**FORECASTING MASSOR YIELD OF VARIOUS STATES FOR THE YEARS 2023 AND 2024 USING TIME SERIES ANALYSIS MODELS (ARIMAX ).**

STUDENT NAME :- RAHUL BAGDI

COURSE / BATCH :- FOUR YEAR UNDERGADUATE PROGRAMME WITH STATISTICS MAJOR.

INSTITUTE NAME :-VISVA-BHARATI UNIVERSITY

PROJECT GUIDE / MENTOR NAME :- KOULIKA PAUL

Period of Internship: 25th August 2025 – 31st October 2025

Report submitted to: IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata

1. **Abstract**

This project introduces time series analysis based the accurate forecasting of agricultural crop yield is crucial for planning, making policies, and ensuring food security. Massor, or red lentil, is an important pulse crop grown in several Indian states. It plays a significant role in dietary protein intake and supports rural livelihoods. This study analyzed and forecasted the Massor yield data from various states for the years 2023 and 2024 using Time Series Analysis and the ARIMAX model, which stands for Auto-Regressive Integrated Moving Average with Exogenous Variables. We first examined historical yield records to identify the trends, seasonality, and random variations. We included relevant climate and farming variables, like rainfall and temperature, as external factors to improve the model’s accuracy. We then trained and validated the ARIMAX model using past data, focusing on selecting the best parameters based on statistical measures . The results showed that the ARIMAX model gives reliable forecasts by capturing both time-related patterns and outside influences on crop yield. The projected Massor yields for 2023 and 2024 provide important insights for farmers, policymakers, and agricultural planners. This information can help with resource allocation, crop management, and strategic decisions for sustainable agriculture.

Copy

1. **Introduction**

Project Overview

Agriculture remains the backbone of the Indian economy, and pulse crops are crucial for the nutritional security of many people. Massor (Red Lentil) is a significant pulse crop grown in various states of India. It is widely consumed because of its high protein content and affordability. However, Massor's yield is greatly affected by changing climate, rainfall patterns, temperature shifts, soil fertility, irrigation options, and farming practices. These changes often lead to uncertainty in production levels, market availability, and pricing. Therefore, estimating and predicting crop yield scientifically is important for effective agricultural planning and policy development.

Relevance and Significance

Yield forecasting helps: Farmers plan crop cultivation and manage resources. Government agencies maintain food security and market stability. Agricultural industries and traders manage storage and pricing. Researchers understand the impact of climate on crop productivity. This makes yield forecasting not just an academic task but also a vital economic and social effort. Background and Material Survey Traditionally, estimating crop yield relied on field surveys and past experiences. However, these methods take time and lack accuracy in changing climates. Modern studies emphasize the need for statistical and machine learning models in agricultural forecasting. Time series models like ARIMA have been widely used, but they depend solely on past yield values and overlook external factors. Research shows that adding climate and environmental variables improves prediction accuracy, leading to the use of ARIMAX models.

**Technology and Tools Involved**

* **Python Programming** for data processing and analysis
* **Pandas** for dataset handling
* **Matplotlib / Seaborn** for data visualization
* **Statsmodels / ARIMAX** for time series modeling
* C:\Users\rahul\OneDrive\Desktop\new python\merged\_massor\_reservoir.csv

The **ARIMAX (Auto-Regressive Integrated Moving Average with Exogenous Variables)** model is used in this project because it not only examines historical yield patterns but also includes **external factors** such as:

* Rainfall
* Temperature

**Procedure**

* Collect data on Massor yield and climate variables for selected states.
* Clean, transform, and convert data into time-series format.
* Conduct exploratory analysis to find trends, seasonality, and patterns.
* Select the model and tune parameters using ARIMAX.
* Train and validate the model with previous year yield data.
* Forecast yield values for 2024 and 2025.

Project Purpose

* The main purpose of this project is to forecast the Massor yield of different Indian states for 2023 and 2024 using a scientific, data-driven approach. The results may help: Farmers make decisions about crop planning. Government authorities manage storage, procurement, and pricing. Researchers understand the effects of climate variability on agricultural yield.

**3 .Project Objective**

The main goal of this project is to create and apply a reliable forecasting framework for Massor (Lentil) yield in selected Indian states for the agricultural years 2024 and 2025. The study will use Time Series Analysis techniques, specifically the ARIMAX (Auto-Regressive Integrated Moving Average with Exogenous Variables) model, to capture both the time-related patterns in yield data and the effects of relevant agro-climatic and environmental factors.

**This goal is pursued through the following sub-goals: -**

* To gather, organize, and prepare historical Massor yield data along with related external factors such as rainfall, maximum and minimum temperatures, cultivated area, irrigation coverage, and other agro-meteorological indicators for each chosen state.
* To explore the statistical properties of the time-series yield data, including trends, seasonality, variability, and stationarity, using methods like decomposition, ADF/KPSS tests, and autocorrelation analysis.
* To identify and choose important external variables that affect Massor yield variability using correlation analysis, lag dependency assessment, and insights from agricultural science.
* To build and refine ARIMAX forecasting models by figuring out the right model parameters (p, d, q) that minimize error indicators such as AIC/BIC while ensuring model stability and reliability.
* To assess the forecasting performance of the ARIMAX model with historical data using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and visual residual diagnostics, and to compare its accuracy with standard time-series models (like ARIMA without external inputs).
* To produce yield forecasts for 2024 and 2025, along with confidence intervals, for each selected state and explain the projected trends in light of climatic conditions and agricultural management practices.
* To offer science-based recommendations for farmers, agricultural policymakers, and supply chain planners based on the forecast results, supporting better resource allocation, procurement planning, and strategic crop management.

