applied to assess the stationarity of the time series, as stationarity is a prerequisite for ARIMA modeling [7]. Non-stationary series were appropriately differenced to achieve stationarity. Multiple ARIMA model configurations were evaluated, and the optimal model was selected based on the lowest Bayesian Information Criterion (BIC) value, ensuring the best balance between model fit and parsimony.

Plot

The original data has been plotted using the R programming language.

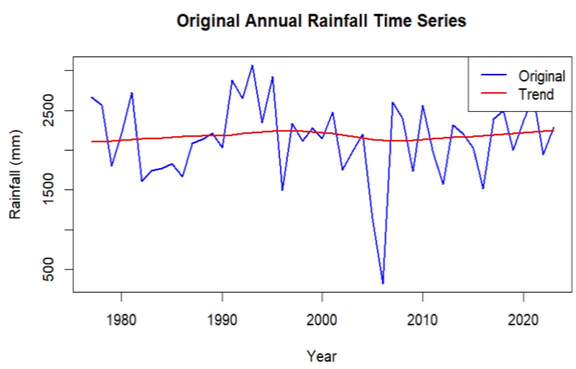

Figure 1

From (Figure 1), we observe that the annual rainfall ranges from a minimum of 324.7 mm in 2006 to a maximum of 3065.7 mm in 1993, with a median of 2201.0 mm and a mean of 2133.8 mm. The lower quartile (25%) is 1817.6 mm, and the upper quartile (75%) is 2438.4 mm this indicates a moderate variability around the central values. The overall trend, based on a LOWESS[1] smoother, shows a slightly increasing pattern over the observed period. The standard deviation of annual rainfall is 494.36 mm which further shows a considerable variability. In the last 10 years (2014–2023), annual rainfall has

[1]LOWESS (Locally Weighted Scatterplot Smoothing) is a method used to reveal the trend in a scatterplot.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 2543-2548 | ISSN:2251-6727

fluctuated between 1510.7 mm and 2674.0 mm, with most years remaining above 1900 mm, suggesting relatively consistent but variable rainfall in the recent decade.

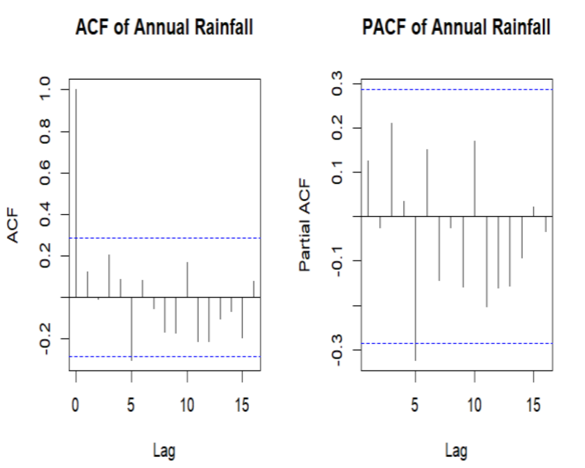

Figure 2

The autocorrelation (ACF) and partial autocorrelation (PACF) plots of the annual rainfall series in (Figure 2), provide insights into the temporal dependence structure of the data and help guide the selection of appropriate time series models.

Stationarizing Data

The data needs to be stationary to result in a reliable prediction [8]. The Augmented Dickey-Fuller (ADF) test was applied to check the stationarity of the original time series. It resulted a test statistic of −2.3632 with a p-value of 0.4293 and a lag order of 3, pointing non-stationarity. After applying first-order differencing, the ADF test on the differenced series yielded a test statistic of –3.5864 with a p-value of 0.0446 and the same lag order of 3, confirming that the series became stationary at the 5% significance level. The following plots represent the first-order differenced annual rainfall data: (i) the differenced time series showing fluctuations around a mean of approximately zero, (ii) the ACF plot indicating significant autocorrelation at certain lags, and (iii) the PACF plot highlighting significant partial autocorrelations, which help to identify ARIMA model .

