MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, Jan-June 2026, Pages: 931-936

ANNUAL RAINFALL FORECASTING USING ARIMA MODEL

[1]S.Prabhakaran, [2]Dr. M. Kannan

[1]Ph.D. Research Scholar, Dept. of CSA,
SCSVMV, Kanchipuram

[1]Assistant Professor, Dept. of Computer Science,
Dharmamurthi Rao Bahadur Calavala Cunnan Chetty's Hindu College, Pattabiram

[2]Assistant Professor, Dept. of CSA,
SCSVMV, Kanchipuram

[1]phdprabha56@gmail.com, [2]saikannan1999@rediffmail.com

Abstract

Rainfall forecasting is important for food, water management and prevention from flood. Forecast rainfall by using variables such as temperature, wind and humidity. We used the Auto-regressive Integrated Moving Average model (ARIMA) to generate rainfall projections for the selected study area of Tamil Nadu. Rainfall data of Tamil Nadu spanning 115 years from 1901 to 2015 were gathered and statistically evaluated.

The Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) approach was used for model identification and diagnostic evaluation for predicting the annual rainfall of the research region. Python programming language was used to identify the appropriate ARIMA $(p, d, q) * (P, D, Q)$ model that fits the rainfall records. The dataset's has been assessed using the Augmented Dickey-Fuller test. The optimal model for forecasting the following decade's rainfall was chosen according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The model's was assessed using root mean square errors (RMSE) and mean squared error (MSE).

Analysis of the rainfall data revealed that ARIMA was the superior model for the annual data with a stationary R-squared value of 0.9978. The optimal ARIMA models were identified, and projections were made for the average annual precipitation for the years 2016 to 2025. The projected rainfall values generated by the ARIMA model were satisfactory and confirmed with normal rainfall data. These findings underscore the reliability of the ARIMA model in climatic studies, particularly in forecasting rainfall patterns. Future research may expand on this model by incorporating additional variables such as temperature, wind and humidity to enhance predictive accuracy.

Introduction

The pattern of the duration of rainfall and temporal and geographical fluctuation of rainfall greatly impact agricultural systems **(Zhang et al. 2012).** Rainfall fluctuations strongly influence the Indian economy which largely relies on agriculture. The diverse seasonal, yearly, and monthly precipitation patterns are crucial for optimising crop management and effectively implementing irrigation. These variations necessitate adaptive strategies among farmers to ensure sustainable yields. Consequently, investment in weather forecasting and water conservation technologies has become increasingly vital to mitigate the risks associated with unpredictable rainfall. Numerous researches have been undertaken in India and internationally about rainfall analysis revealing optimal probability distribution functions for examining rainfall trends **(Sharma and Singh, 2010)**.

Ray et al. (1980) said that the weekly, monthly, and seasonal rainfall patterns and their probability facilitate crop planning by delineating times of drought, regular rainfall, and surplus precipitation.

Advance notification of precipitation aids in the management of water resources and facilitates the implementation of preventative strategies against natural disasters such as floods and droughts. Numerous scholars in India have conducted rainfall forecasting throughout the nation using various spatial and temporal resolutions **(Kaushik &Singh, 2008; Chattopadhyay &Chattopadhyay, 2010; Narayanan et al., 2013).**

Eni and Adeyeye (2015) conducted seasonal ARIMA modelling and forecasting of rainfall in Warri, Nigeria. Various empirical methodologies, such as regression, Autoregressive Integrated Moving Average (ARIMA), fuzzy logic, and artificial neural networks (ANN), are extensively used for rainfall prediction. Empirical methods for rainfall forecasting include analysing historical rainfall data and establishing correlations with either self or other meteorological factors **(Narayanan et al., 2016).**

Valipour (2015) examined the efficacy of seasonal autoregressive integrated moving average (SARIMA) and autoregressive integrated moving average (ARIMA) models for long-term runoff prediction in the United States.