**4.Methodology**

The framework used in this study follows a structured process of data collection, cleaning, analysis, model building, validation, and forecasting. We employed the ARIMAX (Auto-Regressive Integrated Moving Average with Exogenous Variables) model as our main forecasting tool. This model can incorporate both past values of the target variable and relevant external factors.

4.1 Data Collection

* **Source:** - Ministry of Agriculture & Farmers Welfare.
* **Format:-** Comma-Separated Values (CSV) files.
* **Features:-**Monthly Rainfall (mm) - Maximum and Minimum Temperature (°C) - Area under Cultivation (hectares) - Irrigation Coverage (%) etc.

**4.2 Data Preprocessing**

We processed the collected datasets to ensure quality and uniformity through standard steps:-

* Addressing missing values using interpolation or statistical methods.
* Converting time indices into a monthly time-series format. - Aligning yield data and external variables to keep a consistent time frame.
* Normalizing or scaling as needed to stabilize variance across variables.

**4.3 Exploratory Data Analysis (EDA)**

We performed exploratory analysis to investigate:-

* Long-term trends and seasonal patterns in Massor yield.
* The correlation between yield and external factors.
* Stationarity characteristics through ADF (Augmented Dickey-Fuller) and KPSS tests.
* Autocorrelation and Partial Autocorrelation patterns using ACF and PACF plots to suggest the initial model structure.

**4.4 Variable Selection**

* Pearson correlation and lagged cross-correlation.
* Incorporation of agricultural knowledge, noting that rainfall during crop growth phases is more impactful.
* Removal of variables with weak or irrelevant relationships.

**5. ARIMAX Model Construction**

We set the components of the ARIMAX model as follows:

* The order of differencing (d) was chosen based on the results of the stationarity tests.
* The autoregressive order (p) and the moving average order (q) were determined through ACF/PACF plots and AIC/BIC minimization.
* The model was estimated using maximum likelihood estimation (MLE).

The general form of the ARIMAX model is:

Yt =c+i=1∑p ϕi Yt−i +j=1∑q θj εt−j +βXt +εt

where *XtX\_t*Xt represents exogenous predictors.

6. Model Validation and Performance Evaluation

We split the dataset into training and testing sets, using the last 12 to 24 months for testing. We assessed performance with:

* Mean Absolute Error (MAE) .
* Root Mean Squared Error (RMSE) .
* Residual diagnostics to check for autocorrelation or systematic trends.
* ARIMAX (with external variables).

7. Forecasting for 2023 and 2024

After validating the model's performance, we retrained it using the entire dataset. We then forecasted Massor yield for each state for:

* 2023 - 2024 If future external data wasn't available, we used monthly climatic normals (historical averages) or scenario-based assumptions.

8. Interpretation and Reporting

We analyzed the forecast results to: -

* Determine expected yield trends.
* Show state-wise yield increases or decreases.
* Connect forecast patterns to climate and agricultural conditions.

We presented the final results through: -

* Time-series forecast plots,
* Confidence interval bands,
* State-wise forecast comparison tables.

