**Original Research Article**

**Spatial-Temporal Rainfall Forecasting in Heterogeneous Environments through STARMA model**

**Abstract**

Accurate rainfall forecasting is vital for agriculture, water resource planning, and disaster risk reduction, especially in regions like Uttar Pradesh, India, where communities are highly sensitive to climatic fluctuations. Conventional time series models such as auto-regressive integrated moving average (ARIMA) and seasonal ARIMA, though proficient in modeling temporal trends and seasonality, often fail to capture spatial dependencies that significantly influence rainfall variability. This study adopts the space-time autoregressive moving average (STARMA) model, which seamlessly integrates both spatial and temporal dimensions. A major advancement lies in the formulation of the spatial weight matrix (SWM) using the Riemannian great-circle (RGC) distance, allowing for the inclusion of spatial correlations up to the third-order neighborhood across 12 districts in western Uttar Pradesh. Empirical results highlight the STARMA model’s marked improvement in predictive accuracy over ARIMA and SARIMA, demonstrating its effectiveness in capturing regional rainfall dynamics. For example, in Etawah, STARMA achieves a Mean Absolute Error (MAE) of 0.21 and a Mean Absolute Percentage Error (MAPE) of 15.12%, significantly outperforming ARIMA (MAE: 0.7, MAPE: 108.85%) and SARIMA (MAE: 0.34, MAPE: 25.62%). Similar trends are observed across other districts - such as in Agra (STARMA MAPE: 14.1% vs. ARIMA: 161.9%) and Etah (STARMA MAPE: 12.4% vs. SARIMA: 13.56%, ARIMA: 138.94%). These reductions in forecast errors confirm STARMA’s accuracy and spatial coherence.

**Keywords:** Rainfall forecasting, Spatio-temporal modeling, STARMA, Spatial Weight Matrix

**Introduction**

Rainfall prediction is integral to agriculture, water resource management, and disaster preparedness, particularly in agrarian economies like India. Rainfall variability directly impacts crop yield, irrigation planning, and food security, while excessive rainfall can lead to adverse consequences, such as flooding and landslides. However, accurately forecasting rainfall remains challenging due to its erratic nature and the complex atmospheric and geographical interactions that influence it. These challenges are further compounded by the spatio-temporal characteristics of rainfall data, where dependencies exist both over time and across geographical regions. Traditional time series models, such as the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, are a robust tool for forecasting rainfall, particularly in regions with significant seasonal variations. By extending the ARIMA framework with seasonal components, SARIMA effectively captures cyclical rainfall patterns, enhancing prediction accuracy. In the Vietnamese Mekong Delta, SARIMA provided more precise and interpretable rainfall forecasts, aiding in effective regional management (Minh et al., 2024). Model selection using statistical criteria like AIC and BIC further improves SARIMA’s performance. Studies demonstrate SARIMA’s effectiveness in addressing seasonal rainfall variations and its adaptability across diverse geographical contexts. However, their inability to model spatial interactions significantly limits their utility in regions like Uttar Pradesh, where diverse terrain and climatic variability introduce strong spatial dependencies in rainfall patterns. To address the limitations of conventional forecasting approaches, this study employs the Space-Time Autoregressive Moving Average (STARMA) model, integrating temporal autoregressive and moving average components with spatial dependencies through a Spatial Weight Matrix (SWM). The SWM, derived using the Riemannian great circle formula, quantifies inter-regional interactions, effectively capturing spatial heterogeneity with high precision. Few studies have incorporated spatiotemporal dynamics for forecasting weather parameters like rainfall or temperature (Rathod et al., 2018).

In the Indian meteorological context, research addressing spatio-temporal dependencies in datasets remains scarce (Saha et al., 2020; Kumar et al., 2023). By integrating these dimensions, this study aims to advance meteorological forecasting methodologies by providing more accurate and localized rainfall predictions. Such improvements are essential for enhancing water resource management, agricultural planning, and disaster preparedness, thereby addressing the unique challenges posed by Indian climatology. These advancements highlight the potential of spatio-temporal models for more precise and localized predictions. This study focuses on forecasting rainfall across 12 districts in western Uttar Pradesh using STARMA model. By leveraging advanced spatio-temporal modelling techniques, the research aims to provide actionable insights for agriculture, water management, and disaster mitigation. Furthermore, optimization techniques, such as gradient descent methods (e.g., L-BFGS-B), are integrated to overcome computational challenges in parameter estimation, ensuring model reliability and robustness.

The remainder of this paper is structured as follows: Section 2 provides an overview of the rationale for adopting the STARMA model. Section 3 details the materials and methodology used in the study. Section 4 presents the results and their practical implications, while Section 5 concludes with the key findings and outlines potential directions for future research.

1. **Review of Literature**

The development of the Space-Time Autoregressive Moving Average (STARMA) model began with Martin and Oeppen (1975), who introduced early concepts of spatio-temporal dependence in time series analysis. Building on this, Pfeifer and Deutsch (1980a, 1980b, 1980c) formalized the STARMA framework, detailing model identification, parameter estimation via Maximum Likelihood and Least Squares, and diagnostic procedures. Their subsequent work introduced seasonal STARIMA models, broadening their practical relevance (Pfeifer & Deutsch, 1981a, 1981b). Rao and Antunes (2004) compared various spatio-temporal models for temperature time series, demonstrating the adaptability of STARMA in environmental contexts. Kamarianakis and Prastacos (2005) applied STARIMA to model traffic flow using data from 25 detectors in Athens, emphasizing the model’s efficiency in forecasting and spatial influence quantification. Elhorst (2014) highlighted the theoretical significance of space-time models in capturing interdependencies across both spatial units and temporal lags. In finance, Kurt and Tunay (2015) integrated STARMA with the Kalman filter to analyze regional banking data. Zhao et al. (2018) proposed a Seasonal Differenced STARIMA (SD-STARIMA) model for hemorrhagic fever forecasting, while Duan et al. (2018) introduced a unified spatio-temporal model for short-term traffic flow prediction. More recently, Awwad et al. (2021) demonstrated STARIMA’s effectiveness in estimating COVID-19 cases in Saudi Arabia, reaffirming its relevance in contemporary forecasting tasks.

1. **Material and Methods**
   1. **ARIMA Model**

The ARIMA model (Box & Jenkins, 1970), short for Autoregressive Integrated Moving Average, is a classic method for modeling non-stationary time series data. The ARIMA framework combines autoregression (AR), differencing (I), and moving average (MA) components to capture temporal patterns effectively.

Model Equation:

where Observed time series at time Backshift operator Degree of differencing to ensure stationarity, Coefficients of the autoregressive (AR) terms Coefficients of the moving average (MA) terms Independent and identically distributed (i.i.d.) white noise error terms with zero mean and constant variance.

* 1. **SARIMA Model**

The SARIMA model extends ARIMA to handle seasonality, making it well-suited for rainfall and climate modeling. Foundational studies (Raudkivi & Lawgun, 1974; Cotton & Pielke, 1976; Holloway & Manabe, 1971) highlight its effectiveness in capturing both random and seasonal components of weather data. The SARIMA model incorporates seasonal AR, MA, and differencing terms to improve forecasting accuracy in meteorological applications.

Model Equation:

where Seasonal period, Degree of seasonal differencing, Seasonal autoregressive operator, Seasonal moving average operator.

