International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 10 No. 4, 2024
https://www.theaspd.com/ijes.php

Comparative Analysis Of Arima, Sarima And Ets Models For Forecasting Monthly Precipitation In Manipur, India Using High-Resolution Satellite Data

Dr. Ngasepam Pikeshwor Singh[1*], Irom Luckychand Meitei[2], Rishikesh Chongtham[3], Dr. Sumitra Salam[4], Dr. Rajkumari Haripriya Devi[5], Dr. Longjam Ibochoubi Singh[6], Taibangjam Loidang Chanu[7], Manglembi Ningthoujam[8]

[1*]Assistant Professor, Department of Geography Dhanamanjuri University. Email-ngpikeshwor@gmail.com
[2]Assistant Professor, Department of Geography. Nambol L Sanoi College. Email- luckychand565@gmail.com
[3]MA Passed Out Student, Department of Geography, Dhanamanjuri University. Email-rishikeshch26@gmail.com
[4]Assistant Professor, Department Of Botany, Nambol L. Sanoi College, Email-sumitrasalam@gmail.com.
[5]Assistant Professor, Department of Environmental Science, Nambol L. Sanoi College, Email-haripriyark1974@gmail.com
[6]Assistant Professor, Department of Geography, Nambol L. Sanoi College. Email- ibochoubisingh@gmail.com.
[7]Assistant Professor, College of Agricultural Engineering & Post Harvest Technology (CAU), Ranipool, Sikkim. Email- loidangtt@gmail.com
[8]MA (Geography), Passed out Student, Department of Geography, Panjab University, Chandigarh, India. Email- manglembiningthoujam123@gmail.com

Abstract

This research study examines the Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), and Exponential Smoothing State Space Model (ETS) models for the monthly precipitation forecast over Manipur, India with the high resolution satellite data, PERSIANN-PDIR-Now data set of the time span between 2001 and 2023. Spatially consistent preprocessing was carried out for the data by using ArcGIS. Analysis of R for the data sets to analyze time series. Additive Decomposition, which is of type Decomposition, is conducted as seasonality changed without correlation to the trend magnitude. Three models have been developed and validated with different metrics such as RMSE, MAE, and MASE. The best model is an ARIMA(5,0,1) that captures short-term time dependencies, and hence, indicates moderate accuracy but less ability to deal with seasonal effects. The SARIMA(1,0,0)(1,1,0)[12] model is more reliable because it captures both seasonal and non-seasonal components with minimal bias. ETS (A,N,A) shows good positive additive seasonality but with mild residual autocorrelation, which indicates that it lacks proper freedom to deal with irregular fluctuations. There is a steady increase of monsoonal intensity over the years, according to the trend analysis. Seasonal components present predictable peaks in the case of monsoon. The random component presents anomalies, such as extreme rainfall in 2018, which presents global climatic events like El Niño. SARIMA achieves a good balance in terms of being accurate yet uncomplicated for places that exhibit the pronounced seasonality. The implications of this paper are climate resilience and support better water resource management, agricultural planning, and disaster preparedness in the state of Manipur.

Keywords: *ARIMA, RMSE, SARIMA, ETS. PERSIANN-PDIR-Now*

INTRODUCTION

Precipitation forecasting is one of the integral parts of climate analysis and resource management particularly in regions where seasonal variations are more pronounced (Alexander and Block 2022). Precipitation determines the water resources and has a direct impact on agriculture, infrastructure, and disaster vulnerabilities (Ning et al., 2023; Wang et al., 2024; Liu et al., 2023; Lowe et al., 2013). Regions such as the Indian state of Manipur rely solely on monsoon as deciding factors of critical annual precipitation patterns; therefore, it covers much of the temporal and spatial variability (Mathew et al., 2021). Under such conditions, it becomes very important to receive the correct forecast of precipitations for ensuring food security, avoiding flood risk, and in preparing one to deal with deficiencies of precipitations. Even though meteorological modeling has improved, precipitation prediction remains very challenging for climatically complex regions. Traditional methods cannot adequately represent the complexity of local climate variability, especially when this local effect is inextricably linked with a larger-scale global climatic oscillation. The challenges point toward innovative tools and methods that might enhance predictive accuracy while accounting for intricate climate dynamics.

This study aims to surpass these obstacles by applying sophisticated time-series forecasting techniques in order to analyze and predict monthly precipitation for the 22-year period in Manipur (2001–2023). The global high-

resolution satellite dataset of 0.04° x 0.04° or = 4km x 4km PERSIANN-PDIR-Now will be utilized. The satellite data can provide excellent spatial and temporal precipitation estimates accuracy as per the given information of Nguyen et al., 2020. This new product of near-real-time IR precipitation has been developed by the Centre for Hydrometeorology and Remote Sensing at the University of California, Irvine. This is a development of the earlier PERSIANN algorithm of PERSIANN–Cloud Classification System (Gorooh et al., 2022). In regions with hardly any land surface meteorological observation sites, PERSIANN-PDIR-Now datasets are particularly valuable in combining the characteristics and variability in the precipitation phenomenon. Such datasets are able to enable the observation of local details of precipitation dynamics, yet balance those with regional influences of the climate. Its principal advantage compared with other real-time precipitation products is that it incorporates high frequency-sampled IR imagery. This method guarantees the least delay possible, while the data becomes available for the rainfall within a 15-to-60 minutes interval after an event (Nguyen et al., 2020). The other error of IR imagery along with its uncertainties is greatly diminished in PDIR-Now. All the three primary statistical models used in this paper: ARIMA, SARIMA, and ETS are a three most prevalent statistical models based on time series research and prediction techniques. Such individual models provide a number of strengths and weaknesses in such a manner that they can easily be applied to various features concerning precipitation variability. The ARIMA model is a good method for forecasting an univariate time series since it captures short-term temporal dependencies.

It further combines AR, I, and MA components to deal with non-stationarity in data. It produces important temporal relations and deals with residual variability. Its flexibility and statistical precision make it extremely popular for precipitation forecasting in most contexts. Many studies successfully applied this model, including Narayanan et al. (2013) and Lai and Dzombak (2020). However, ARIMA models have limitations in dealing with large seasonal variations that limit their applicability to high-periodicity datasets. To cope with the problem of seasonality with ARIMA, the SARIMA model was developed. Because SARIMA could model both the non-seasonal as well as seasonal changes, SARIMA is the most effective tool for datasets displaying periodic patterns, such as the monthly precipitation under the influence of monsoonal cycles. With the addition of parameters on seasonal autoregressive (SAR), seasonal differencing (SD) and seasonal moving averages (SMA), the SARIMA model more precisely captures the details of precipitation variability. It is thus this model's capacity to absorb local climatic dynamics into broad seasonal trends that has made it such a useful tool for regions such as Manipur, whose annual precipitation patterns are dominated by monsoonal rainfall. The ETS model is especially effective for data sets that exhibit seasonal variation, and it applies smoothing parameters to estimate components of this sort in an adaptive way.