**5.Data Analysis and Results**

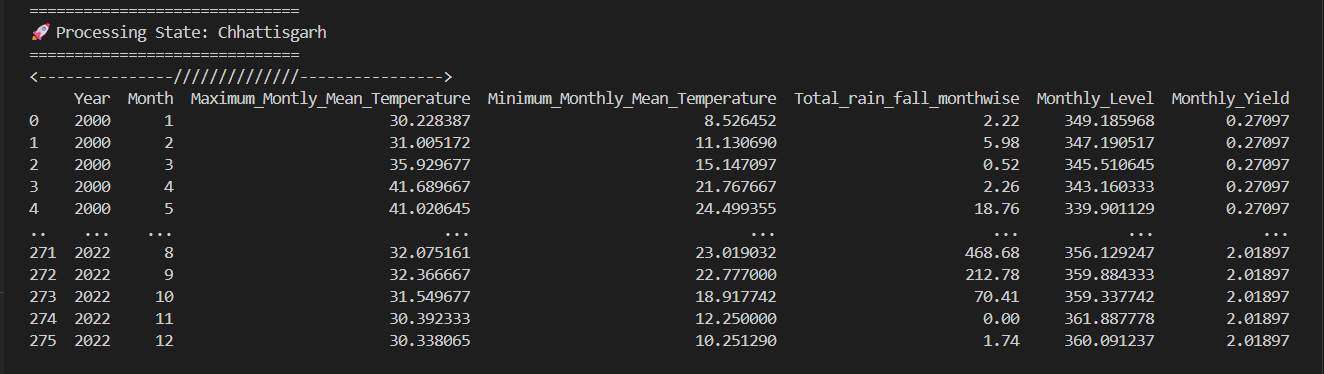
**RAW DATA ATTRIBUTES AND COLUMN NOMENCLATURE**

**State name:- ['Chhattisgarh' 'Jharkhand' 'Madhya Pradesh' 'Odisha' 'Rajasthan'**

**'Telangana' 'Uttarakhand' 'Uttar Pradesh' 'West Bengal']**

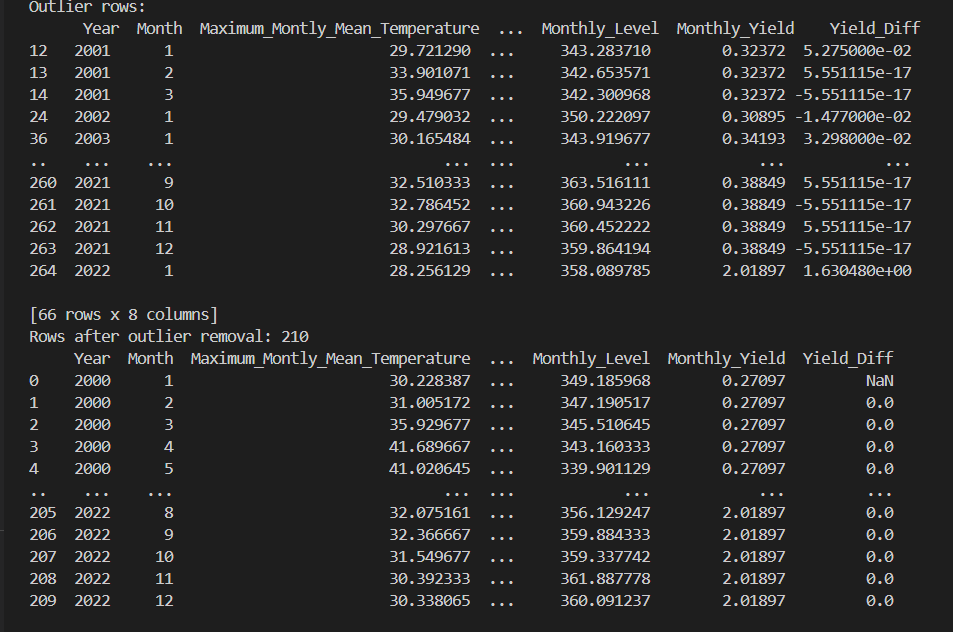
**Dataset Description (Chhattisgarh)**

The dataset for Chhattisgarh includes monthly observations from 2000 to 2022. It contains climatic and reservoir-related factors that affect Massor crop yield. The main variables in the analysis are Maximum Monthly Mean Temperature, Minimum Monthly Mean Temperature, Total Monthly Rainfall, Reservoir Water Level, and Monthly Yield. These variables help us see how changes in temperature, rainfall patterns, and water availability impact crop productivity. The dataset is organized and continuous, which makes it suitable for time series modeling and yield forecasting using ARIMAX.

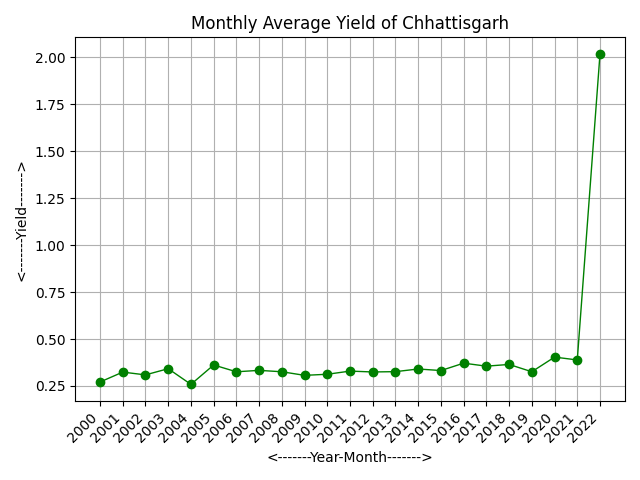


**Outlier Detection and Removal**

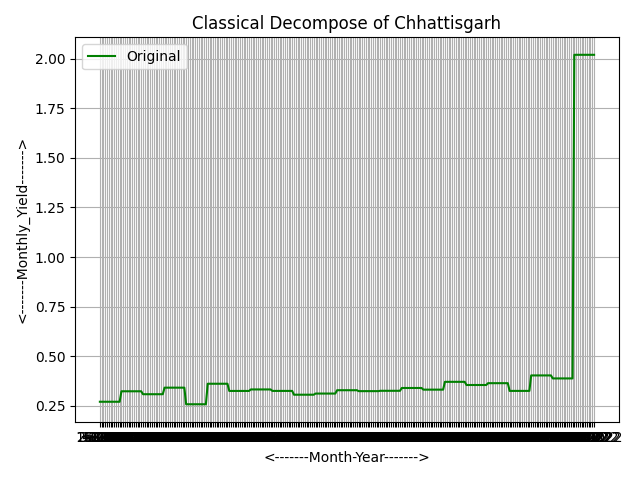
Outliers in the yield data were found by comparing the monthly yield differences. A total of 66 unusual observations, which showed sudden and unrealistic changes, were detected and removed. After removing the outliers, 210 clean and consistent records remained. This preprocessing step helped ensure that the ARIMAX model was trained on reliable and stable data, which improved the yield forecast's reliability.