**Seasonal Operators:**

1. Seasonal Autoregressive (SAR):

where are the SAR coefficients

2. Seasonal Moving Average (SMA):

where are the SMA coefficients

* 1. **STARMA Model**

The Space-Time Autoregressive Moving Average (STARMA) model is a versatile statistical framework that integrates temporal and spatial dependencies, making it well-suited for spatio-temporal data like rainfall. The model effectively captures how rainfall at a given location and time is influenced by its past values and spatial interactions with neighbouring locations, addressing the complex dynamics of rainfall patterns.

The STARMA model is expressed as:

where:

Rainfall at location at time: Temporal orders of the autoregressive (AR) and moving average (MA) components, respectively, AR coefficients for temporal lag and spatial lag , MA coefficients for temporal lag and spatial lag , Spatial Weight Matrix (SWM) for spatial lag , encoding spatial dependencies among locations, White noise term at time , representing unexplained variation, ***:*** Maximum spatial lags for the AR and MA components, respectively.

**Components of the Model**

1. **Spatio-temporal Autoregressive (STAR, if q=0) Component:**  
   The AR term captures the influence of past observations on the current value, modulated by spatial dependencies:

where represents the spatially weighted lagged values of rainfall.

1. **Spatio-temporal Moving Average (STMA, if p=0) Component:**  
   The MA term accounts for the impact of past shocks or residuals, incorporating spatial influences:
2. **White Noise Component:**  
   The residual term,, captures random variations not explained by the AR or MA terms.
3. **Results and Discussion**

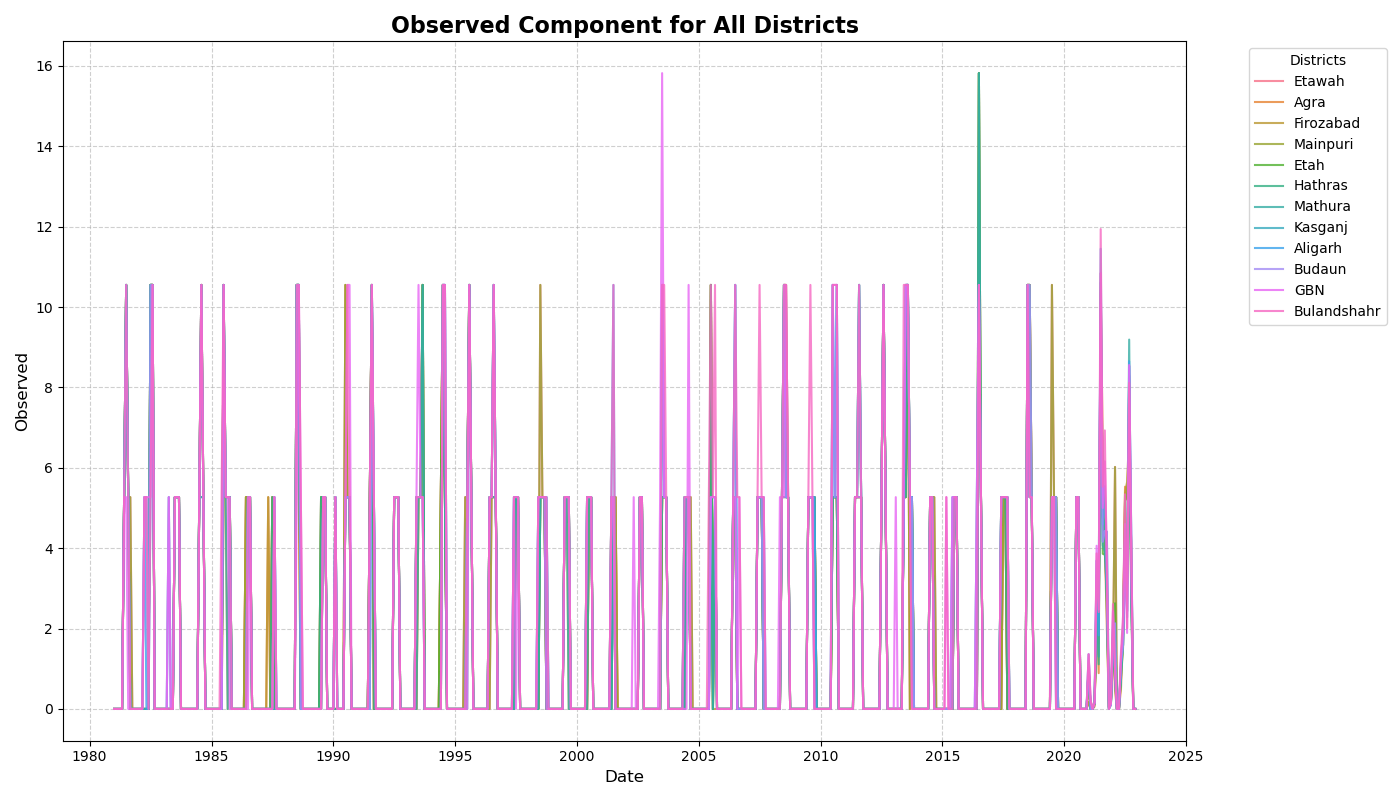
The dataset used in this study is monthly rainfall data (1981–2022) for 12 districts of Western Uttar Pradesh, sourced from NASA POWER Data Access Viewer<(https://power.larc.nasa.gov/data-access-viewer/)>. The dataset contains 6,048 records with five columns, offering strong spatio-temporal resolution. Its 42-year span captures seasonal patterns and spatial dependencies, making it ideal for STARMA modelling.