It has additive error and seasonal formulations like ETS (A,N,A) configuration. This model is suitable for analysis of precipitation data with relatively stable seasonal patterns. ETS models might not do very well in case of irregularities or anomalies, but simplicity and interpretability make them excellent choices for capturing additive seasonal dynamics in precipitation. For evaluating the performance of the ARIMA, SARIMA, and ETS models, a set of statistical metrics will be applied. These are RMSE, MASE, and residual autocorrelation. These would be the criteria by which one model would compare the other and show how the two can be good for the precipitation amount forecast at a place such as Manipur. The results, therefore, conclude the importance of model selection against the specific characters of the dataset. This, in itself, is the culmination of the fact that statistical measures of evaluation will help in building up credibility and validity concerning the forecasting model. Besides model evaluation, this study will also be enlightening enough on the climatic trends and anomalies of the past two decades in Manipur. The implications of the results of this study are going to be crucially important for building climate resilience and sustainable development in Manipur. Precise precipitation forecasts would help in informing agricultural planning. Thus, crops selected and the strategies of irrigation will be optimally determined. Precipitation periods with a lot of rain can be foreseen, enabling precautionary measures to be undertaken in protecting fragile communities and infrastructures. Precipitation deficits forecasting can guide water resource management to ensure supplies during the dry seasons. Policymakers and planners may use the results of this research as a basis for developing adaptive strategies that minimize impacts arising from climatic variability and extreme weather events.

Objectives

This study aims to improve the understanding and prediction of precipitation variability in Manipur by using advanced time-series models. It tries to identify and analyze long-term trends, seasonal patterns, and irregularities in precipitation data over a 22-year period. It has also been attempted to provide reliable forecasts of monthly precipitation, enabling informed decision-making for water resource management, agriculture, and disaster preparedness. Various attempts have been made to evaluate the performance of each model in capturing the nuances of precipitation variability, including short-term dependencies, seasonal fluctuations, and anomalies..

METHODOLOGY

1. Data Source and Characteristics

PERSIANN-PDIR-Now is a high-resolution global satellite precipitation product obtained from the Centre for Hydrometeorology and Remote Sensing, University of California, Irvine; these precipitation data are used here in this study. The dataset can be used to produce precipitation estimates with spatial resolution at 0.04° x 0.04°, that is, around 4 km x 4 km with 15–60 minutes of latency. It is, thus very appropriate for near-real-time hydrological applications and particularly for flood forecasting and inundation mapping (Nguyen et al., 2020). The coverage spans from March 1, 2000, to date and has coverage from 60°S to 60°N. The most important advantage of the PERSIANN-PDIR-Now dataset is that it can correct the errors in the infrared imagery-based precipitation estimates through techniques like dynamic shifting of brightness temperature-rain rate (Tb-R) curves using rainfall climatology (Nguyen et al., 2020). Data for this study were downloaded from the CHRS data portal in October 2024 and included monthly aggregated rainfall estimates from January 2001 to December 2023.

2. Data Pre-processing in ArcGIS

2.1 Spatial Data Processing

The downloaded precipitation raster files were processed using ArcGIS to ensure consistency and accuracy. All raster datasets were reprojected to Projected Coordinate System UTM Zone 46N using the "Raster Projection" tool, enabling accurate spatial analysis and zonal statistics.

2.2 Zonal Statistics

To derive the mean monthly rainfall for each spatial unit, the 'Zonal Statistics as Table' tool was applied to the reprojected raster datasets. This operation summarized the spatial rainfall distribution for each month, yielding tabular outputs of mean monthly rainfall values for the entire study period. The resultant monthly tables were combined into a single comprehensive dataset using the 'Merge' Tool in ArcGIS. This step ensured the temporal alignment and integration of monthly precipitation values, forming the basis for time-series analysis.

3. Data Transformation for Time-Series Analysis

The merged dataset was exported to a CSV format and subsequently imported into R for statistical analysis. Precipitation data were converted to a time-series format, enabling detailed exploration of seasonal patterns, trends, and variability. Steps included handling missing values, duplicate removal, and conversion of monthly data to a "tsibble" object indexed by time which is a data structure optimized for time series analysis in the R-Software.

4. Time Series analysis

A time series analysis was performed for the monthly precipitation in Manipur for the period from January 2001 to December 2023 using R software. To gain a deeper understanding of the underlying patterns in the data, a decomposition of the time series was also performed. Adjusted Dickey-Fuller (ADF) test was first applied to test the stationarity of the time series data. Accordingly, the time series was decomposed into trend, seasonal and random or residual components. Trend component reflects the long-term progression of the data, seasonal component shows the repeating patterns or cycles in the data while residual is the noise or unexplained variance in the data. Furthermore, for a better interpretation of the results, several visualizations such as time-series Plot, decomposed time-series plot and forecast comparison plot were generated.

5. Model Development

Three forecasting models were used to predict future precipitation values: ARIMA, SARIMA, and ETS. For the Augmented Dickey-Fuller test, the time series data was tested for stationarity. The Augmented Dickey-Fuller test is crucial before using any of the forecast models because it determines if a time series is stationary-an assumption that needs to be satisfied by these models. Stationarity means the statistical properties of the series, such as mean and variance, remain constant over time. The ARIMA and SARIMA models require a stationary series to ensure accurate modeling and forecasting because non-stationary data leads to unreliable results. The ADF test determines whether there is a unit root in the series, which is a sign of non-stationarity. If the series is non-stationary, differencing or transformation is applied to make it stationary. On the other hand, ETS models handle non-stationary data differently, but understanding stationarity helps choose the most suitable model and improve forecasting accuracy.

The ADF test evaluates the null hypothesis that a time series has a unit root, indicating non-stationarity, against the alternative hypothesis of stationarity. The test is based on the following regression equation:

$$y_t - y_{t-1} = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i (y_{t-i} - y_{t-i-1}) + \epsilon_t$$

Here, y_t represents the time series, $y_t - y_{t-1}$ is the first difference of y_t, t is the time trend, γ tests for the presence of a unit root, α is a constant, β represents the coefficient of the time trend, δ_i are the coefficients of

lagged differences, and ϵ_t is the error term. The parameter γ is of particular interest, as the null hypothesis "$H_0: \gamma = 0$" corresponds to the presence of a unit root (non-stationarity), and the alternative hypothesis "$H_0: \gamma < 0$" corresponds to stationarity.

5.1 ARIMA Model

The ARIMA model would be used for the time series forecasting and analysis, with consideration of the historical patterns of precipitation based on past data. This univariate time series modeling technique utilizes the three primary components involved in effectively capturing the characteristics of the data in question. The autoregressive component models how a certain value depends on some or all of the previous values to account for the temporal dependencies. The integrated component allows the time series to be stationary. Differencing removes the trends or seasonal effects that obscure underlying patterns in a time series. The MA component adds influence of past errors from forecasting, so the models have a good predictive capability as well as correcting for residual variability. All these together create a well-structured framework of modeling and forecasting of time-series data. The order (5,0,1) ARIMA model with nonzero mean is used in this paper for time series analysis and forecasting. The model applied here is the ARIMA(p, d, q), with p=5 representing five autoregressive terms, d=0 as there is no differencing, and q=1 for one moving average term. This is because the observed data characteristics have a mean level that is consistent and not zero. The best model selection is based on statistical evaluation criteria, which balances the model fit and complexity. One of the most commonly used metrics is the Akaike Information Criterion (AIC), which tries to find the model that best explains the data without overfitting. It does so by penalizing models that have more parameters. Lower values of the AIC correspond to a better fit to trade for the complexity of the model. That means good and simple models get more favored. The BIC, short for Bayesian Information Criterion, is the same in philosophy but it uses an even stricter penalty term to increase its focus on simplicity of models. It considers the number of parameters and sample size in addition to the above, which makes it very useful for large datasets with lots of observations. Since the number of parameters and dataset size appear in its penalty term, the BIC generally will favor simpler models if the top candidates differ little in their explanatory power. These criteria, AIC and BIC, validate the selected model such that it is not only an adequate fit to the data but also avoids excessive complexity, which might tend to overfit and reduce the generalizability of the model by optimal minimizations in the evaluation of a model that gives considerable predictions with optimum interpretability. The general form of the ARIMA (5,0,1) model is given by:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_3 y_{t-3} + \varphi_4 y_{t-4} + \varphi_5 y_{t-5} + \theta_1 \varepsilon_{t-1} + \mu + \varepsilon_t$$

Here, y_t represents the observed time series at time t, and ε_t is the white noise error term. This equation encapsulates the relationship among the time series observations, their past values, and the error components, providing a reliable framework for forecasting and analysis.