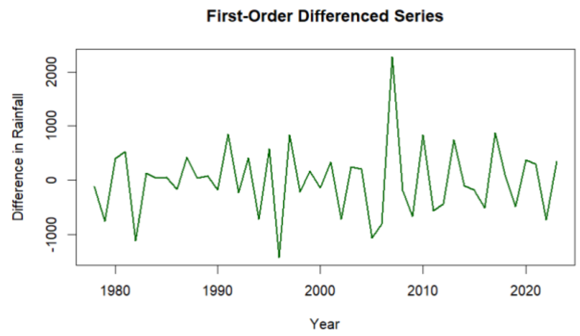

Figure 3

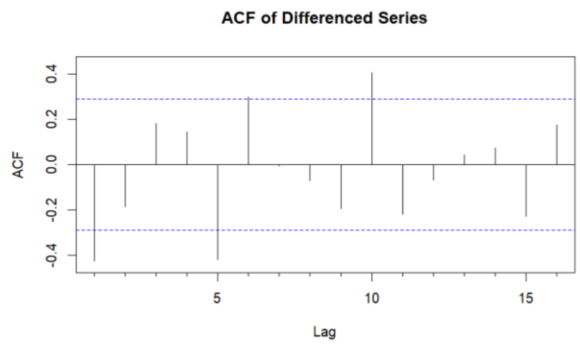

Figure 4

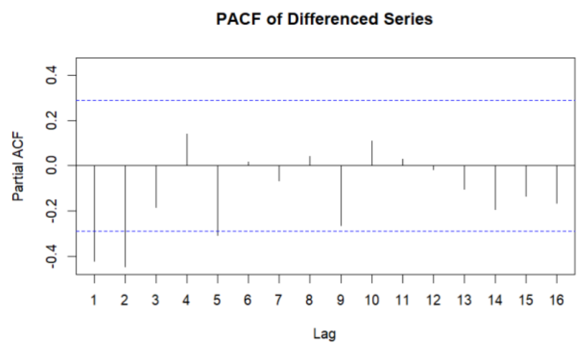

Figure 5

The first-order differenced (Figure 3) time series appears to fluctuate randomly around zero, suggesting that the non-stationarity in the original data has been cleared. The absence of any visible trend or seasonal component confirms that the differenced series is stationary. The ACF plot (Figure 4) of differenced series shows significant autocorrelation at lag(s): 1, 5, 6, 10. The PACF plot (Figure 5) of differenced series shows significant partial autocorrelation at lag(s): 1, 2, 5.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 2543-2548 | ISSN:2251-6727

Best-Fit ARIMA Model Selection

The Bayesian Information Criterion (BIC) was used to compare a range of ARIMA models applied to the first-order differenced rainfall series to find the most suitable model for the data, . The model with the lowest BIC value was chosen. The ARIMA (2,1,0) was chosen the best ARIMA model with BIC value 718.2886.

Trend Detection

The Mann-Kendall test was applied to the annual rainfall data to analyze the trend and LOWESS smoothing. This test is commonly used in to detect trends in time series data, especially when the data may not follow a normal distribution [4]. It helps to determine whether there is a consistent upward or downward trend over time.

In the Mann-Kendall test, the sign of Kendall's tau (τ) indicates the direction of the trend:

• Positive tau ($\tau > 0$): suggests an upward (increasing) trend.

• Negative tau ($\tau < 0$): suggests a downward (decreasing) trend.

• Tau near zero ($\tau \approx 0$): suggests no clear trend or a very weak trend.

In this analysis, the test returned a Kendall's tau of approximately 0.0009 with a p-value of 1, suggesting that there is no significant monotonic trend in the rainfall data over the observed period.

Results and Discussion

The ARIMA(2,1,0) model was employed to forecast the monthly rainfall for the year 2024 in Kollam district. The observed and predicted monthly rainfall values are presented in Table 1. The cumulative forecasted annual rainfall for 2024 was 794.45 mm, compared to the observed annual rainfall of 946.8 mm. The model achieved an overall forecasting accuracy of 83.89%, indicating a satisfactory level of predictive performance. Minor deviations were noted during the pre-monsoon and southwest monsoon months, particularly in May and June, reflecting the inherent variability and nonlinearity in rainfall patterns characteristic of the region. Overall, the ARIMA(2,1,0) model effectively captured the general rainfall trend, demonstrating its suitability for short-term rainfall prediction in a monsoon-dominated climatic setting such as that of Kerala.

Table 1: Monthly Actual, Forecasted, and Error in Rainfall for 2024

Month	Actual Rainfall(mm)	Forecasted rainfall(mm)	Error
January	10.2	53.86	-43.66
February	4.1	73.48	-69.38
March	35.1	67.08	-31.98
April	42.5	64.33	-21.83
May	457.6	68.39	389.21
June	335.7	66.71	268.99
July	10.2	66.37	-56.17
August	10.2	67.18	-56.98
September	10.2	66.77	-56.57
October	10.2	66.75	-56.55
November	10.2	66.91	-56.71
December	10.2	66.81	-56.61

The following plot (Figure 6) represents a comparison of actual and forecasted rainfall for 2024.