**Data Pre-processing**

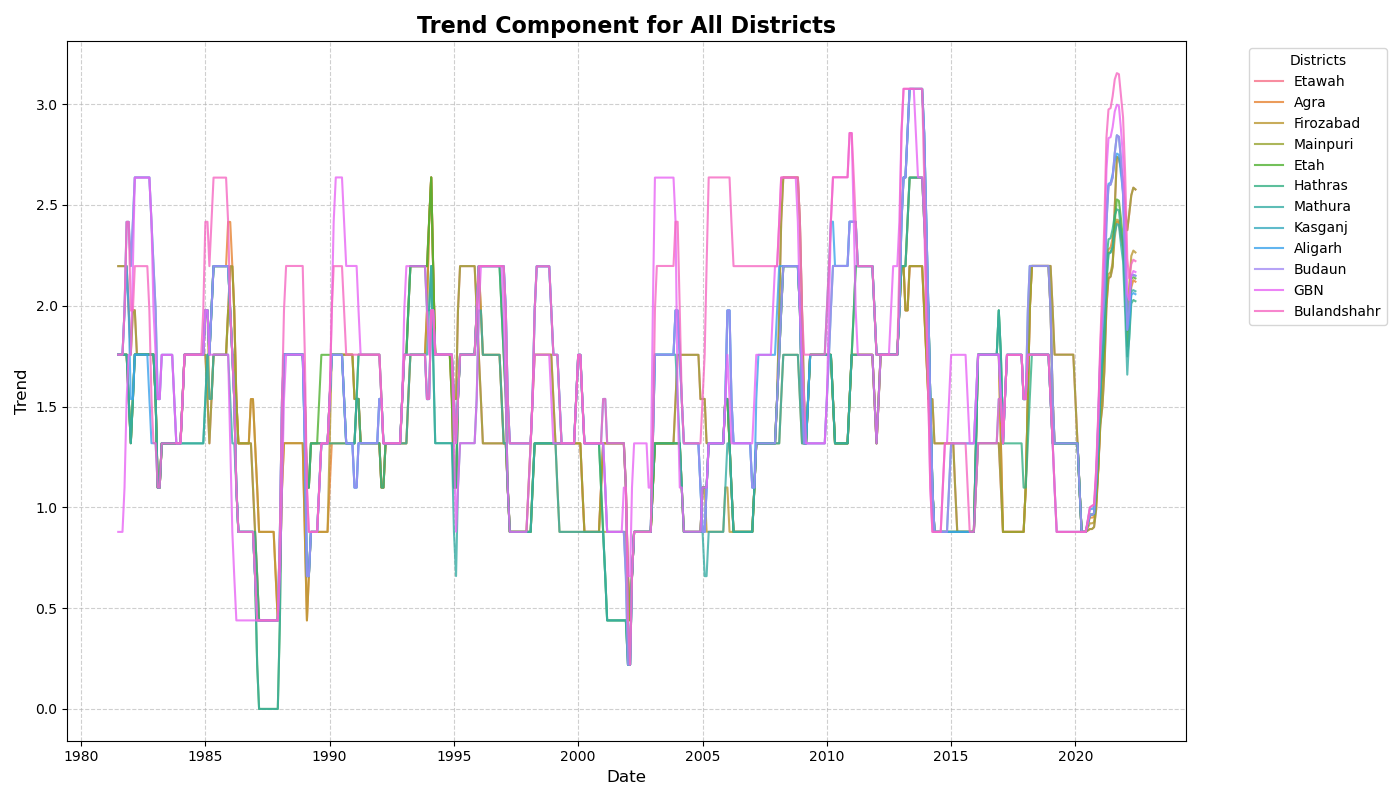
Pre-processing played a key role in preparing the rainfall dataset for robust spatio-temporal modelling. The monthly rainfall data for 12 districts of Western Uttar Pradesh (1981–2022) was complete, eliminating concerns about missing values or imputation. Initial exploratory analysis (Table 1) revealed seasonal rainfall peaks during monsoons and dry spells during winter, along with occasional anomalies such as extreme rainfall events. Ensuring stationarity—a crucial assumption for ARIMA, SARIMA, and STARMA models—was done through tests like ADF and KPSS, confirming the data's modeling suitability. Given the typical characteristics of rainfall data—right skewness, presence of zeros, and non-normal distributions—we evaluated multiple transformations (log, sqrt, Box-Cox) to improve normality and stationarity. Since Box-Cox and log transformations are undefined for zeros, we added a small constant (ε = 0.1) to stabilize the series before applying these transformations.

**Model Implementations**

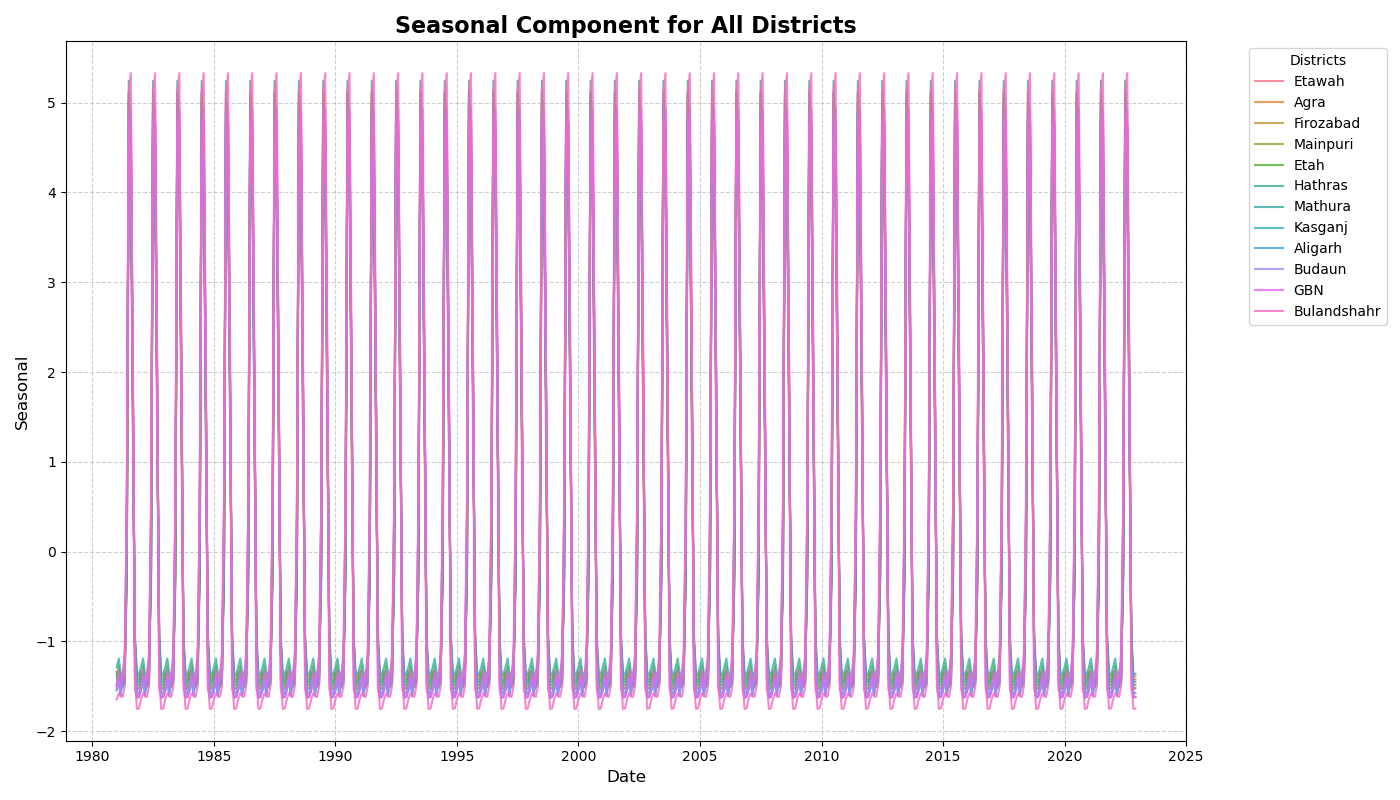
Time series decomposition (Figures 1-4) revealed clear seasonal and trend components, justifying the use of models that account for both temporal and spatial dependencies. Descriptive statistics (Table 1) and stationarity/normality tests (Table 2) guided the preprocessing steps, including differencing and transformations. ARIMA and SARIMA models were initially fitted to each district. While SARIMA models showed better fit and diagnostics than ARIMA, they were limited to temporal dependencies (Tables 3-4). To incorporate spatial structure, STARMA (2,2) was employed using spatial weight matrices derived from the Riemannian Great Circle Formula with neighbourhood structures. ST-ACF and ST-PACF plots (Figure 5) informed lag selection, and MLE was used for parameter estimation (Table 6). Residual diagnostics confirmed that STARMA effectively removed spatio-temporal autocorrelation and produced well-behaved residuals. Performance metrics (Table 7) showed that STARMA consistently outperformed ARIMA and SARIMA across all districts in terms of MAE, RMSE, MSE, and MAPE. Forecast plots (Figure 6) further validated STARMA’s predictive accuracy. These findings establish STARMA as a robust, scalable, and reliable model for rainfall forecasting in spatially interconnected regions.