5.2 SARIMA Model

The SARIMA model (Seasonal Autoregressive Integrated Moving Average) extends the ARIMA model by incorporating seasonal components. The notation for a SARIMA model is typically expressed as:

SARIMA (p,d,q)(P,D,Q)[s]

Where, p, d, and q are non-seasonal autoregressive (AR), Differencing (I), and Moving Average (MA) parameters, respectively; P, D and Q are Seasonal autoregressive (SAR), differencing (SI), and moving average (SMA) parameters, respectively; and s is seasonal period.

In this study, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model is applied to analyse and forecast the time series data by capturing both seasonal and non-seasonal components. The model was selected in the same way as with the ARIMA model based on AIC and BIC metrics. The selected model, SARIMA (1,0,0)(1,1,0)[12] with drift, is specifically tailored to reflect the seasonal nature of the data, with a seasonal period of 12. The general form of the SARIMA model is expressed as:

$$\phi(B)(1-B)^d\left(1-B^s\right)^D Y_t = \Psi(B)\varepsilon_t$$

In this equation, Y_t represents the observed time series at time t, d denotes the degree of non-seasonal differencing, and s indicates the seasonal period. $\phi(B)$ and $\Psi(B)$ are polynomials in the backshift operator B, which capture the autoregressive (AR) and moving average (MA) components, respectively. The parameter D accounts for seasonal differencing, and ε_t denotes the error term. The SARIMA(1,0,0)(1,1,0)[12] configuration specifies its components clearly. The model includes a single non-seasonal autoregressive term (AR(1)) and a seasonal autoregressive term (SAR(1)) with a lag of 12. The degree of non-seasonal differencing (d=0) indicates that the data is already stationary, while the seasonal differencing (D=1) addresses the seasonal patterns. No moving average terms (MA(0)) are incorporated in the model. Additionally, the inclusion of a drift term accounts for any systematic linear trends over time.

5.3 ETS Model

This study employs the Exponential Smoothing State Space Model (ETS) to perform time series forecasting, focusing specifically on the ETS(A,N,A) configuration, which incorporates additive error, no trend, and additive seasonality. The objective is to analyze the model's effectiveness in predicting future values of the given time series dataset by estimating its parameters and evaluating its forecasting accuracy through statistical measures such as AIC, AICc, and BIC.

The ETS(A,N,A) model is mathematically expressed as,

$$y_t = l_{t-1} + s_{t-m} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$
$$s_t = s_{t-m} + \gamma\varepsilon_t$$

Where, y_t is the observed value at time t; l_t is the level at time t; s_t is the seasonal component from 'm' periods ago (for monthly data, m=12); ε_t is the error term at time 't' with zero mean and variance σ^2, assumed to follow a normal distribution. The smoothing parameters α and γ determine the contributions of the level and seasonal adjustments, respectively, to the model.

A zero mean error term ensures that the errors are not systematically biased in one direction (positive or negative). They average out to zero, implying that the model is, on average, correct in capturing the relationship between variables.

The dataset was split into two subsets for applying forecast models. A training data set was formed with the data of historical precipitation records from January 2001 to December 2022. The testing data consist of precipitation records from January 2023 to December 2023. This two data-subsets were used for model validation and performance evaluation. The training data is used to fit the forecasting models, while the testing data is reserved for evaluating the accuracy of the predictions.

5.4 Forecasting

Once the models were trained on the historical data, forecasts were generated for the period from January 2023 to December 2023, using the test data. The forecasted values of precipitation from each model (ARIMA, SARIMA, and ETS) were extracted. Each of the models provided a series of forecasted precipitation values for the 12-month period. These forecasted points were combined into a single data frame for comparison and further analysis.

5.5 Accuracy Testing

The performance of the ARIMA, SARIMA and ETS models in forecasting monthly precipitation was evaluated using standard accuracy metrics, including Mean Error (ME), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE), Autocorrelation of Residuals at lag 1 (ACF_1) and Adjusted Rand Index (ARI). These metrics are briefly described below:

5.6 Mean Error (ME):

ME measures the average bias in the model's predictions. A value close to zero indicates minimal systematic overprediction or underprediction. A positive ME indicates the model tends to underestimate the actual values, while negative values indicate overestimation. ME measures bias in the predictions but does not provide information about the overall accuracy. ME alone is insufficient for evaluating model performance as positive and negative errors may cancel each other out.

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})$$

Where, n is the total number of observations; y_i is the actual value of the i^{th} observation; $\hat{y_i}$ is the predicted value for the i^{th} observation.

5.7 Root Mean Square Error (RMSE):

RMSE evaluates the magnitude of prediction errors by penalizing large errors more heavily due to squaring them before averaging. It provides insight into the average error magnitude in the same units as the original data, with smaller values indicating better model performance.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$$

Where, n is the total number of observations; y_i is the observed value for the i^{th} observation; $(y_i - \hat{y_i})^2$ is the squared difference between the actual and predicted values, which eliminates negative signs and emphasizes errors and $\hat{y_i}$ is the predicted value for the i^{th} observation

5.8 Mean Absolute Error (MAE):

MAE calculates the average magnitude of prediction errors without considering their direction. Unlike RMSE, it treats all errors equally, making it an efficient measure for evaluating overall accuracy.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$$

Where, n is the total number of observations; y_i is the observed value for the i^{th} observation; $\hat{y_i}$ is the predicted value for the i^{th} observation; $|y_i - \hat{y_i}|$ is the absolute value of the difference between actual values and predicted values

5.9 Mean Percentage Error (MPE):

MPE assesses the average percentage deviation of the forecast from actual values. Positive MPE indicates that, on average, predictions are underestimating the actual values, while negative MPE indicates that predictions are overestimating the actual values.

$$\text{MPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y_i}}{y_i} \times 100$$

Where, n is the total number of observations; y_i is the observed value for the i^{th} observation and $\hat{y_i}$ is the predicted value for the i^{th} observation

5.10 Mean Absolute Percentage Error (MAPE):

MAPE is the average of absolute percentage errors, providing an intuitive measure of prediction accuracy in relative terms. However, it can be unreliable when data values are small, leading to disproportionately large percentages.

$$\text{MPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y_i}}{y_i}\right| \times 100$$

Where, n is the total number of observations; y_i is the observed value for the i^{th} observation and $\hat{y_i}$ is the predicted value for the i^{th} observation

5.11 Mean Absolute Scaled Error (MASE):

MASE compares the model's error to the average error of a baseline (seasonal naïve) model. Values less than 1 indicate better performance than the baseline, while values greater than 1 suggest underperformance. MASE is scale-independent, making it suitable for comparing datasets with different units. It is calculated using the formulae:

$$\text{MASE} = \frac{1}{N}\sum_{t-1}^{N}\frac{|\hat{y}_t - y_t|}{\text{mean}\,(|y_{t-1} - y_t|)}$$

Where, 'N' represents the number of forecasting points; $\hat{y}_t$ is the forecasted value at time t; y_t is the actual value at time t; $y_{t-1} - y_t$ represents the absolute difference between consecutive observations and $\text{mean}\,(|y_{t-1} - y_t|)$ is the mean of these absolute differences, which serves as a scale factor reflecting the level of persistence in the time series.