Liu et al. (2017) have examined two time series models such as ARIMA and ARIMA-GARCH concluded that both techniques are adequately effective. They determined that the ARIMA-GARCH model is better appropriate when the variability of the variables is inconsistent across the data range.

Dayal et al. (2019) have developed an ARIMA model to estimate the monthly rainfall in Betwa River Basin, India using India Meteorological Department (IMD) Pune date of 1960–2012. Researchers used Thiessen polygon technique and determined basin-wide monthly average precipitation. The precipitation time series from 1960 to 2000 was used for training to create the best model and 2001–2012 was used for testing and validation. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to identify optimal parameter values and the parsimonious model was ARIMA.

Khan et al. (2023) have used ARIMA modelling to predict rainfall in the Klang River Basin, Selangor for both short-term and long-term periods. They used the Box-Jenkins technique for ARIMA modelling which included Model Identification, Parameter Estimation, Diagnostic Checking, and Forecasting. For data analysis and ARIMA modelling, monthly rainfall data from 1984 to 2019 was used. Analysis of rainfall data revealed ARIMA as the best model for monthly series R2 value is 0.78 and ARIMA for yearly series R2 value is 0.52.

Singh et al (2024) have employed SVM (Support Vector Machine) and SARIMA (Seasonal Autoregressive Integrated Moving Average) models to enhance rainfall forecasts. Understanding precipitation patterns and dynamics is crucial for tackling climate challenges and the SARIMA model anticipates future values. Precise precipitation control ensures fresh water supply and sustainability, underlining its importance in environmental management and resource economics.

Yavuz (2025) have analysed monthly rainfall and temperature changes in Van Province, Türkiye using ARIMA and SARIMA models from 1955 to 2023. The ARIMA temperature models were chosen according to Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values where AIC scores of 788.224 and 172.077 respectively. BIC values of 672.061 and 163.669 were obtained using SARIMA models for rainfall and temperature respectively to handle seasonality.

Hussain et al. (2025) have applied ARIMA and SARIMA models to estimate the climate variables like precipitation, cloud cover, and temperature trends over 80 years (2020-2100). The forecasts indicated a 2.0℃ increase in mean temperature by the end of the 21st century compared to 2019 values. The forecasts indicated a significant decrease in winter months with mean temperatures below 2.0℃, perhaps ending by 2048, thereby harming the region's environment. This study's findings could help the government and stakeholders establish effective climate change plans.

Khan et al. (2025) have examined monthly rainfall forecasts in District Dir (Lower) using multiple ARMA models to determine the best accurate method. ARMA models, used in the Box-Jenkins approach, effectively analysed and forecasted rainfall patterns in varied time series. ARMA model was the best fit for the dataset, as evaluated by SIGMASQ, AIC, and SC criteria, yielding the smallest values and validated by Ljung-Box Q-test. The study accurately predicted monthly rainfall in District Dir (lower) from June 2022 to May 2028 using the ARMA model.

Zhang (2025) have analysed local precipitation patterns using ARIMA model estimated yearly precipitation, and made comparisons between anticipated and measured values. The yearly variation trend of the anticipated value matched the actual precipitation trend, indicating local precipitation features and useful for short-term prediction and trend analysis. To avoid forecasting deviations from long forecasting periods, the ARIMA model used real-time correction year by year for short and medium-term rainfall forecasts.

Ismail et al. (2025) have used historical climate data from 1995-2024 to analyse patterns and predict future changes in climate. They used the ARIMA model to predict climatic trends using the Mann-Kendall test and Sen's slope estimator to measure monotonic temperature and precipitation trends. Results showed an average annual rainfall of 2,653.97 mm.

Materials and Method

The rainfall data for the period from 1901 to 2015 for the state of Tamil Nadu was collected from open government data (OGD) platform India. Monthly rainfall data have been consolidated to generate annual rainfall forecasts. The statistical procedure namely Autoregressive Integrated Moving Average Method (ARIMA) using python linear Ordinary Least Square Method was used to study rainfall forecasting. Finally, to find the best fit model the Root Mean Square Error (RMSE) has been calculated using observed and forecasted rainfall data from 2016 to 2025.