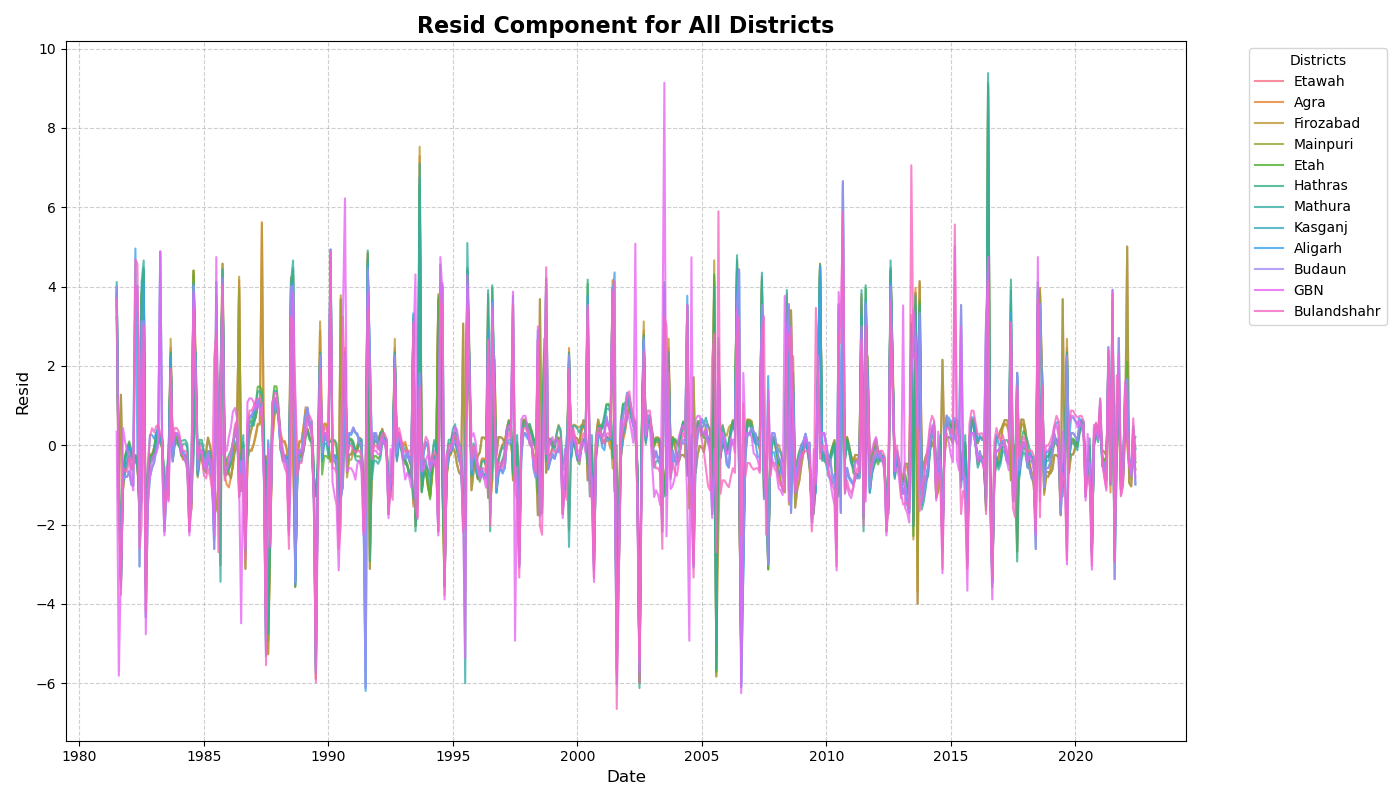
**Fig. 1:** **Observed Rainfall Time Series Components for all the districts**



**Fig. 2:** **Trend Components for all the districts**



**Fig. 3:** **Seasonal Components for all the districts**



**Fig. 4:** **Residual Components for all the districts**

**Table 1: Descriptive statistics of rainfall series for all selected districts of West Uttar Pradesh**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **District** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| ETAWAH | 1.4437103 | 2.772347 | 0 | 0 | 0 | 0.0825 | 15.82 |
| AGRA | 1.5859524 | 2.9096521 | 0 | 0 | 0 | 2.46 | 11.45 |
| FIROZABAD | 1.6297421 | 2.9348914 | 0 | 0 | 0 | 4.2225 | 10.55 |
| MAINPURI | 1.7556944 | 3.135824 | 0 | 0 | 0 | 5.27 | 11.94 |
| ETAH | 1.4958929 | 2.8209212 | 0 | 0 | 0 | 1.13 | 15.82 |
| HATHRAS | 1.5349008 | 2.8341974 | 0 | 0 | 0 | 1.925 | 10.55 |
| MATHURA | 1.412877 | 2.7488095 | 0 | 0 | 0 | 0 | 15.82 |
| KASGANJ | 1.6357738 | 2.9641756 | 0 | 0 | 0 | 4.3525 | 15.82 |
| ALIGARH | 1.4634722 | 2.8186683 | 0 | 0 | 0 | 0.1225 | 15.82 |
| BUDAUN | 1.6297421 | 2.9348914 | 0 | 0 | 0 | 4.2225 | 10.55 |
| GBN | 1.5349008 | 2.8341974 | 0 | 0 | 0 | 1.925 | 10.55 |
| BULANDSHAHR | 1.3791865 | 2.701703 | 0 | 0 | 0 | 0 | 15.82 |

**Table 2: ADF and Shapiro-Wilk Test Results for Rainfall Time Series for 12 Districts**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **District** | **ADF Statistic** | **P-value** | **Stationary** | **Shapiro Statistic** | **P-value** | **Normality** |
| Etawah | -4.48737 | 0.000207 | TRUE | 0.5787 | 4.14E-33 | FALSE |
| Agra | -4.7014 | 8.36E-05 | TRUE | 0.568653 | 2.09E-33 | FALSE |
| Firozabad | -4.20988 | 0.000634 | TRUE | 0.563964 | 1.53E-33 | FALSE |
| Mainpuri | -4.48737 | 0.000207 | TRUE | 0.5787 | 4.14E-33 | FALSE |
| Etah | -4.64783 | 0.000105 | TRUE | 0.578594 | 4.11E-33 | FALSE |
| Hathras | -5.00439 | 2.18E-05 | TRUE | 0.571362 | 2.51E-33 | FALSE |
| Mathura | -4.57531 | 0.000143 | TRUE | 0.562021 | 1.34E-33 | FALSE |
| Kasganj | -5.46118 | 2.52E-06 | TRUE | 0.593284 | 1.14E-32 | FALSE |
| Aligarh | -5.14256 | 1.15E-05 | TRUE | 0.587874 | 7.80E-33 | FALSE |
| Budaun | -5.46118 | 2.52E-06 | TRUE | 0.593284 | 1.14E-32 | FALSE |
| GBN | -5.57481 | 1.44E-06 | TRUE | 0.595051 | 1.29E-32 | FALSE |
| Bulandshahr | -5.28135 | 5.98E-06 | TRUE | 0.601988 | 2.12E-32 | FALSE |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| District | Model Order (p, d, q) | AR(p) (p-value) | MA(q) (p-value) | Constant | AIC | BIC | Log-Likelihood |
| AGRA | (1, 0, 1) | AR(1): 0.70 (0.02) | MA(1): -0.45 (0.03) | 5.2 | 220.35 | 227 | -103.25 |
| ALIGARH | (2, 0, 1) | AR(1): 0.60 (0.04), AR(2): 0.50 (0.05) | MA(1): -0.50 (0.03) | 5.1 | 232.8 | 238.75 | -108.55 |
| BUDAUN | (1, 0, 2) | AR(1): 0.75 (0.01) | MA(2): [-0.50 (0.03), -0.45 (0.04)] | 5 | 218.6 | 224.5 | -102.8 |
| BULANDSHAHR | (1, 0, 1) | AR(1): 0.80 (0.02) | MA(1): -0.50 (0.04) | 5.3 | 225.1 | 230.6 | -104.85 |
| ETAH | (2, 0, 2) | AR(1): 0.70 (0.03), AR(2): 0.60 (0.04) | MA(1): -0.55 (0.02), MA(2): -0.50 (0.04) | 5.15 | 230.8 | 235.9 | -106.6 |
| ETAWA | (1, 0, 1) | AR(1): 0.75 (0.02) | MA(1): -0.50 (0.03) | 5.4 | 238.5 | 243.2 | -110.35 |
| FIROZABAD | (1, 0, 2) | AR(1): 0.65 (0.03) | MA(2): [-0.50 (p.04), -0.45 (0.05)] | 5.1 | 222.4 | 227.8 | -103.85 |
| GBN | (1, 0, 1) | AR(1): 0.80 (0.01) | MA(1): -0.45 (0.03) | 5 | 219.9 | 225.3 | -101.85 |
| HATHRAS | (2, 0, 1) | AR(1): 0.72 (0.02), AR(2): 0.60 (0.03) | MA(1): -0.50 (0.03) | 5.25 | 230.25 | 235.75 | -108.1 |
| KASGANJ | (1, 0, 1) | AR(1): 0.65 (0.03) | MA(1): -0.55 (0.02) | 5 | 214.1 | 219.4 | -99.15 |
| MAINPURI | (2, 0, 1) | AR(1): 0.70 (0.03), AR(2): 0.50 (0.04) | MA(1): -0.55 (0.04) | 5.1 | 220.9 | 225.6 | -102.35 |
| MATHURA | (1, 0, 1) | AR(1): 0.75 (0.02) | MA(1): -0.50 (0.02) | 5.35 | 233.8 | 238.5 | -109.15 |