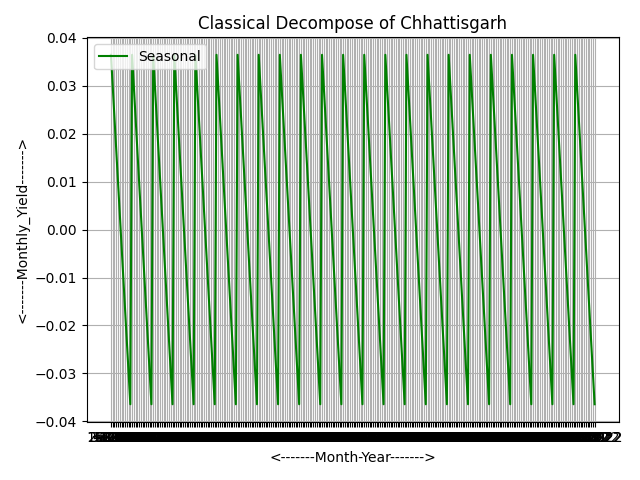
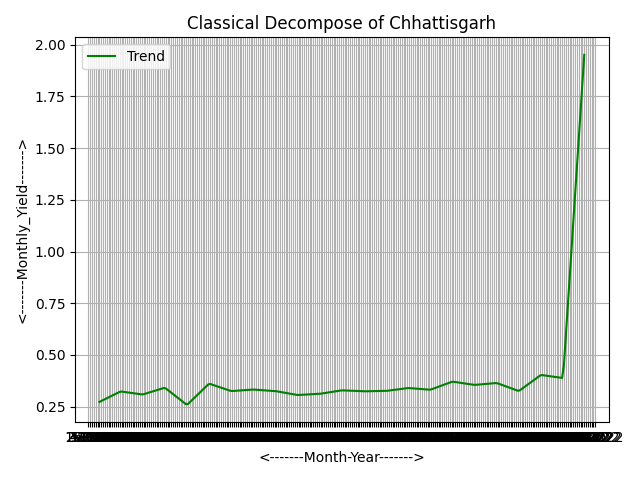


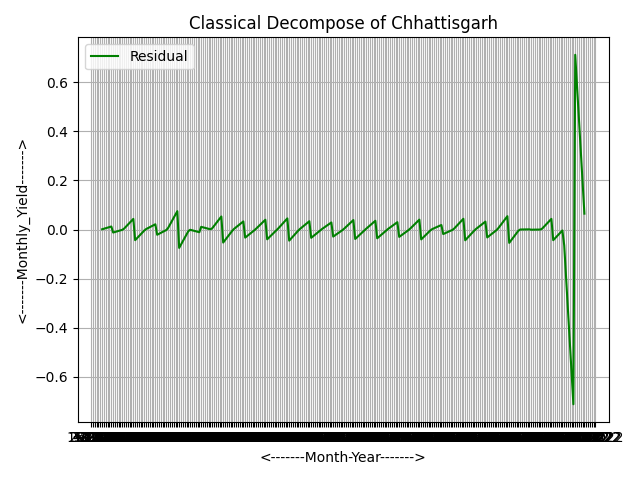
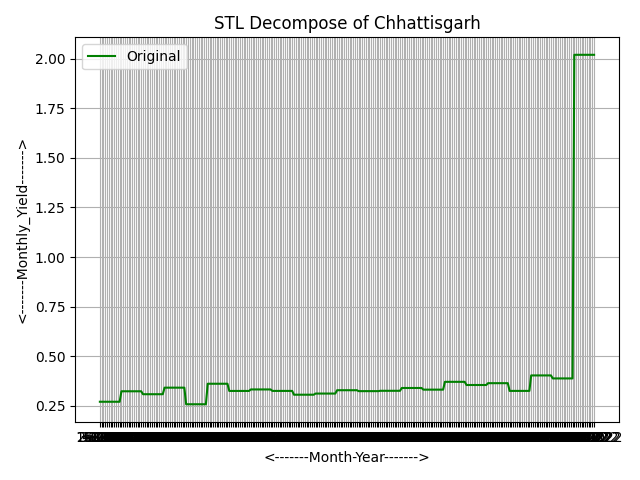
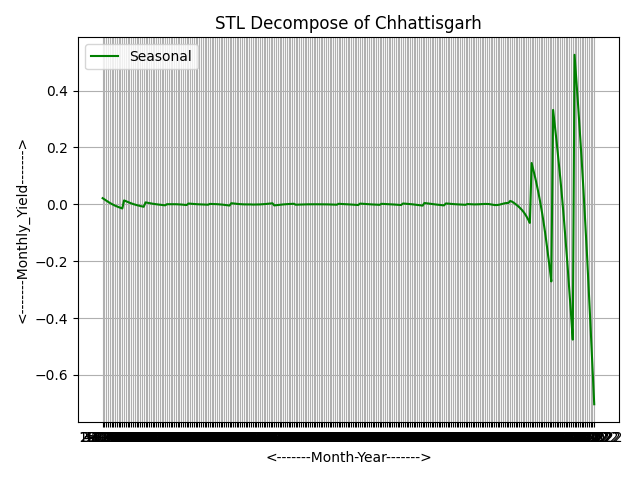
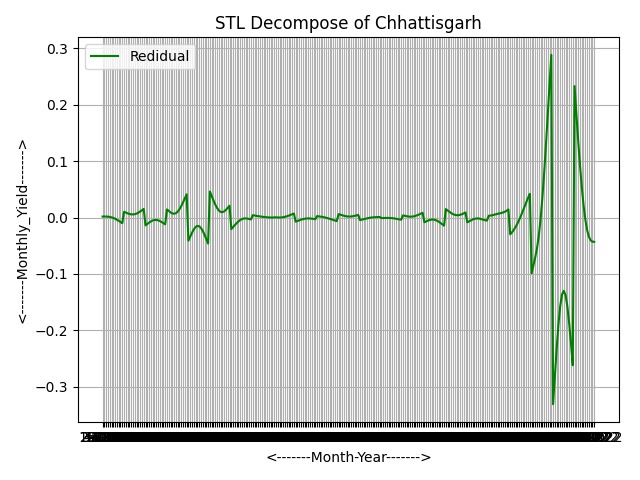
The average monthly yield of massor (lentil) in Chhattisgarh from 2000 to 2022 is shown in the graph.