5.12 Autocorrelation of Residuals at Lag 1 (ACF_1):

ACF_1 evaluates the correlation of residuals (errors) at lag 1, indicating whether patterns in the data remain unaccounted for by the model. Values close to zero suggest well-behaved residuals, indicating that the model has effectively captured temporal dependencies. It helps assess the extent to which the residuals are correlated with their previous values, providing insight into the model's adequacy.

The equation for the autocorrelation of residuals at lag 1 (ACF1) is given by:

$$\text{ACF1} = \frac{\sum_{T=2}^{N}(\hat{e}_t - \bar{\hat{e}})(\hat{e}_{t-1} - \bar{\hat{e}})}{\sum_{t=1}^{N}(\hat{e}_t - \bar{\hat{e}})^2}$$

Where, $\hat{e}_t$ are the residuals from the model at time t; $\bar{\hat{e}}$ is the mean of the residuals; The numerator captures the covariance between the residuals at lag 1 and the denominator measures the variance of the residuals.

The methodology provide above outlines the process of forecasting precipitation data using three distinct time series models: ARIMA, SARIMA, and ETS. By comparing the forecasted values against actual observations from January 2023 to December 2023, we assess the accuracy of each model using standard metrics. The decomposition of the time series also provides a valuable perspective on the trend and seasonality in the precipitation data. The results from this study can be used to inform decision-making in sectors such as agriculture, water resource management, and climate modelling.

RESULTS AND DISCUSSION

Analysis of monthly precipitation time series in Manipur from January 2001 to December 2023 was made to check the trend of average monthly rainfall. For decomposition of the time series, it is important to establish whether the given time series is additive or multiplicative. A plot of time series (Fig. No. 2) showing the values of the series in millimeters of precipitation exhibited trends and variability across time; thus, an effective check of seasonal as well as trend components is in order.

From an initial look at the time series plot in Fig. No. 2, variations in the levels of precipitation with significant seasonal patterns are seen. These seasonal patterns point to a requirement for decomposition so that trend, seasonality, and residual components could be identified separately. Additive and multiplicative models were both used to this effect. The multiplicative model, however, could not be applied since zero and near-zero values appeared in the precipitation data. This has the limitation that multiplicative models demand strictly positive values in order to calculate proportional effects due to trend and seasonality.

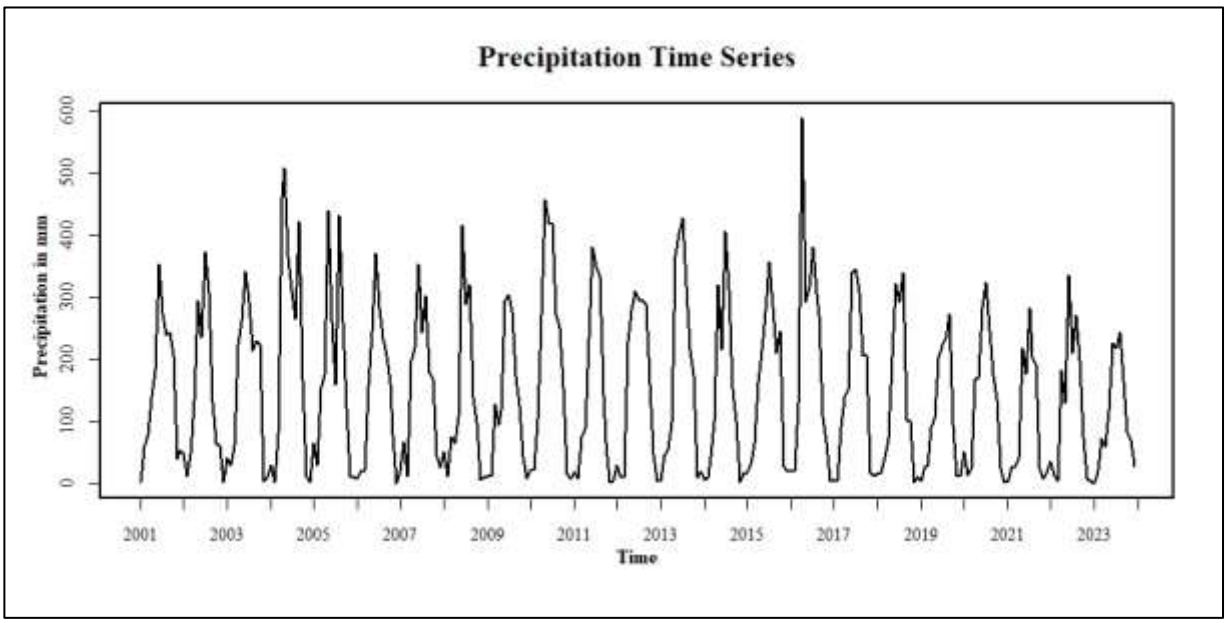

Fig. No. 2

On the contrary, the additive decomposition model was well-suited to this time series. The seasonal variations were nearly constant in magnitude, meaning the fluctuations did not scale directly with trend. This characteristic fits the assumptions of the additive model, wherein the seasonal and residual are assumed to be independent of the size of the trend. The decomposition results using the additive model successfully isolated the key components, which enabled a more detailed understanding of the underlying patterns in the precipitation data. To verify the additive decomposition of time series, further analysis was done after the additive decomposition. First, the seasonal component was compared against the trend to identify whether it changes proportionally with the trend's magnitude. A scatter plot of normalized seasonal and trend (Fig. No. 4) components revealed no discernible proportional relationship, which was further corroborated by a low correlation coefficient of 0.0276. This result shows that the seasonal variations are not significantly affected by changes in the trend level. Relationship between the random component and the trend was also assessed. Random component vs. trend scatter plot (Fig. No. 5) shows a random scatter without an apparent pattern, thus indicating independence between the two. This was corroborated by a low correlation coefficient of 0.0667, which indicated that the fluctuations of the random component are not influenced by the magnitude of the trend.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 10 No. 4, 2024
https://www.theaspd.com/ijes.php