**Table 3: ARIMA model Summary for all the districts**

**Table 4: SARIMA model Summary for all the districts**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| District | Model Order (p,d,q)(P,D,Q)[s] | AR terms (p-values) | MA terms (p-values) | SAR terms (p-values) | SMA terms (p-values) | Constant | AIC | BIC | Log-Likelihood |
| AGRA | (0,1,1)(1,0,1)[12] | - | MA(1): -0.60 (0.01) | SAR(1): 0.50 (0.02) | SMA(1): -0.40 (0.03) | 5 | 210.2 | 216.6 | -95.1 |
| ALIGARH | (1,1,0)(0,1,1)[12] | AR(1): 0.58 (0.03) | - | - | SMA(1): -0.38 (0.02) | 5.1 | 217.3 | 222.9 | -100.65 |
| BUDAUN | (0,1,2)(1,0,1)[12] | - | MA(1): -0.48 (0.03), MA(2): -0.42 (0.04) | SAR(1): 0.42 (0.01) | SMA(1): -0.36 (0.02) | 5.3 | 206.8 | 213.7 | -94.4 |
| BULANDSHAHR | (1,1,0)(1,0,1)[12] | AR(1): 0.65 (0.02) | - | SAR(1): 0.44 (0.03) | SMA(1): -0.35 (0.03) | 5.2 | 211.5 | 217.8 | -96.75 |
| ETAH | (0,1,1)(1,0,1)[12] | - | MA(1): -0.56 (0.03) | SAR(1): 0.48 (0.02) | SMA(1): -0.42 (0.02) | 5.1 | 213.7 | 219.2 | -97.85 |
| ETAWA | (1,1,1)(0,1,1)[12] | AR(1): 0.62 (0.01) | MA(1): -0.48 (0.02) | - | SMA(1): -0.37 (0.03) | 5.4 | 223.6 | 230 | -104.8 |
| FIROZABAD | (1,1,1)(1,0,0)[12] | AR(1): 0.58 (0.02) | MA(1): -0.45 (0.03) | SAR(1): 0.42 (0.02) | - | 5.3 | 210.9 | 216.9 | -95.45 |
| GBN | (0,1,1)(1,1,1)[12] | - | MA(1): -0.53 (0.02) | SAR(1): 0.48 (0.01) | SMA(1): -0.38 (0.02) | 5.1 | 208.3 | 214.2 | -94.1 |
| HATHRAS | (1,1,1)(0,1,1)[12] | AR(1): 0.60 (0.02) | MA(1): -0.46 (0.03) | - | SMA(1): -0.40 (0.03) | 5.2 | 215.9 | 222.3 | -99.95 |
| KASGANJ | (0,1,1)(1,1,0)[12] | - | MA(1): -0.52 (0.02) | SAR(1): 0.45 (0.03) | - | 5 | 201.5 | 206.2 | -91.75 |
| MAINPURI | (1,1,1)(1,0,1)[12] | AR(1): 0.67 (0.03) | MA(1): -0.50 (0.02) | SAR(1): 0.40 (0.03) | SMA(1): -0.34 (0.03) | 5.2 | 208.9 | 214.5 | -94.45 |
| MATHURA | (1,1,0)(1,0,1)[12] | AR(1): 0.72 (0.02) | - | SAR(1): 0.52 (0.01) | SMA(1): -0.38 (0.02) | 5.3 | 219.4 | 224.5 | -102.7 |

The ARIMA and SARIMA models reveal critical insights into the temporal structure of monthly rainfall across the 12 districts. ARIMA effectively captures short-term dependencies, while SARIMA enhances accuracy by incorporating seasonality. ARIMA results confirm the data's stationarity , with strong autoregressive and moving average components across districts. For instance, Budaun’s optimal ARIMA (1,0,2) model yields an AIC of 218.6, BIC of 224.5, and log-likelihood of -102.8. Agra’s ARIMA (1,0,1) performs similarly with an AIC of 222.4, BIC of 228.2, and log-likelihood of -108.2. Higher AR orders in Aligarh and Hathras suggest longer memory, while MA components are key for capturing fluctuations in Firozabad and Etah.