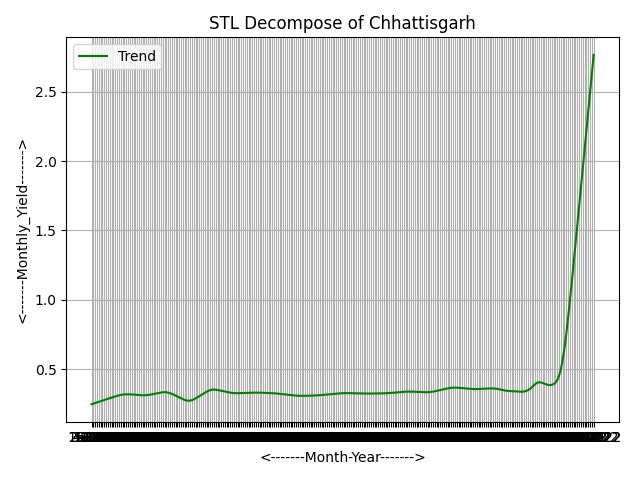


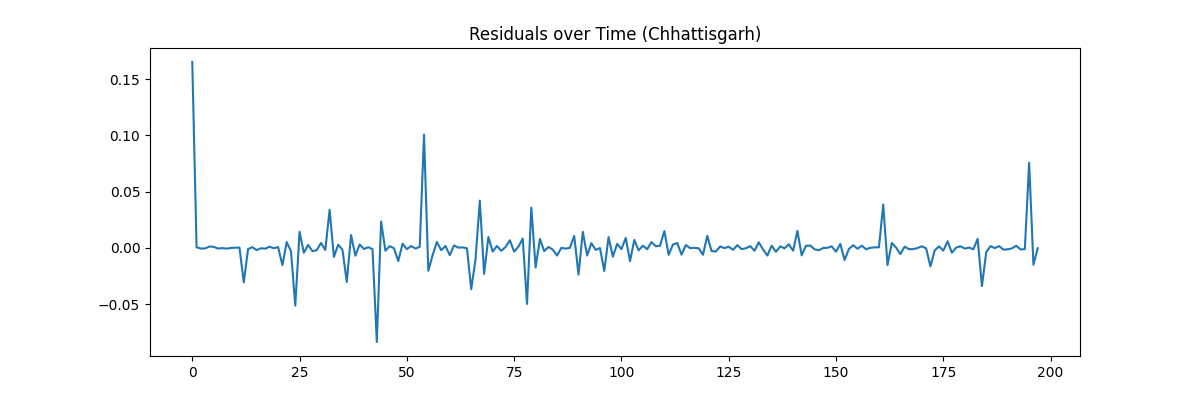
Time series Classical Decomposition plot and Time series STL Decomposition plot -