Augmented Dickey–Fuller (ADF) test

Researcher formulated hypotheses as a basis for studying the relationship of the variable is rainfall time series data in Tamil Nadu. Augmented Dickey–Fuller (ADF) test was performed on the annual rainfall series to test whether the data is stationary or not. The test statistic t value is −6.29and p value is 3.66×10^{-8} shows that that p-value is far below 0.05. Since the ADF statistic value −6.29 is more negative than all the critical values and the p-value is much smaller than 0.05, the null hypothesis of a unit root is rejected; therefore, the series can be treated as stationary and no differencing is required.

Serial Correlation Analysis

The Serial correlation, or autocorrelation, is used to identify a relationship between the present value of a variable and its preceding values that need to be examined. It is a genuine method for uncovering hidden trends and patterns in time series data that would otherwise remain undetected **(Alam&Majumder2022).**

In using ARIMA for rainfall forecasting, it is expected that the observed time-series data exhibits serial independence. Significant serial correlation coefficients in time series rainfall data may occur, requiring the examination of serial correlation effects when analysing historical data.

The Autocorrelation function (ACF) and Partial Autocorrelation functions (PACF) were plotted to assess significant autocorrelation coefficients across different lagged values at a 0.05 confidence level (Figure 1 & Figure 2). Lag-1 autocorrelation is frequently utilised to investigate the impact of serial correlation in time series data **(Gourieroux et al. 1985).**

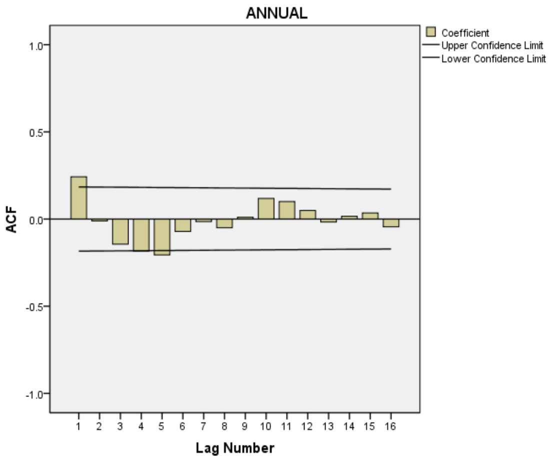

Figure 1 Autocorrelation (ACF) Function for Annual Rainfall Data

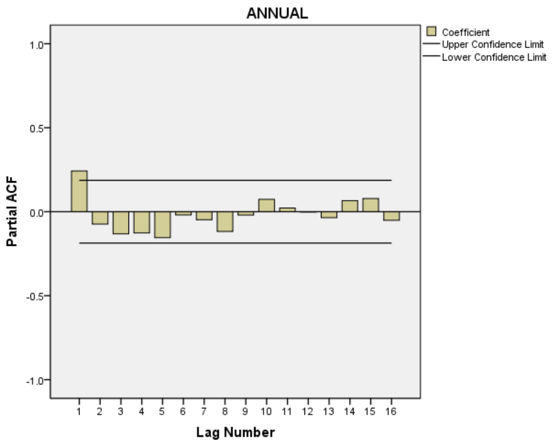

Figure 2 Partial Autocorrelation(PACF) Function for Annual Rainfall Data

In the ACF and PACF charts presented above the horizontal axis represents the lag numbers between the elements of the datasets, while the vertical axis indicates the value of the auto correlated function, which can vary from -1 to 1. A vertical line representing each lag is referred to as a spike on the graph, indicating the value of the autocorrelation function for that specific lag. The autocorrelation at lag zero is consistently equal to one, as it represents the correlation of each term with itself. Each spike that exceeds or falls short of the significance zone is assigned a level of statistical significance.