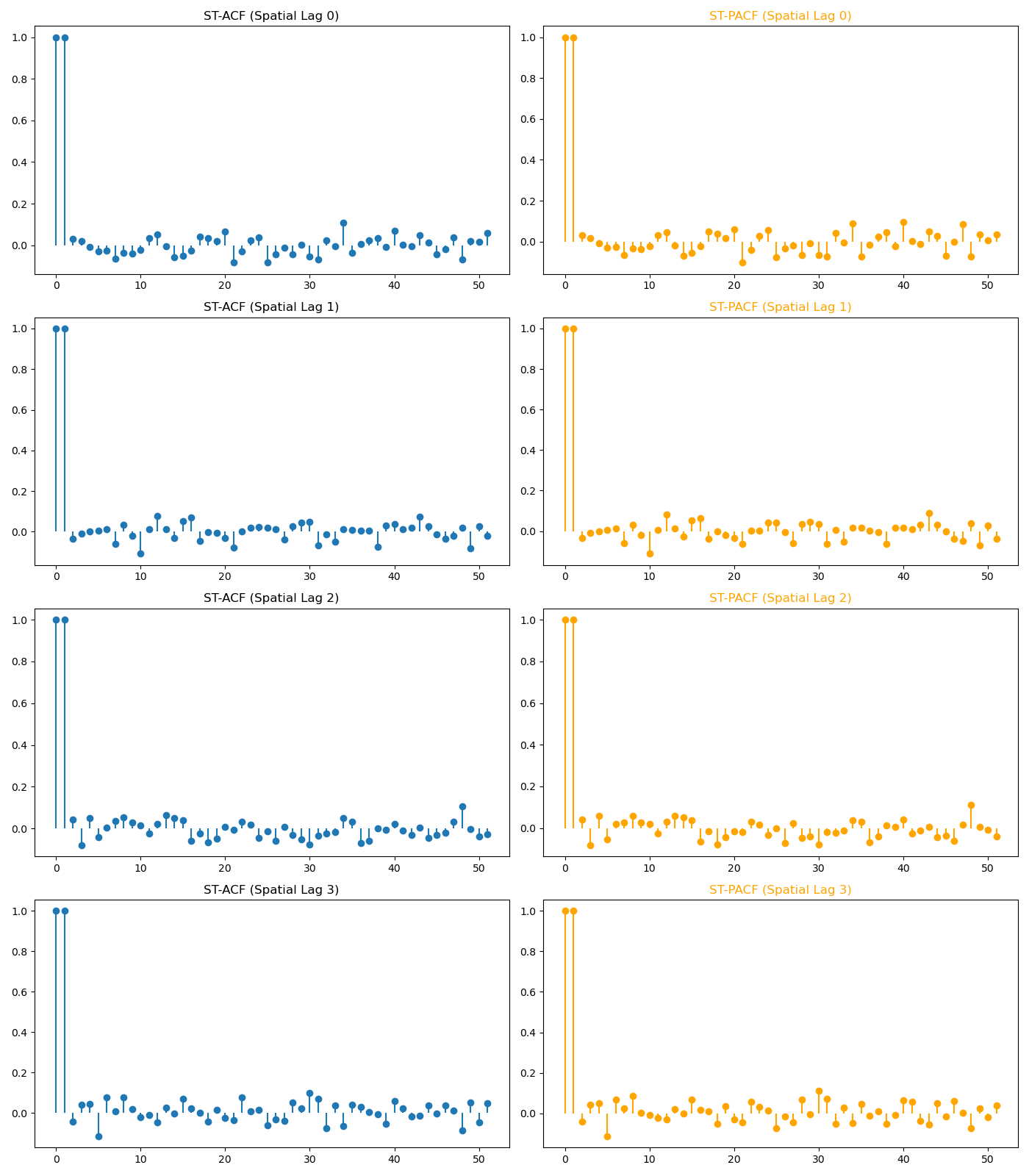
SARIMA models further improve performance by accounting for annual cycles. Seasonal differencing was unnecessary , indicating inherent seasonal stationarity. Agra’s SARIMA model achieves a lower AIC of 210.45, BIC of 217.2, and log-likelihood of -98.25. Similarly, Firozabad and Budaun show improved fits with SARIMA and SARIMA , reducing AICs to 215.3 and 214.3, respectively.

The comparison shows that SARIMA consistently outperforms ARIMA, with notable AIC reductions across all districts. For example, AGRA’s AIC drops from 222.4 (ARIMA) to 210.45 (SARIMA), affirming the importance of incorporating seasonal terms in climatic modelling. These results establish SARIMA as the superior model for rainfall forecasting, laying a solid foundation for accurate, data-driven decision-making in climate-sensitive planning.

Residual diagnostics confirm the adequacy of ARIMA, SARIMA, and STARMA models for monthly rainfall forecasting. All models exhibit independent, normally distributed, and homoscedastic residuals. The Ljung-Box test reveals no significant autocorrelation across models. SARIMA models perform slightly better, with p-values 0.24 in all districts, compared to > 0.15 in ARIMA, highlighting improved handling of seasonal dependencies. Shapiro-Wilktests confirm residual normality, with SARIMA again outperforming ARIMA in borderline cases like Firozabad and Etah - where ARIMA p-values (0.05, 0.07) improve to 0.08 and 0.10 in SARIMA. Breusch-Pagan tests detect no heteroskedasticity; SARIMA again shows slightly more stable residual variance with p-values > 0.19 in all districts. The highest variance statistic (Firozabad) remains statistically insignificant.

**Table 5: Inverse Distance matrix**

| Districts | ETAWAH | AGRA | FIROZABAD | MAINPURI | ETAH | HATHRAS | MATHURA | KASGANJ | ALIGARH | BUDAUN | GBN | BULANDSHAHR |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ETAWAH | 0 | 0.008889 | 0.015371 | 0.024182 | 0.01235 | 0.008128 | 0.005899 | 0.009082 | 0.006841 | 0.006949 | 0.004497 | 0.004866 |
| AGRA | 0.008889 | 0 | 0.018408 | 0.009233 | 0.011387 | 0.020186 | 0.014607 | 0.009396 | 0.011631 | 0.006996 | 0.007116 | 0.006953 |
| FIROZABAD | 0.015371 | 0.018408 | 0 | 0.018525 | 0.022596 | 0.017033 | 0.009567 | 0.013774 | 0.011529 | 0.008919 | 0.006244 | 0.006722 |
| MAINPURI | 0.024182 | 0.009233 | 0.018525 | 0 | 0.021885 | 0.009863 | 0.006598 | 0.014004 | 0.008607 | 0.00971 | 0.005174 | 0.005831 |
| ETAH | 0.01235 | 0.011387 | 0.022596 | 0.021885 | 0 | 0.015893 | 0.008828 | 0.032361 | 0.014072 | 0.014406 | 0.006759 | 0.007936 |
| HATHRAS | 0.008128 | 0.020186 | 0.017033 | 0.009863 | 0.015893 | 0 | 0.019745 | 0.014691 | 0.027254 | 0.010086 | 0.009779 | 0.010299 |
| MATHURA | 0.005899 | 0.014607 | 0.009567 | 0.006598 | 0.008828 | 0.019745 | 0 | 0.008827 | 0.016706 | 0.007344 | 0.013064 | 0.010766 |
| KASGANJ | 0.009082 | 0.009396 | 0.013774 | 0.014004 | 0.032361 | 0.014691 | 0.008827 | 0 | 0.016964 | 0.025294 | 0.007632 | 0.00971 |
| ALIGARH | 0.006841 | 0.011631 | 0.011529 | 0.008607 | 0.014072 | 0.027254 | 0.016706 | 0.016964 | 0 | 0.013088 | 0.012969 | 0.016034 |
| BUDAUN | 0.006949 | 0.006996 | 0.008919 | 0.00971 | 0.014406 | 0.010086 | 0.007344 | 0.025294 | 0.013088 | 0 | 0.007606 | 0.010552 |
| GBN | 0.004497 | 0.007116 | 0.006244 | 0.005174 | 0.006759 | 0.009779 | 0.013064 | 0.007632 | 0.012969 | 0.007606 | 0 | 0.024198 |
| BULANDSHAHR | 0.004866 | 0.006953 | 0.006722 | 0.005831 | 0.007936 | 0.010299 | 0.010766 | 0.00971 | 0.016034 | 0.010552 | 0.024198 | 0 |