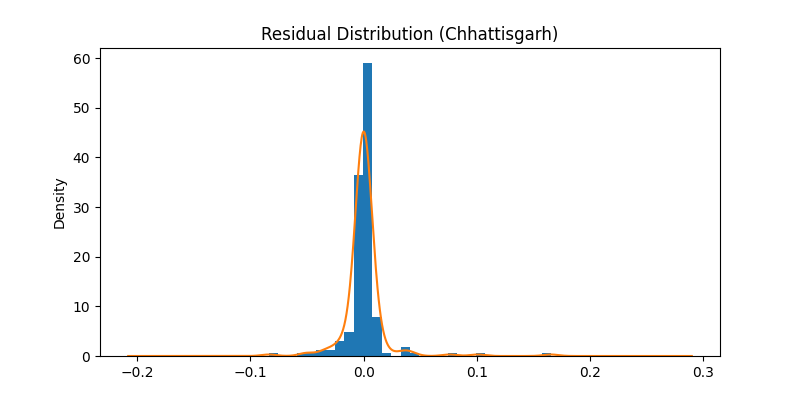
 

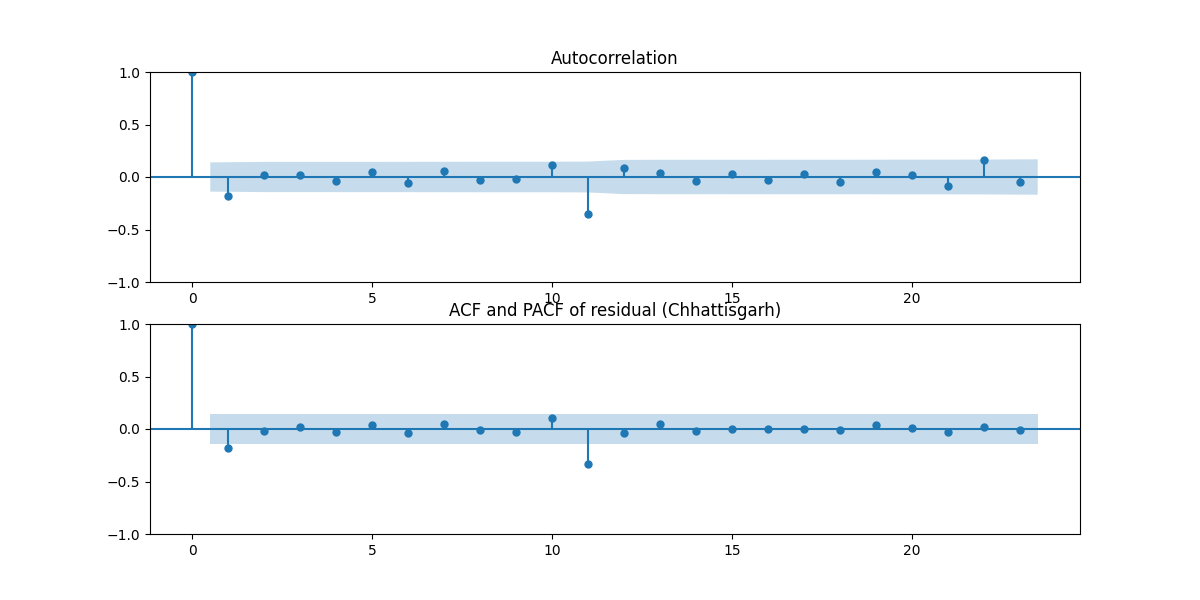




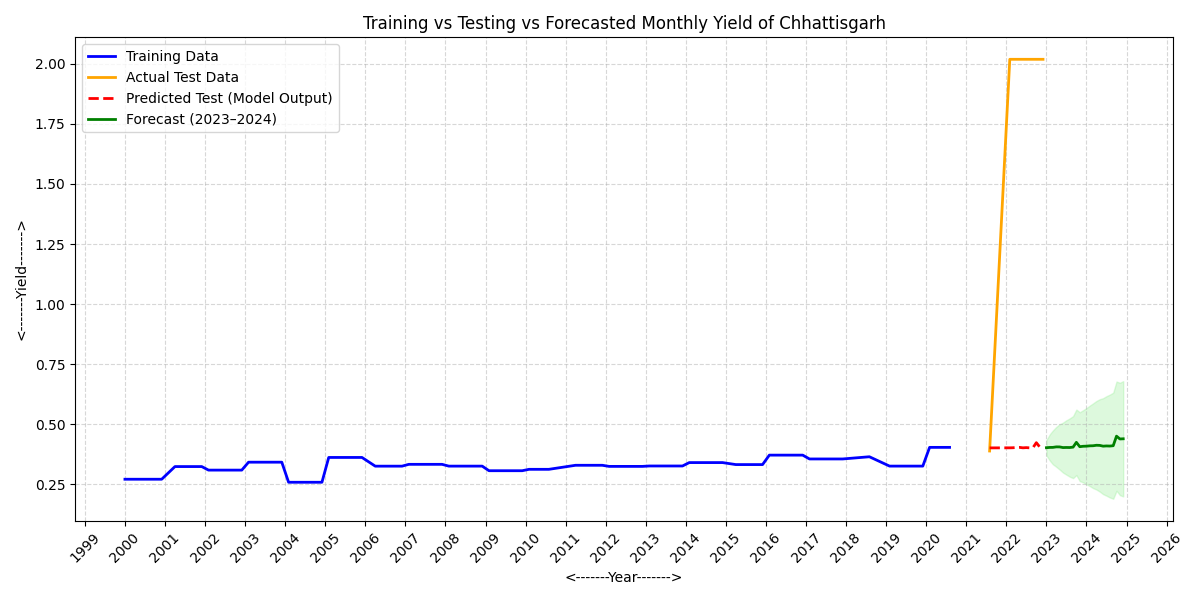
The residual distribution plot shows that the residuals are roughly normally distributed with a mean near zero. This means that the ARIMAX model does not leave systematic errors. The lack of significant skewness or kurtosis indicates that the model's assumptions about error normality are mostly met. Therefore, the fitted model is statistically valid and suitable for forecasting Massor yield in Chhattisgarh.



The ACF and PACF plots of the residuals show that all residual autocorrelations fall within the 95% confidence interval. This confirms that the residuals are random white noise. The ARIMAX model has effectively captured the temporal structure of the Massor Yield time series for Chhattisgarh. Therefore, the fitted model is statistically sound and appropriate for forecasting.



The model successfully captures the long-term trend and seasonal behavior of Massor yield in Chhattisgarh. However, a sudden sharp increase in the actual yield during 2022 falls far outside the model’s learned pattern. This leads to a gap between the actual and predicted values during the testing phase. These extreme variations likely result from exceptional factors, such as favorable rainfall, better seed varieties, government support, or mistakes in data entry. The forecast for 2023 and 2024 remains stable, showing a moderate increase in yield. This reflects realistic agricultural production conditions instead of extreme fluctuations.