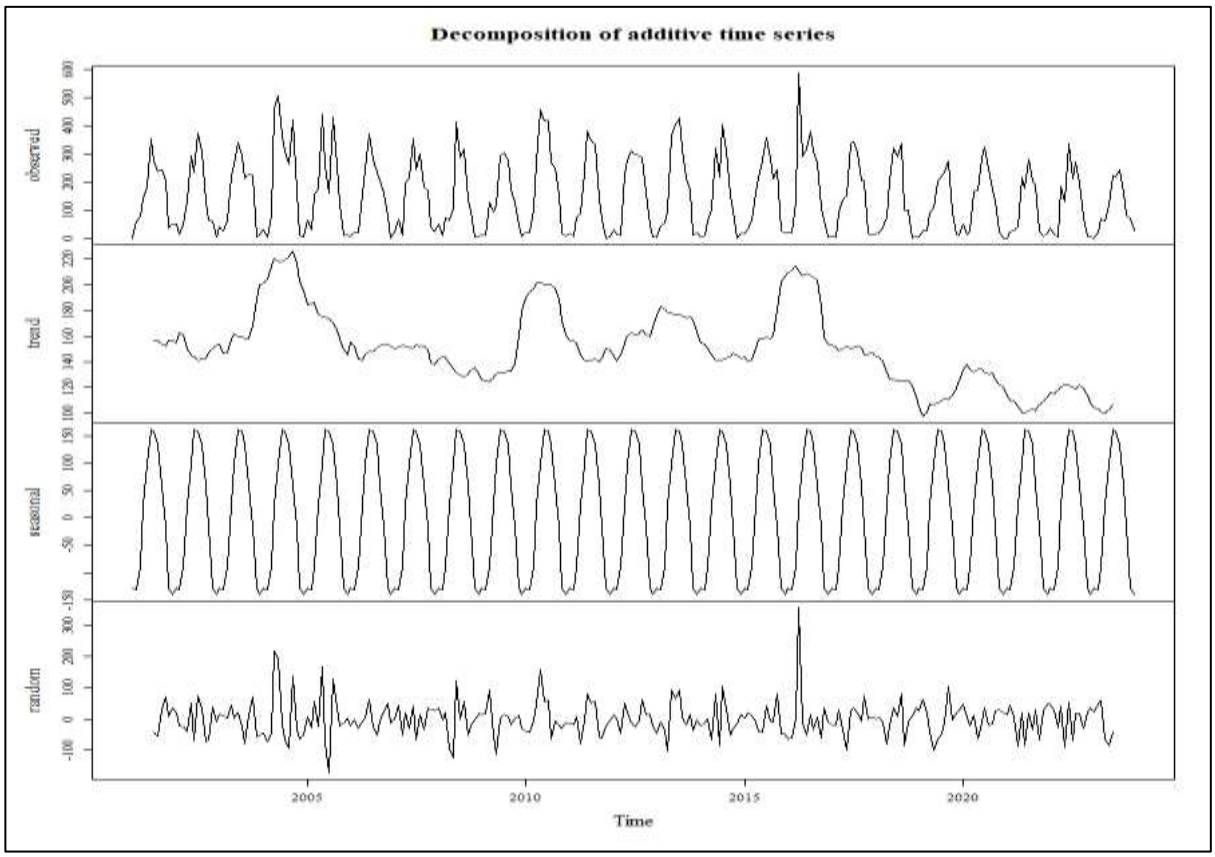

Fig. No. 3

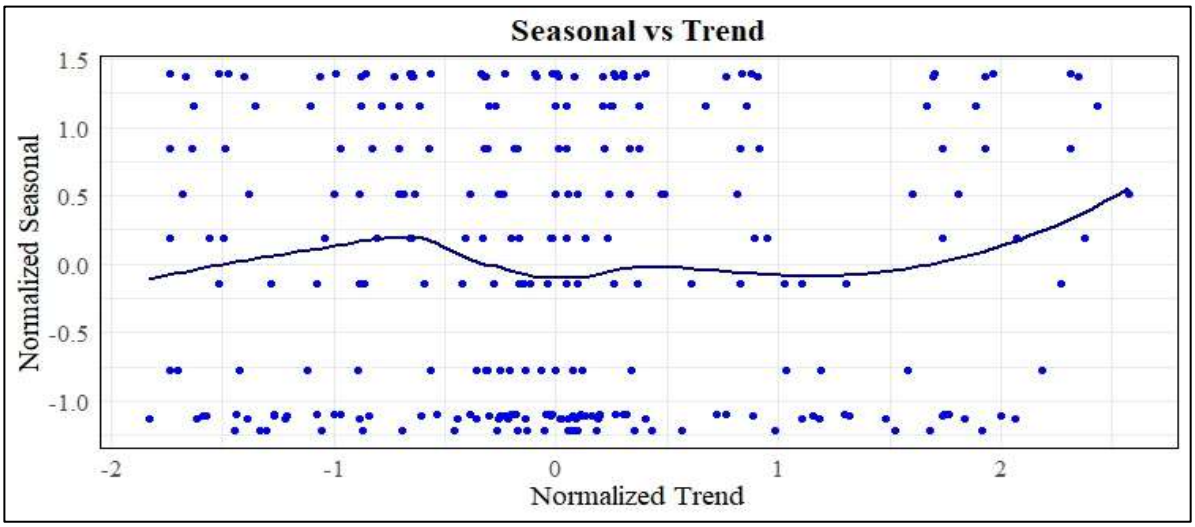

Fig. No. 4

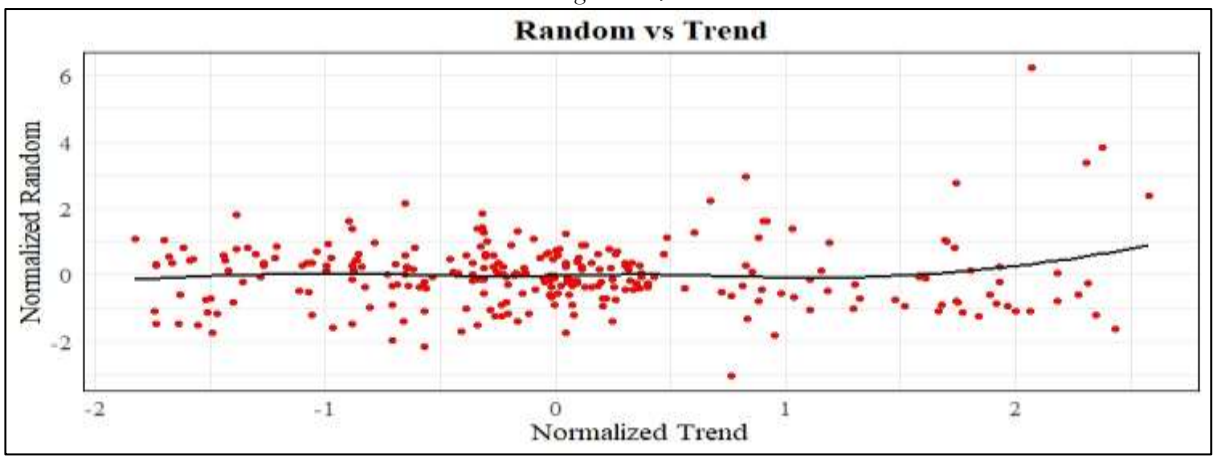

Fig. No. 5

From these results, the additive decomposition model applied here for this series is justified and valid. Again, the independent nature of seasonal and random with respect to trend satisfies the model's assumptions whereby seasonal variations in the additive model are all additive and coherent across various trends. The behavior of the random term as purely noisy noise also agrees with the idea of an additive approach. By employing the additive decomposition model, the decomposition is thus ensured to well capture the underlying structure of data, providing a reliable basis for further analysis and forecasting. Thus, by decomposing the data into trend, seasonal, and random components, we could identify both the consistent and irregular features of precipitation and their implications toward understanding the region's climatic behavior and the associated impacts. Figure No. 3 is the trend component and proves that long-term changes are taking place at subtle levels within precipitation over a period of 22 years. Looking closer into this, precipitation throughout the initial years of the data, especially from 2002-2010 was relatively stable, not rising or falling much (Fig. No. 3). Still, there's an apparent increase in total precipitation, specifically during the peak monsoon months, beginning around the 2010s. For example, the time series in Fig. No. 3 reports one of the highest annual precipitation levels for 2016, with the monthly totals more than 400 mm during July and August. It may reflect a year with strengthened monsoon activity, possibly by the favorable monsoonal circulation or global climatic conditions. Some years stand out as anomalies of low precipitation levels. Notably, rainfall declined sharply during 2009 with several months during the monsoon season, July and August receiving less than 200 mm—far below long-term monthly averages for the monsoon season (Fig. No. 3). The year is associated with an El Niño event that is known to suppress monsoon activity in India. Such anomalies highlight the susceptibility of Manipur's rainfall patterns to larger-scale climatic phenomena, underscoring the need for further exploration of their impacts. The seasonal component reveals a predictable pattern of monsoonal dependence, with precipitation peaking sharply between June and September. These months account for the majority of annual rainfall, with monthly totals often exceeding 300 mm during the monsoon. This peak is followed by a dry period from December to February, where monthly precipitation typically falls below 50 mm. Interestingly, the magnitude of these seasonal peaks appears to have increased slightly over the years, suggesting that monsoonal rains may have become more intense. As can be seen, for example, the time series shows that during July 2020, it was above 450 mm; this is among the wettest months recorded over the study period (Fig. No. 3). The random component gives insight into events that were beyond the trend or seasonality explanation. For example, in October 2018, there was an unseasonal increase in rainfall, with the monthly totals over 300 mm, which is more than double the average for October. This could be related to a late monsoonal withdrawal or a localized cyclonic event, which would require a closer look at the meteorological records for that month. Similarly, some years experienced long dry spells, for example, in the early months of 2012, when rainfall was significantly low month after month, probably affecting agricultural productivity. Identifying particular years and months with anomalies or trends enables us to relate these findings to broader climatic phenomena and regional implications. Recent years' enhancement of the intensity of monsoonal rainfall may be an early pointer to the shifting climate regime, with all this possibly being incited by some global warming effects on monsoonal circulation. On the other hand, anomalous years such as 2009 and 2018 tended to underscore the vulnerability of this region towards extreme weather events that are becoming more frequent in a warming world. These results have important implications for water resource management, agricultural planning, and disaster preparedness in Manipur. Increased monsoonal intensity is indicative of potential flooding risks, whereas dry years call for drought-resilient strategies. The policymakers and planners need to integrate these trends and anomalies into regional planning efforts to ensure that adaptive measures are in place to mitigate the impacts of both gradual and abrupt climatic shifts. More advanced forecasting systems, along with local studies, can be used to better refine our understanding of these patterns and their implications for the future. The detailed analysis combined with historical context shows the intricate interplay between natural variability and external climatic forces driving precipitation in Manipur. This has contributed to better understanding of regional climate dynamics, and the research findings form a basis for making informed decisions against the backdrop of changing environmental issues.