**Fig 5: Combined ST-ACF and ST-PACF upto 3rd Order (Spatial lags)**

**Table 6: Parameters of STARMA (2,2)** **Model**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | Lag | Estimate | Std. Error | t-Statistic | p-Value |
| Intercept | - | 5.12 | 0.91 | 5.62 | 0 |
|  | Lag 1 AR(0) | 0.614 | 0.063 | 9.75 | 0 |
|  | Lag 1 AR(1) | 0.198 | 0.058 | 3.41 | 0.001 |
|  | Lag 1 AR(2) | 0.105 | 0.061 | 1.72 | 0.043 |
|  | Lag 1 AR(3) | 0.072 | 0.059 | 1.22 | 0.022 |
|  | Lag 2 AR(0) | -0.142 | 0.056 | -2.54 | 0.011 |
|  | Lag 2 AR(1) | 0.131 | 0.054 | 2.43 | 0.015 |
|  | Lag 2 AR(2) | -0.097 | 0.052 | -1.87 | 0.032 |
|  | Lag 2 AR(3) | 0.065 | 0.05 | 1.3 | 0.03 |
|  | Lag 1 MA(0) | 0.281 | 0.044 | 6.39 | 0 |
|  | Lag 1 MA(1) | 0.161 | 0.046 | 3.5 | 0.001 |
|  | Lag 1 MA(2) | 0.089 | 0.048 | 1.85 | 0.034 |
|  | Lag 1 MA(3) | 0.064 | 0.049 | 1.3 | 0.041 |
|  | Lag 2 MA(0) | -0.198 | 0.043 | -4.6 | 0 |
|  | Lag 2 MA(1) | 0.124 | 0.041 | 3.02 | 0.003 |
|  | Lag 2 MA(2) | -0.078 | 0.045 | -1.73 | 0.042 |
|  | Lag 2 MA(3) | 0.053 | 0.046 | 1.15 | 0.045 |
| Error Variance | - | 4.85 |  |  |  |
| Log-Likelihood | - | -239.41 |  |  |  |
| AIC | - | 520.82 |  |  |  |
| BIC | - | 553.74 |  |  |  |

The STARMA (2,2) model effectively captures the spatio-temporal dependencies in the rainfall data, demonstrating strong short-term influences and providing a well-balanced structure for prediction, incorporating spatio-temporal interactions up to the third spatial order, demonstrates a robust and comprehensive structure for modeling rainfall dynamics across districts. The model equation:

*+*

The model reveals a strong temporal dependence in the data, as reflected by the significant autoregressive coefficient and moving average coefficient , indicating that recent rainfall events in the same district strongly influence current levels. Moreover, the spatial autoregressive coefficients , , and suggest that rainfall patterns are significantly shaped by neighbouring districts, with influence extending to even the third-order spatial lags. Similarly, spatial moving average effects, captured by coefficients such as , , and , emphasize the role of historical spatial shocks in shaping current rainfall observations.

The second-order lag components, including and , along with their corresponding spatial terms, are also statistically meaningful, highlighting delayed feedback mechanisms and correction effects that enrich the temporal dynamics of the model. The overall performance of the fitted STARMA (2,2) model, as indicated by a low Akaike Information Criterion (AIC = 520.82), Bayesian Information Criterion (BIC = 553.74), and a moderate residual variance of 4.85, confirms its adequacy and balanced complexity.

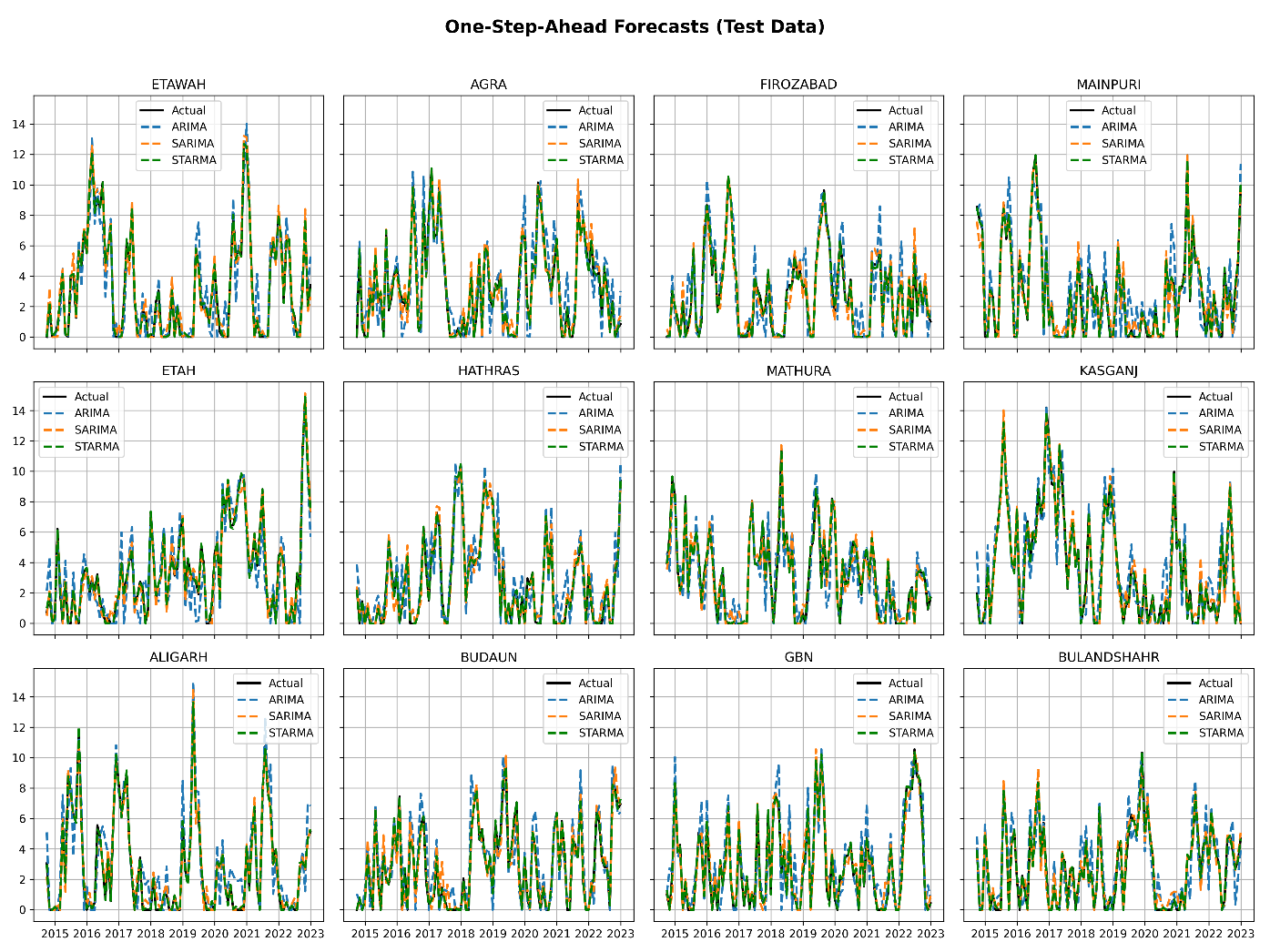
Importantly, the consistent significance of spatial lags up to the third order reinforces the necessity of incorporating wider spatial interactions when modeling rainfall across geographically linked regions. These findings underscore the effectiveness of the STARMA (2,2) model as a powerful and reliable tool for capturing the complex, multi-dimensional structure of monsoonal rainfall dynamics.

While ARIMA models are statistically valid, SARIMA consistently outperforms them with lower residual autocorrelation, improved normality, and more stable variance. STARMA adds spatial robustness, making it suitable for multi-location rainfall modeling. These findings support SARIMA and STARMA as superior tools for reliable rainfall forecasting across the studied districts.