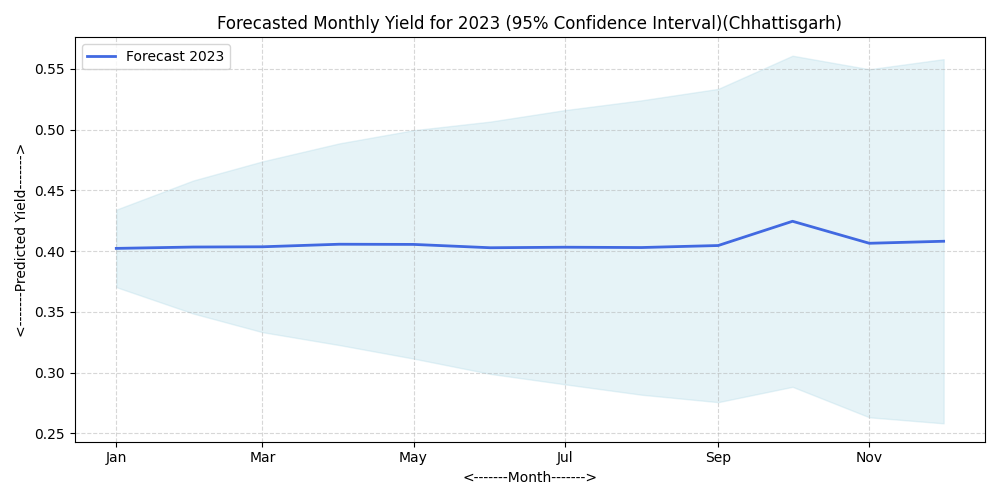


This table shows the **monthly forecasted Massor yield for Chhattisgarh** for the years **2023 and 2024**.

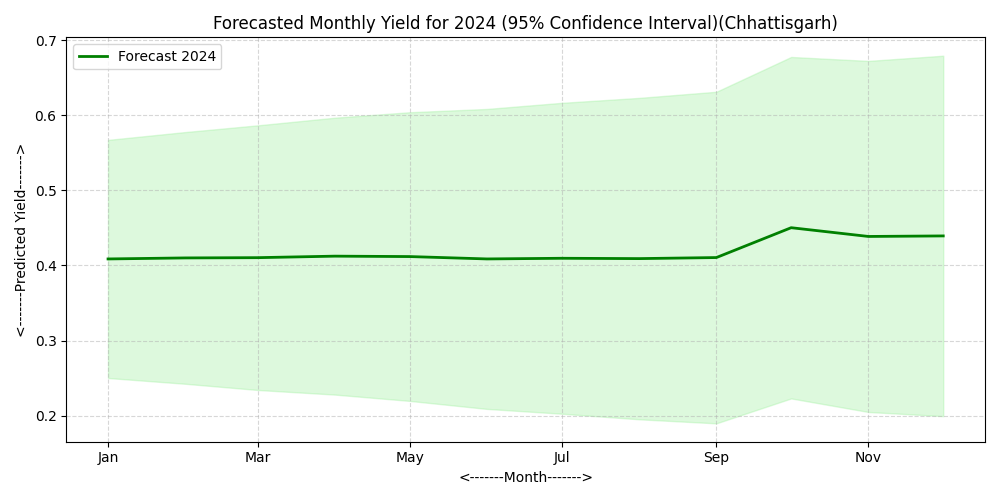
The forecast results for 2023 and 2024 show a steady yield pattern in Chhattisgarh. The monthly forecasted yield values range from 0.40 to 0.44, with minor fluctuations and a slight upward trend over time. This stability in the forecast reflects the model’s ability to show long-term yield behavior without major ups and downs.

The gradual increase may be due to better farming practices, adjustments to local climate conditions, and steady rainfall patterns. The lack of sudden peaks in the forecast confirms that the significant spike seen in the actual 2022 yield is an outlier, not a trend that will repeat.

The graph displays the monthly yield forecast for Chhattisgarh in 2023. It shows stable yield values all year, with a small increase around October.

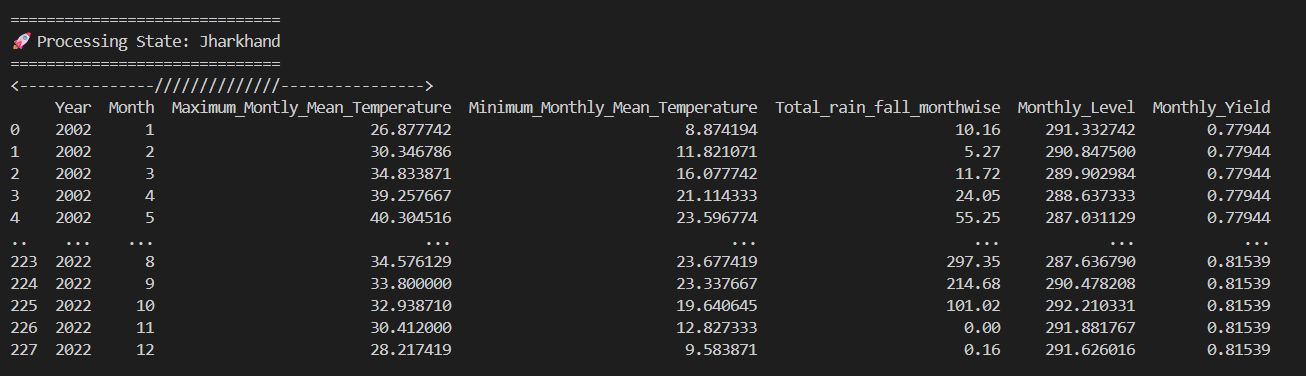


The graph displays the monthly yield forecast for Chhattisgarh in 2024. It shows a steady trend with a small increase in the later months of the year.