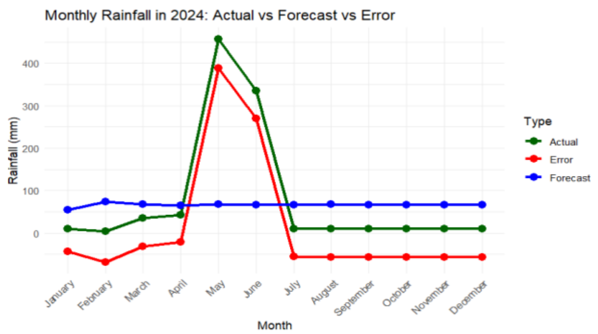

Figure 6

Rainfall in Kollam from 2025-2040 has been forecast using ARIMA (2,1,0) model. The result is tabulated and a plot on the same is shown in (Table 2) and Figure 7.

Table 2: Forecasted Annual Rainfall (2025–2040) using ARIMA (2,1,0)

Year	Forecasted rainfall(mm)
2025	2358.92
2026	2171.04
2027	2266.91
2028	2280.34
2029	2232.35
2030	2258.53
2031	2260.84
2032	2248.63
2033	2255.75
2034	2256.04
2035	2252.94
2036	2254.87
2037	2254.86
2038	2254.08
2039	2254.60
2040	2254.57

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 2543-2548 | ISSN:2251-6727

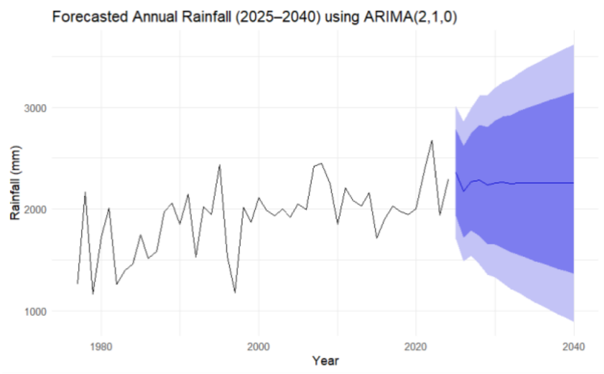

Figure 7

The forecasting performance of the ARIMA(2,1,0) model demonstrates that time series-based approaches can effectively capture the underlying temporal dynamics of rainfall in monsoon-affected regions such as Kollam, Kerala. Although the model achieved a commendable accuracy of 83.89% for 2024, the discrepancies observed during the months of May and June indicate the limitations of purely linear models in representing the complex, nonlinear nature of monsoonal rainfall influenced by regional and global climatic oscillations. Such deviations may also be attributed to short-term anomalies, including localized convection and changes in sea surface temperatures associated with ENSO events [9]. Despite these challenges, the model successfully reproduced the overall annual trend, suggesting its reliability for medium-term rainfall forecasting. Future research could integrate hybrid modeling approaches—combining ARIMA with machine learning or seasonal decomposition techniques—to enhance the predictive accuracy and account for nonlinearity and seasonality more effectively [10]. The findings underscore the potential of ARIMA-based forecasting as a valuable tool for regional climate assessment, water resource management, and disaster preparedness planning in Kerala.

The projected annual rainfall for the period 2025–2040, estimated using the ARIMA(2,1,0) model, exhibits moderate interannual variability. The highest forecasted annual rainfall occurs in 2025 (2358.92 mm), whereas the lowest is predicted for 2026 (2171.04 mm). Beyond 2026, the forecasts display a relatively stable pattern, with annual rainfall values fluctuating narrowly between 2250 mm and 2260 mm, indicating the absence of a pronounced increasing or decreasing trend. The mean forecasted rainfall for the period is 2253.54 mm, and the median value, 2254.48 mm, both corresponding to the year 2036, further emphasize the long-term stability projected by the model. However, the prediction intervals progressively widen with the increasing forecast horizon, particularly beyond 2030, reflecting the expected accumulation of uncertainty inherent in long-term time series forecasts. This divergence underscores the limitations of the ARIMA model in capturing complex climatic fluctuations over extended periods while still providing a reasonable estimate of the overall rainfall trend. The long-term rainfall projections derived from the ARIMA(2,1,0) model suggest a period of relative climatic stability in Kollam district during 2025–2040, with only moderate interannual fluctuations. The absence of a significant upward or downward trend in the forecasted series indicates that, based on historical patterns, the region is likely to maintain a stable rainfall regime under similar climatic conditions. However, the widening of prediction intervals beyond 2030 highlights an increase in model uncertainty, which is characteristic of statistical forecasting methods when extrapolated over extended horizons [10]. This uncertainty may stem from potential shifts in large-scale climatic drivers such as the Indian Ocean Dipole (IOD), El Niño–Southern Oscillation (ENSO), or changes in monsoon dynamics, none of which are explicitly captured by univariate ARIMA models [12]. Despite this limitation, the model provides a useful baseline for anticipating general rainfall behavior and supporting long-term water resource and agricultural planning. Incorporating additional climatic predictors or adopting hybrid approaches that combine ARIMA with machine learning or global circulation model (GCM)-based inputs could further enhance predictive reliability and adaptability to future climatic variability [13].