The ACF and PACF charts are consistent with a weakly stationary series with short-memory AR structure and no strong need for differencing. In ACF chart Lag 1 shows a positive autocorrelation clearly above the confidence band, indicating that rainfall in one year is positively related to rainfall in the immediately preceding year.

Beyond lag 1, the ACF bars are small and mostly within the confidence limits, with a few modest negative values around lags 3 to 5 and small positives near lags 9 to 11. This pattern suggests quickly decaying dependence rather than a slow, monotonic decay, so the series behaves like a stationary process rather than requiring differencing. PACF has a strong significant spike at lag 1 and smaller negative spikes up to about lag 4 with everything after that inside the confidence bands.A large partial autocorrelation at lag 1 plus weaker contributions from lags 2 to 4, and no clear seasonal pattern, supports an AR model of low order formula subtraction between Auto Regressive of first value and Auto Regressive of Second value possibly combined with an MA term as in the ARIMA (2, 0, 1) model.

ARIMA Model

ARIMA model is chosen for this Tamil Nadu rainfall series (1901 - 2015) by combining visual diagnostics of time plot, ACF and PACF with information criterion based model selection for the given data. The time-series plot of annual rainfall shows fluctuations around a roughly constant mean with no persistent upward or downward trend or changing variance. A formal Augmented Dickey–Fuller test on the 1901–2015 series strongly rejects a unit root p-value less than 0.05. So no differencing is required and d=0.

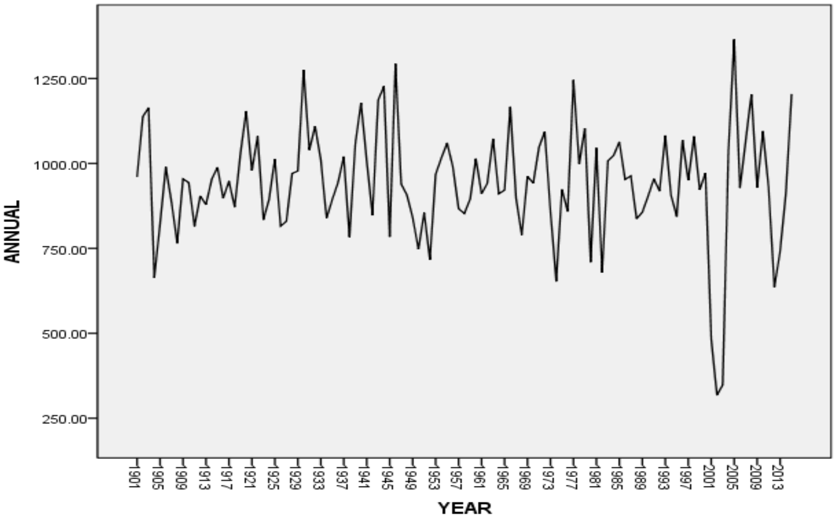

Figure 3 Time series Plot of Tamil Nadu Rainfall Data (1901 – 2015)

The time-series plot shows large year-to-year variability in annual rainfall around a roughly constant mean level with no obvious long-term upward or downward trend. The ACF has one clearly significant positive spike at lag 1 and then small, quickly decaying autocorrelations, with some modest negatives around lags 3 to 5 and small positives near lags 9 to 11 consistent with short-memory AR-type behaviour. The PACF has a strong spike at lag 1 and noticeable but smaller spikes up to about lag 2 to 3, after which partial autocorrelations lie within the confidence bands, suggesting a low-order AR model around subtraction between Auto Regressive of first value and auto regressive of second value.

Rainfall forecasting with the several low-order ARIMA models are fitted for ARIMA models with parameters (1,0,0), (1,0,1), (2,0,0), (2,0,1), (0,0,1) by using Akaike Information Criterion (AIC) and Bayesian information criterion (BIC) values. Among these values, ARIMA model gives the lowest AIC value 498.07 and BIC value 1511.80. In the ARIMA model fit the constant first Auto Regressive value, Second Auto Regressive value and Mean Regressive value of first value coefficients are all of reasonable magnitude and residuals behave approximately like white noise with small autocorrelations inside the confidence limits. This confirms that ARIMA model captures the main structure implied by the ACF and PACF while leaving no strong pattern in the residuals. Therefore it is an appropriate model for this annual rainfall data.