Augmented Dickey-Fuller (ADF) test results

Augmented Dickey-Fuller (ADF) test was carried out before applying the forecast models. In the alternative hypothesis, it is specified as "stationary". The Dickey-Fuller statistic obtained was -11.531, with a lag order of 6, and a p-value of 0.01. ADF test results are interpreted in terms of the p-value. In this scenario, the p-value was below the standard level of significance, 0.05, which rejects the null hypothesis. The data, therefore, is stationary with no need for transformation, hence retaining the raw time series for further analysis. If the p-value were higher than 0.05, then the series would be labeled non-stationary and in need of transformation for stationarity to be achieved. Such cases require first differencing, in which each observation would be subtracted from its

previous value. Such transformed data then will be used for further analyses. In this particular study, since the data was already determined to be stationary, no transformation of such a type was needed.

.

ARIMA Forecast Results

The ARIMA time series model used in this research is of the type ARIMA (5,0,1) mean nonzero. It captures time series data from Manipur where monthly levels of rainfall vary in mm during the period from January 2001 to December 2023. The model specification being p=5, d=0, and q=1 accounts for five autoregressive terms, no differencing, and one moving average term, which reflects the intricacy of the dataset. Due to the persistent bias of the mean level of the data from zero during exploratory analysis, the nonzero mean was required (μ=150.2935). In addition, other performance measures confirm the adequate fit of the model. The log-likelihood is -1583.62; along with the Akaike Information Criterion (AIC = 3183.24), corrected AIC (AICc = 3183.78), and Bayesian Information Criterion (BIC = 3212.21), the model is sufficiently fitted to the data. These measures confirm the appropriateness of the time series representation performed by the model.

The estimated coefficients really bring out the effectiveness with which the ARIMA model captures the underlying structure of the series. The estimated values of the autoregressive coefficients, ϕ_1 to ϕ_5, stand at 0.8700, -0.1166, -0.0922, -0.0822, and -0.2271, respectively. Their associated standard errors are in the range of 0.0691 and 0.1002. These indicate a strong dependence on the immediate past values and the gradual diminution over higher lags. The magnitude of the moving average coefficient (θ_1=-0.3771) and its standard error (0.0894) further reinforce the importance of considering prior error terms in explaining the observed fluctuations. Residual variance (σ^2 = 5738) indicates that the model is an appropriate representation of the data variability.

The predictive performance of the model is indicated by the error measures computed from the training data set of precipitation from January 2001 to December 2022. RMSE = 74.78424 and MAE = 52.86737 indicate moderate predictive accuracy, and MASE = 0.8906256 indicates that the model is better than a naïve benchmark model. The Mean Error (ME = -0.1010514) indicates that there is an almost negligible tendency to underpredict, and the autocorrelation of residuals at lag 1 (ACF1 = -0.01924824) confirms that the residuals are close to white noise, which is a critical assumption for model validity.

Substituting the estimated coefficients in the general ARIMA equation, the equation becomes:

$$y_t = 0.8700y_{t-1} - 0.1166y_{t-2} - 0.0922y_{t-3} - 0.0822y_{t-4} - 0.2271y_{t-5} - 0.3771\varepsilon_{t-1} + 150.2935 + \varepsilon_t$$

Here, y_t represents the observed time series at time t, and ε_t is the white noise error term. This equation reflects the relationship among the time series observations, their past values, and the error components, providing a reliable framework for forecasting and analysis.

It can thus be concluded that, the ARIMA (5,0,1) model with a nonzero mean effectively captures the structure and dynamics of the precipitation data. Its predictive performance and adherence to assumptions make it a dependable tool for forecasting and analysis, providing deeper understanding of the temporal patterns of precipitation in Manipur.