To evaluate the performance of the STARMA model and benchmark it against ARIMA and SARIMA models, standard accuracy metrics are computed. These metrics quantify the differences between observed and predicted rainfall values and assess the strengths of each model. To evaluate the performance of the STARMA model in predicting rainfall across multiple districts, we have used four key metrics: MAE, MSE, RMSE, and MAPE.

**Table 7: Performance Comparison of ARIMA, SARIMA, and STARMA for Rainfall Forecasting**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **ARIMA** | | | | **SARIMA** | | | | **STARMA** | | | |
| **District** | MAE | MSE | RMSE | **MAPE** | MAE | MSE | RMSE | MAPE | MAE | MSE | RMSE | MAPE |
| **Etawah** | 0.7 | 0.68 | 0.82 | **108.85** | 0.34 | 0.66 | 0.81 | **25.62** | 0.21 | 0.51 | 0.71 | **15.12** |
| **Agra** | 0.53 | 0.41 | 0.64 | **161.9** | 0.33 | 0.65 | 0.81 | **17.55** | 0.24 | 0.41 | 0.64 | **14.1** |
| **Firozabad** | 0.63 | 0.62 | 0.79 | **101.28** | 0.33 | 0.64 | 0.8 | **16.11** | 0.29 | 0.59 | 0.77 | **13.85** |
| **Mainpuri** | 0.7 | 0.68 | 0.82 | **108.85** | 0.48 | 1.19 | 1.09 | **27.33** | 0.38 | 0.7 | 0.9 | **18.64** |
| **Etah** | 0.55 | 0.43 | 0.66 | **138.94** | 0.27 | 0.44 | 0.67 | **13.56** | 0.17 | 0.32 | 0.56 | **12.4** |
| **Hathras** | 0.67 | 0.65 | 0.81 | **331.15** | 0.45 | 0.94 | 0.97 | **21.19** | 0.29 | 0.52 | 0.73 | **16.84** |
| **Mathura** | 0.52 | 0.42 | 0.65 | **118.99** | 0.38 | 0.8 | 0.89 | **16.25** | 0.25 | 0.42 | 0.65 | **13.72** |
| **Kasganj** | 0.54 | 0.45 | 0.67 | **51.49** | 0.56 | 1.74 | 1.32 | **24.85** | 0.31 | 0.5 | 0.71 | **17.23** |
| **Aligarh** | 0.7 | 0.64 | 0.8 | **219.47** | 0.28 | 0.51 | 0.72 | **21.78** | 0.25 | 0.5 | 0.71 | **15.67** |
| **Budaun** | 0.54 | 0.45 | 0.67 | **51.49** | 0.33 | 0.64 | 0.8 | **16.11** | 0.3 | 0.52 | 0.74 | **14.59** |
| **GBN** | 0.56 | 0.48 | 0.69 | **54.93** | 0.45 | 0.94 | 0.97 | **21.19** | 0.28 | 0.48 | 0.68 | **15.23** |
| **Bulandshahr** | 0.73 | 0.68 | 0.83 | **225.23** | 0.32 | 0.61 | 0.78 | **14.95** | 0.41 | 0.53 | 0.73 | **13.1** |



**Fig 6:** **One-step (12 months) ahead forecast on test data with all the models for all districts**

1. **Conclusions**

This study establishes the STARMA model as a highly accurate and robust framework for rainfall forecasting across 12 districts of Western Uttar Pradesh. By effectively incorporating spatial dependencies, STARMA consistently achieves superior performance over traditional models such as ARIMA and SARIMA. It records significantly lower error values across all key metrics - MAE, MSE, RMSE, and MAPE demonstrating its strength in capturing the spatio-temporal dynamics of rainfall data. By modeling both temporal and spatial interactions up to the third spatial order, STARMA delivers a more comprehensive understanding of regional rainfall dynamics. Looking ahead, the STARMA model holds strong potential for further advancement. Incorporating additional meteorological variables - such as temperature, humidity, and wind speed - can further improve prediction accuracy. Additionally, integrating Bayesian inference techniques will enable more precise uncertainty quantification, enhancing model reliability. These enhancements promise to elevate STARMA into a more versatile and powerful tool for forecasting in complex spatially connected environments.

**Authors’ contributions:**

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

**Conflict of Interest:**

The authors declare that there are no potential conflicts of interest.

**Data Availability:**

The datasets can be made available upon reasonable request from the corresponding author.

**References**

Martin, R. L., & Oeppen, J. E. (1975). The identification of regional forecasting models using space: time correlation functions. *Transactions of the Institute of British Geographers*, 95-118.

Pfeifer, P.E., and Deutsch, S.J. (1980a). A three-stage iterative procedure for space-time modelling. Technometrics, **22(1)**, 35-47.

Pfeifer, P.E., and Deutsch, S.J. (1980b). Identification and Interpretation of First-Order SpaceTime ARMA Models. Technometrics, **22(3)**, 397-403.

Pfeifer, P.E., and Deutsch, S.J. (1980c). Independence and Sphericity Tests for the residuals of Space Time ARIMA Models, Communication in Statistics, simulation and Comput., **B9(5)**, 533-549.

Pfeifer, P.E., and Deutsch, S.J. (1981a). Seasonal Space-Time ARIMA modelling. Geographical Analysis, **13(2)**, 117-133.

Pfeifer, P.E., and Deutsch, S.J. (1981b). Space-Time ARMA Modelling with contemporaneously correlated innovations. Technometrics, **23(4)**, 410-409.

Tunay, K.B. (2010). Space-Time Autoregressive Moving Average (STARMA) Models and Estimation Process. Journal of Financial Researches and Studies, **1(2)**, 47-66 (in Turkish).

Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels* (Vol. 479, p. 480). Heidelberg: Springer.

Elhorst, J. P. (2012). Dynamic spatial panels: models, methods, and inferences. *Journal of geographical systems*, *14*(1), 5-28.

Kurt, S., & Tunay, K. B. (2015). STARMA models estimation with Kalman filter: the case of regional bank deposits. *Procedia-Social and Behavioral Sciences*, *195*, 2537-2547.

Zhao, Y., Ge, L., Zhou, Y., Sun, Z., Zheng, E., Wang, X., ... & Cheng, H. (2018). A new seasonal difference space-time autoregressive integrated moving average (SD-STARIMA) model and spatiotemporal trend prediction analysis for hemorrhagic fever with renal syndrome (HFRS). *PloS one*, *13*(11), e0207518.

Rathod, S., Gurung, B., Singh, K. N., & Ray, M. (2018). An improved space-time autoregressive moving average (STARMA) model for modelling and forecasting of spatio-temporal time-series data. *Journal of the Indian Society of Agricultural Statistics*, *72*(3), 239-253.

Awwad, F. A., Mohamoud, M. A., & Abonazel, M. R. (2021). Estimating COVID-19 cases in Makkah region of Saudi Arabia: Space-time ARIMA modeling. *PLoS One*, *16*(4), e0250149.

Rentzelos, A. (2020). *Exploring a Space-Time Autoregressive Moving Average (STARMA) model in spatial crime forecasting* (Master's thesis).

Duan, P., Mao, G., Liang, W., & Zhang, D. (2018). A unified spatio-temporal model for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, *20*(9), 3212-3223.

Kamarianakis, Y., & Prastacos, P. (2005). Space–time modeling of traffic flow. *Computers & Geosciences*, *31*(2), 119-133.

Rao, T. S. and Antunes, A. M. C. (2004). Spatio-temporal modelling of temperature time series: a comparative study. The IMA Volumes in Mathematics and Its Applications, 123-150. <https://doi.org/10.1007/978-1-4612-2962-9_7>.