Level, Variability, and outliers

Annual rainfall mostly fluctuates between about 700 mm and 1200 mm, indicating stationary mean but substantial inter-annual volatility. There are a few extreme years, especially around the early 2000s where rainfall drops below 300 mm followed by a sharp rebound above 1200 mm, which appear as outliers that increase the overall variance. Because the mean level is visually stable over time and there is no persistent trend or changing variance, the series is suitable for modelling with a stationary ARIMA model. The pronounced short-run fluctuations and occasional extremes justify including autoregressive and moving-average terms to capture short-memory dependence and shock effects rather than relying on a simple mean model.

Model Validation

The ARIMA model fitted well for the 1901–2015 annual rainfall data and yielded the following in-sample model fit statistics.

Table 1 Model Validation

Model Fit statistics	Mean
Stationary R-squared	0.9978
R-squared	0.1077
RMSE	156.14
MAPE	14.30
MaxAPE	156.74
MAE	115.44
MaxAE	498.42
Normalized BIC	13.15

Stationary R-squared near 1.0 shows the model captures nearly all predictable variation relative to the unconditional process variance. Standard R-squared is modest at 0.11 because annual rainfall has high inherent volatility whereas historical standard deviation approximately 200 mm limiting explained variance in levels. Error metrics like RMSE is 156 mm and MAE is 115 mm indicate typical prediction errors, with MAPE around 14% reflecting percentage accuracy across the series range.

10-year Projection of 1901 – 2015 Rainfall Data Series

In order to generate the forecasts for the ten-year 2016 to 2025, the yearly rainfall in a time-series format from 1901 to 2015 was plotted to check visually for trend, seasonality or changing variance. However, the plot shows fluctuations around a roughly constant mean with no strong trend. Augmented Dickey–Fuller (ADF) test on the indifference series generated ADF statistic value is −6.29 with p-value 3.7×10^{-8} which is more negative than the 1%, 5%, and 10% critical values. So the null of a unit root is rejected and the series is treated as stationary where d = 0. The plot of the ACF and PACF of the level series shows one clear positive spike at lag 1 and then small, quickly decaying correlations then the PACF has a strong spike at lag 1 and smaller spikes up to about lag from 2 to3 suggesting a low-order AR process with a possible Mean average. Several low-order ARIMA models (1,0,0), (2,0,0), (1,0,1), (2,0,1), (0,0,1) were fitted to the 1901–2015 data to estimate the candidate ARIMA models and select the best one. A comparison between AIC and BIC, ARIMA model with lag (2,0,1) gave the lowest AIC value is 1498.07 and a good BIC value is 1511.80. So it was chosen as the preferred model.

Upon inspection of the chosen ARIMA (2,0,1) output constant approximate value is 943.99. Auto Regressive of first value is 1.05, Auto Regressive of second approximate value is −0.33 and Mean Average of first value is −0.81 with residual variance value is 24334 and Ljung–Box tests showed that residuals were approximately white noise confirming the model adequately captures temporal dependence.

Using the fitted ARIMA model forecast for 10 years from 2016 to 2025 were generated based on the annual rainfall data series from 1901 to 2015 with 95% confidence intervals.

Table 2

10-Year Forecasts

Year	Forecast (mm)	Lower 95% (mm)	Upper 95% (mm)
2016	1047.4	741.7	1353.1
2017	966.7	652.2	1281.1
2018	933.8	618.4	1249.1
2019	925.8	606.7	1244.9
2020	928.3	606.3	1250.3
2021	933.6	610.2	1256.9
2022	938.2	614.5	1262.0
2023	941.4	617.5	1265.2
2024	943.1	619.3	1267.0
2025	944.0	620.1	1267.8

Forecasts converge toward the long-run mean of approximately 944 mm per year, consistent with the model's constant term of 943.99 mm. Confidence intervals widen gradually due to accumulating forecast uncertainty but remain reasonably narrow relative to the data's historical range from 318 to1365 mm.