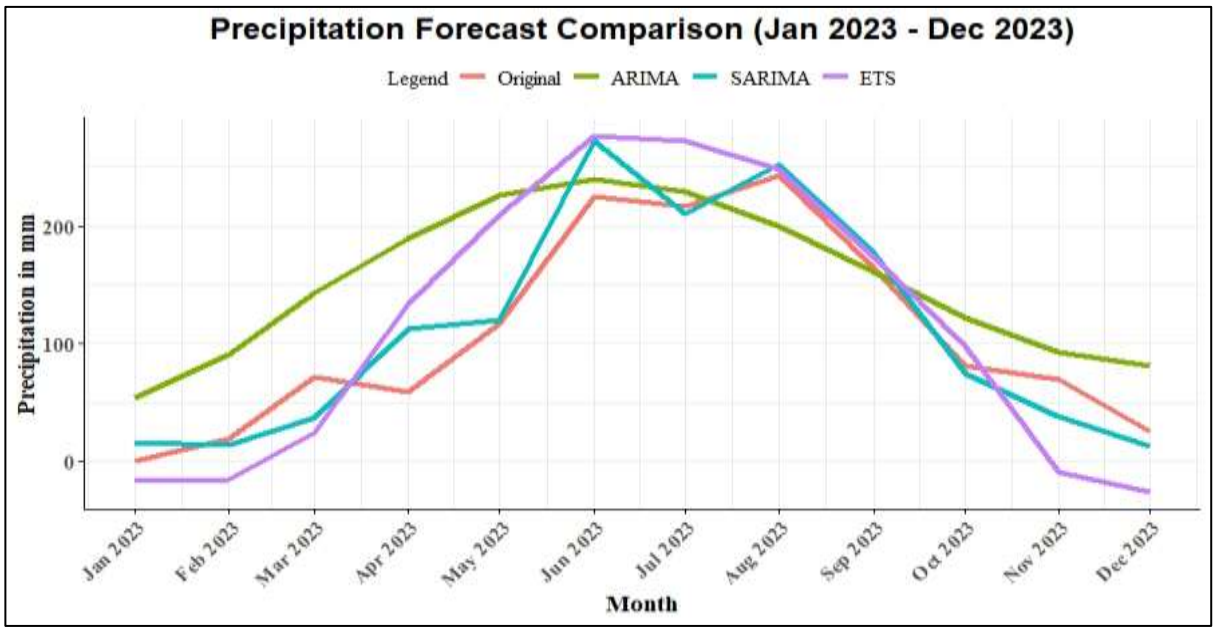

Fig. No. 6. Model accuracy

SARIMA forecast

The estimated coefficients for the model include an AR(1) term of 0.1619, a SAR(1) term of -0.4562, and a drift term of -0.1491. The standard errors for these coefficients are 0.0607, 0.0546, and 0.3265, respectively, indicating their statistical reliability. Substituting these values into the SARIMA equation, the model is expressed as:

$$(1 - 0.1619B)(1 - 0.4562B^{12})Y_t = \varepsilon_t - 0.1491$$

This equation captures both the non-seasonal dependency at lag 1 and the seasonal dependency at lag 12, effectively modelling the data's underlying structure.

The model's performance is assessed using several statistical criteria. It achieves a log-likelihood of -1521.5, with an Akaike Information Criterion (AIC) of 3051, corrected AIC (AICc) of 3051.15, and Bayesian Information Criterion (BIC) of 3065.3. These values indicate a reasonable balance between model complexity and goodness-of-fit, confirming the model's ability to capture the data's dynamics.

Further evaluation of the model is conducted using training set error measures. The Mean Error (ME) is -0.1540816, reflecting a slight under-prediction bias. The Root Mean Square Error (RMSE) of 74.9434 and the Mean Absolute Error (MAE) of 49.36739 highlight moderate predictive accuracy. However, the Mean Percentage Error (MPE) is -Infinity (-ve ∞), and the Mean Absolute Percentage Error (MAPE) is infinity (∞), due to the challenges posed by near-zero or zero actual values in the data. The Mean Absolute Scaled Error (MASE) is 0.8316635, suggesting that the model outperforms a naïve benchmark. Additionally, the first-order autocorrelation of residuals (ACF_1) is -0.009117953, indicating minimal residual autocorrelation.

It has thus been observed that, the SARIMA (1,0,0)(1,1,0)[12] model with drift effectively captures the seasonal and nonseasonal patterns in the time series, providing a robust framework for forecasting. The error measures and statistical criteria support the model's validity and its ability to generate reliable predictions.

ETS forecast

After fitting the ETS (A,N,A) model to the given data, the estimated parameters were determined as α=0.08 and γ=0.0001. The initial state values included l_0 = 157.8374 and the seasonal components for 12 periods as: s = {−139.7587, −122.9436, −15.8572, 59.6454, 133.8799, 157.7386, 162.4346, 94.8183, 19.3799, −89.5293, −129.5648, −130.2431}.

Substituting these values into the general equations, the model equations become:

$$y_t = l_{t-1} + s_{t-m} + \varepsilon_t$$
$$l_t = l_{t-1} + 0.08\varepsilon_t$$
$$s_t = s_{t-m} + 0.0001\varepsilon_t$$

The estimated error variance (σ^2) of the model was found to be σ^2= 64.2174. Model selection criteria yielded AIC = 3864.439, AICc =3866.285 and BIC = 3918.745, confirming the model's adequacy for representing the underlying time series data.

Additionally, error metrics from the training set were examined to assess the model's forecasting performance. The Mean Error (ME) was -1.977615, and the Root Mean Squared Error (RMSE) was calculated as 62.56755. The Mean Absolute Error (MAE) was found to be 42.77073, suggesting moderate accuracy. However, instability in forecasting was evident in the Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE), which were NaN (not a number) and infinity respectively due to zero precipitation values. The Mean Absolute Scaled Error (MASE) was 0.7205335, indicating that the model outperformed a naïve forecast. The autocorrelation of the residuals at lag 1 (ACF_1) was 0.1385229, suggesting mild correlation in residuals.

Comparative assessment of Models in terms of Accuracy Metrics

In time series forecasting, accuracy metrics are essential to evaluate and compare the predictive performance of different models. Each metric provides unique understanding of how well a model captures the underlying patterns of the data and how reliable its forecasts are. The explanation of the metrics used in this study and their significance are given below:

Root Mean Square Error (RMSE)

The RMSE measures the square root of the average squared differences between predicted and observed values. It is highly sensitive to large errors due to the squaring of residuals, making it an ideal metric when large deviations from actual values are especially undesirable. A lower RMSE value indicates better predictive performance. In this study, SARIMA's RMSE of 74.9434, comparable to ARIMA's RMSE of 74.78424, suggests moderate accuracy but reflects the model's ability to handle larger deviations more effectively than ETS in some scenarios.

Mean Absolute Error (MAE)

MAE is the average of the absolute differences between predicted and observed values, offering a straightforward measure of prediction accuracy. Unlike RMSE, it treats all deviations equally and does not disproportionately penalize larger errors. The lower MAE value of SARIMA (49.36739) compared to ARIMA (52.86737) signifies

that SARIMA is more accurate on average across all data points, while ETS shows the lowest MAE (42.77073), indicating that it excels in reducing the average prediction error.

Mean Error (ME)

ME quantifies the average bias in the predictions. A value close to zero implies that the model does not systematically over-predict or under-predict the actual values. For SARIMA, the ME of -0.1540816 indicates a slight tendency toward under-prediction, but this bias is minimal and unlikely to significantly affect its reliability. In contrast, ETS has a larger ME (-1.977615), suggesting a stronger bias in its forecasts, which might reduce its overall applicability in some scenarios. The Mean Error (ME = -0.1010514) of ARIMA suggests a minimal bias toward under-prediction.