Conclusion

This study demonstrates the applicability of the ARIMA(2,1,0) model for modeling and forecasting rainfall in Kollam district, Kerala. The model effectively captured the temporal dynamics of historical rainfall data and produced reliable short-term forecasts with an accuracy of 83.89% for the year 2024. The long-term projections for 2025–2040 indicated moderate interannual variability and a stable rainfall regime, suggesting the absence of a pronounced trend in the forthcoming years. Although forecast uncertainty increased with the prediction horizon, the results affirm

the model's potential for short to medium-term rainfall forecasting in monsoon-dominated regions. The study highlights the importance of statistical modeling as a practical tool for regional climate assessment, agricultural planning, and water resource management. Future research should focus on integrating ARIMA with hybrid or multivariate approaches that incorporate additional climatic and atmospheric parameters to improve predictive accuracy and account for nonlinear influences on rainfall variability.

Acknowledgement

Grateful acknowledgment is extended to the India Meteorological Department (IMD), Pune, for providing access to the historical rainfall data of Kollam district, which served as the primary dataset for this study.

References

[1] M. Kour, "Modelling and forecasting of carbon-dioxide emissions in South Africa by using ARIMA model," *International Journal of Environmental Science and Technology,* pp. 11267-11274, 2022.

[2] Jain, Gourav and Jusleen Kaur Rekhi and Preeti Nagrath and Rachna, "Forecasting Air Quality of Delhi Using ARIMA Model," 2019.

[3] M. O. Adenomon, "Modelling and Forecasting unemployment rates In Nigeria using Arima Model," 2017.

[4] Pandit, Chandan and Lal, Anamol, "Rainfall prediction using Arima Model and Trend analysis of future rainfall over Ranchi district, Jharkhand, India," vol. 14, pp. 13-21, 10 2023.

[5] Hyndman, R.J., and Athanasopoulos, G., Forecasting: Principles and Practice (2nd ed) OTexts, 2018.

[6] Shumway, R. H., and Stoffer, D. S., Time Series Analysis and Its Applications: With R Examples (4th ed.), Springer, 2017.

[7] Qasim T. B., Ali H., Malik, N., & Liaquat M, "Forecasting inflation applying ARIMA Model with GARCH innovation: The Case of Pakistan.," *Journal of Accounting and Finance in Emerging Economies,* vol. 7, pp. 313-324.

[8] C. Vijayalakshmi, K. Sangeeth, R. Josphineleela, R. Shalini, K. Sangeetha, and D. Jenifer, "Rainfall prediction using ARIMA and linear regression," in *in International Conference on Computer, Power and Communications (ICCPC)*, 2022.

[9] Medha Khole and Uday Shankar De, "Floods and droughts in association with cold and warm ENSO events and related circulation features," *MAUSAM,* 2021.

[10] Ali Abdulhafidh Ibrahim and Bilal Saeed and Marwa A. Fadil, "Forecasting stock prices with an integrated approach combining ARIMA and Machine Learning techniques ARIMAML," *Journal of Computer and Communication,* 2023.

[11] Jin Liu, "Navigating the Financial Landscape: The power and limitations of the ARIMA Model," *Highlights in Science, Engineering and Technology,* 2024.

[12] Identifying the influence of El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) Phenomena on rainfall in The Aceh Region, "Identifying the influence of El Nino Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) Phenomena on rainfall in The Aceh Region, Indonesia," *Journal of Geoscience, Engineering, Environment, and Technology,* 2024.

[13] Sadhan Malik and Subodh Chandra Pal and Ashim Sattar and Sudhir Kumar Singh and Biswajit Das and Rabin Chakrabortty and Pir Mohammad, "Trend of extreme rainfall events using suitable Global Circulation Model to combat the water logging condition in Kolkata Metropolitan Area," *Urban Climate,* 2020.

[14] "S. Swain, S. Nandi and P. Patel, "Development of an ARIMA Method for Monthly Rainfall Forecasting over".

[15] Angelina Geetha and G. M. Nasira, "Time-series modelling and forecasting: modelling of rainfall prediction using ARIMA model," *International Journal of Society Systems Science,* vol. 8, p. 361, 2016.