The line graph of the observed fitted, and forecast annual rainfall from 1901 to 2025 is presented in Figure 4.

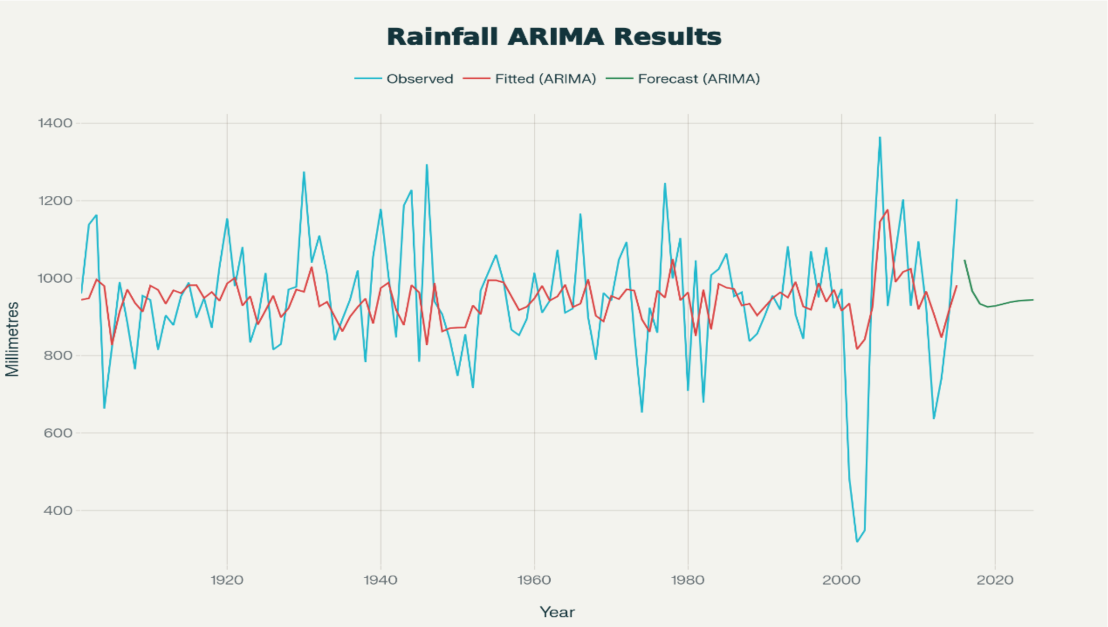

Figure 4 Observed, Fitted, and ARIMA-Forecast Annual Rainfall (1901–2025)

The blue line indicated Fitted ARIMA tracks the observed series over 1901–2015 but is smoother, representing the model's expected rainfall after accounting for autoregressive and moving-average structure. The red line segment indicates Forecast ARIMA from 2016–2025 extends beyond the historical data and gradually converges toward the long-run mean around 940–950 mm reflecting the ARIMA model's projection that future rainfall will fluctuate around this historical average.

Conclusions

The main aim of this research was to evaluate the effectiveness of the ARIMA model for predicting rainfall in Tamil Nadu based on the annual historical time series data. The study used a dataset of 115 years of long-term annual precipitation data spanning from 1901 to 2015. The results of the study proved that traditional ARIMA satisfied suitable model for Forecasting rainfall. It does not necessarily imply that it will consistently be the optimal model for forecasting precipitation. The study rainfall forecasting is vital for the state of Tamil Nadu for reasons ranging from appropriate planning of agriculture activities and avoiding unnecessary losses due to flooding during heavy rainfall. In addition, identifying the suitable model for rainfall prediction is necessary for successful water management.