Mean Absolute Scaled Error (MASE)

MASE compares the model's predictive performance to that of a naïve benchmark model, which assumes that future values will be the same as the most recent observed values. A MASE value less than 1 implies that the model outperforms the naïve approach. SARIMA's MASE of 0.8316635 demonstrates that it is superior to a naïve forecast and is slightly better than ARIMA (0.8906256). ETS, with the lowest MASE (0.7205335), outperforms both ARIMA and SARIMA in this respect, though other metrics highlight limitations in its performance.

Autocorrelation of Residuals at Lag 1 (ACF_1)

ACF_1 measures the correlation between residuals and their immediate lag. Ideally, residuals should exhibit no significant autocorrelation, meaning they resemble white noise, as this indicates that the model has captured all patterns in the data. SARIMA's ACF_1 value of -0.009117953 confirms near-zero autocorrelation, suggesting that the residuals are random and the model sufficiently captures the underlying structure. In comparison, ARIMA also demonstrates minimal autocorrelation (ACF_1 = -0.01924824), while ETS exhibits mild residual correlation (ACF_1 = 0.1385229), indicating potential for improvement in accounting for the underlying patterns.

Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE)

MPE and MAPE measure forecast accuracy as a percentage of observed values. These metrics become problematic when the dataset contains near-zero or zero values, leading to undefined or infinite results. This limitation is evident in the SARIMA and ETS models for this dataset. Despite their absence in this analysis, the consistent performance of other metrics, such as RMSE and MAE, ensures reliable evaluation of the models.

Significance of Metrics in the context of this Study lies in the fact that each of them provides a unique perspective of each model's performance. RMSE and MAE are crucial for assessing overall prediction accuracy, particularly in cases where, avoiding large deviations or understanding average performance is important. SARIMA's competitive RMSE and low MAE indicate its reliability in capturing trends. ME is critical for identifying and correcting systematic biases. SARIMA's minimal bias suggests reliable model calibration. MASE highlights the practical advantage of SARIMA over a naïve forecast, reinforcing its utility in real-world applications. ACF_1 ensures the statistical validity of the residuals, which is vital for ensuring that the model has effectively captured the data's structure. These metrics collectively highlights SARIMA's superior performance, balancing predictive accuracy, minimal bias, and residual randomness. This comprehensive study provides confidence in SARIMA's capability to handle the complex seasonal patterns of precipitation data effectively.

CONCLUSION

The research has been conducted by determining how appropriate the models of ARIMA, SARIMA, and ETS are for the state of Manipur in India through monthly precipitation on high-resolution satellite data for forecasting purposes. In fact, some general strengths and weaknesses of the given models are ascertained. The model of ARIMA (5,0,1) captured all short-term dependencies perfectly and strongly but failed to explain the phenomenon of high-seasonality-related factors in the data; the best performing was SARIMA (1,0,0)(1,1,0)[12], explaining all seasonal pattern and gave very reliable points of forecasts. ETS (A, N, A) has well captured additive seasonality but showed residual autocorrelation to be rather low. Due to this, smoothing of the data's irregularities cannot be very possible. Decomposition additive also made it much more effective to define the trends, cyclical, and anomalous trends that characterize this data set. The global climatic event known as El Niño during the recent years, is responsible for the strong magnitude of monsoon rainfall and the most significant anomalies. Seasonal analysis brought forth the role of monsoon-driven rain, and in all series, peaks appeared predictable between June and September; however, drier months could be seen in others. Results present the adequacy of the use of SARIMA in the modeling and precipitation forecasting of such complexly seasonal regimes. Also, time-series analysis of information about precipitation helps bring out some crucial inputs for climatic behavior of Manipur. Such subtle long-term trends are showing enhanced strength of monsoons whereas anomaly types like the 2009 drought and unseasonal rainfall in 2018 are suggested to be vulnerable to extreme

events. The conclusions of such results indicate a need for adaptive strategies in water management, agriculture, and disaster preparedness. Regional planning enables policymakers better to predict climatic variability and its impacts by the inclusion of reliable forecasting models. This research further supports the necessity to refine further more the techniques used for forecasting purposes and their further integration with localized studies meant for more excellent predictive accuracy regarding the sustainability-based strategies under changing climatic conditions.

REFERENCES

1. Alexander, S., & Block, P. (2022). Integration of seasonal precipitation forecast information into local-level agricultural decision-making using an agent-based model to support community adaptation. Climate Risk Management, 36, 100417. https://doi.org/10.1016/j.crm.2022.100417
2. Gorooh, V. A., Shearer, E. J., Nguyen, P., Hsu, K., Sorooshian, S., Cannon, F., & Ralph, M. (2022). Performance of New Near-Real-Time PERSIANN Product (PDIR-Now) for Atmospheric River Events over the Russian River Basin, California. Journal of Hydrometeorology, 23(12), 1899–1911. https://doi.org/10.1175/jhm-d-22-0066.1
3. Hong, Y., Hsu, K., Sorooshian, S., & Gao, X. (2004). Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. Journal of Applied Meteorology, 43(12), 1834–1853. https://doi.org/10.1175/jam2173.1
4. Lai, Y., & Dzombak, D. A. (2020). Use of the Autoregressive Integrated Moving Average (ARIMA) model to forecast Near-Term regional temperature and precipitation. Weather and Forecasting, 35(3), 959–976. https://doi.org/10.1175/waf-d-19-0158.1
5. Liu, K., Wang, Q., Wang, M., & Koks, E. E. (2023). Global transportation infrastructure exposure to the change of precipitation in a warmer world. Nature Communications, 14(1). https://doi.org/10.1038/s41467-023-38203-3
6. Lowe, D., Ebi, K., & Forsberg, B. (2013). Factors Increasing Vulnerability to Health Effects before, during and after Floods. International Journal of Environmental Research and Public Health, 10(12), 7015–7067. https://doi.org/10.3390/ijerph10127015
7. Mathew, M. M., K, S., Mathew, M., Arulbalaji, P., & Padmalal, D. (2021). Spatiotemporal variability of rainfall and its effect on hydrological regime in a tropical monsoon-dominated domain of Western Ghats, India. Journal of Hydrology Regional Studies, 36, 100861. https://doi.org/10.1016/j.ejrh.2021.100861
8. Narayanan, P., Basistha, A., Sarkar, S., & Kamna, S. (2013). Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. Comptes Rendus Géoscience, 345(1), 22–27. https://doi.org/10.1016/j.crte.2012.12.001
9. Nguyen, P., Ombadi, M., Gorooh, V. A., Shearer, E. J., Sadeghi, M., Sorooshian, S., Hsu, K., Bolvin, D., & Ralph, M. F. (2020b). PERSIANN Dynamic Infrared–Rain Rate (PDIR-Now): a Near-Real-Time, Quasi-Global satellite precipitation dataset. Journal of Hydrometeorology, 21(12), 2893–2906. https://doi.org/10.1175/jhm-d-20-0177.1
10. Ning, T., Feng, Q., Li, Z., Li, Z., Xi, H., Yang, L., & Chang, X. (2023). Precipitation changes and its interaction with terrestrial water storage determine water yield variability in the world's water towers. The Science of the Total Environment, 880, 163285. https://doi.org/10.1016/j.scitotenv.2023.163285
11. Wang, J., You, Z., Song, P., & Fang, Z. (2024). Rainfall's impact on agricultural production and government poverty reduction efficiency in China. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-59282-2