References

- Alam, M.J. and Majumder, A., 2022. A Comparative Analysis of ARIMA and other Statistical Techniques in Rainfall Forecasting: A Case Study in Kolkata (KMC), West Bengal. *Research Square*, *ARIMA BOX-JENKINS*, pp.1-19.
- Gourieroux, C., Monfort, A. and Trognon, A., 1985. A general approach to serial correlation. *Econometric Theory*, *1*(3), pp.315-340.
- Zhang, Q., Sun, P., Singh, V.P. and Chen, X., 2012. Spatial-temporal precipitation changes (1956–2000) and their implications for agriculture in China. *Global and Planetary Change*, *82*, pp.86-95.
- Sharma, M.A. and Singh, J.B., 2010. Use of probability distribution in rainfall analysis. *New York Science Journal*, *3*(9), pp.40-49.
- Kaushik, I. and Singh, S.M., 2008. Seasonal ARIMA model for forecasting of monthly rainfall and temperature. *Journal of Environmental Research and Development*, *3*(2), pp.506-514.
- Chattopadhyay, S. and Chattopadhyay, G., 2010. Univariate modelling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. *ComptesRendus Geoscience*, *342*(2), pp.100-107.
- Narayanan, P., Basistha, A., Sarkar, S. and Kamna, S., 2013. Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. *ComptesRendus Geoscience*, *345*(1), pp.22-27.
- Narayanan, V.L., Gurubaran, S., Shiokawa, K. and Emperumal, K., 2016. Shrinking equatorial plasma bubbles. *Journal of Geophysical Research: Space Physics*, *121*(7), pp.6924-6935.
- Valipour, M., 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*, *22*(3), pp.592-598.
- Khan, M.M.H., Mustafa, M.R.U., Hossain, M.S., Shams, S. and Julius, A.D., 2023. Short-term and long-term rainfall forecasting using arima model. *International Journal of Environmental Science and Development*, *14*(5), pp.292-298.
- Liu, Y., Wang, B., Zhan, H., Fan, Y., Zha, Y. and Hao, Y., 2017. Simulation of nonstationary spring discharge using time series models. *Water Resources Management*, *31*(15), pp.4875-4890.
- Dayal, D., Swain, S., Gautam, A.K., Palmate, S.S., Pandey, A. and Mishra, S.K., 2019, May. Development of ARIMA model for monthly rainfall forecasting over an Indian River Basin. In *World Environmental and Water Resources Congress 2019* (pp. 264-271). Reston, VA: American Society of Civil Engineers.
- Singh, P., Hasija, T. and Ramkumar, K.R., 2024, July. Enhancing agricultural sustainability: SARIMA and SVM models for precise rainfall forecasting and environmental management. In *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)* (pp. 359-365). IEEE.
- Yavuz, V.S., 2025. Forecasting monthly rainfall and temperature patterns in Van Province, Türkiye, using ARIMA and SARIMA models: a long-term climate analysis. *Journal of Water and Climate Change*, *16*(2), pp.800-818.
- Hussain, K., Farooq, S.U. and Altaf, I., 2025. Time-Series Analysis for Forecasting Climate Parameters of Kashmir Valley Using ARIMA and Seasonal ARIMA Model. *Jordan Journal of Earth & Environmental Sciences*, *16*(1).
- Khan, A., Khan, H.U. and Ismail, M., 2025. Forecasting of Monthly Rainfall in Dir (L) KP Pakistan with ARMA Models. *Journal of Asian Development Studies*, *14*(1), pp.734-750.
- Ismail, I.N., Kamarudin, M.K.A., Abdullah, S.N.F., Hamzah, F.M. and Sunardi, S., 2025, October. Forecasting of Tropical Climate Using Integration on Mann Kendall Test and ARIMA Model for Production of Rice Husk Particle Board. Proceeding of International Exchange and Innovation Conference on Engineering & Sciences, pp. 390-395.
- Zhang, Y., 2025. Spatial characteristics analysis and prediction of precipitation based on ARIMA model. *Procedia Computer Science*, *262*, pp.1316-1321.