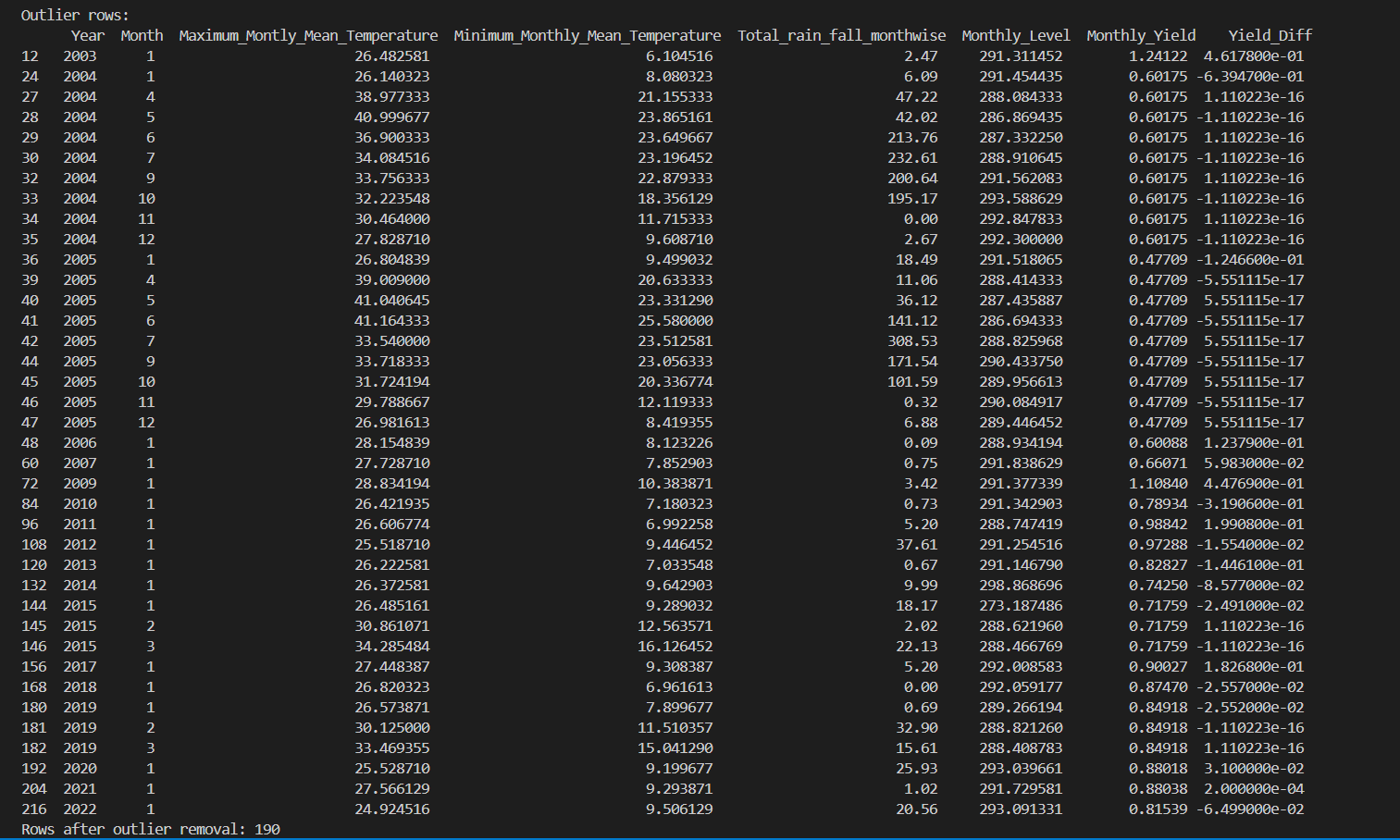
**Dataset Description (Jharkhand)**

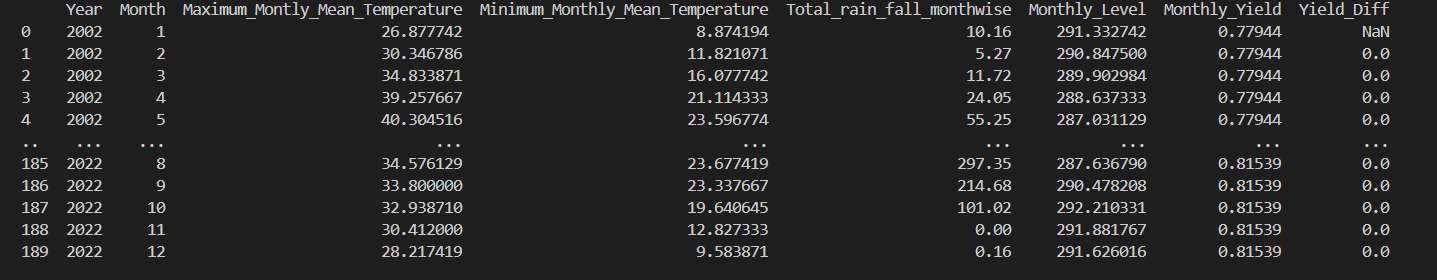
This dataset includes monthly data for Jharkhand from 2002 to 2022. It covers maximum and minimum temperatures, total monthly rainfall, water levels, and crop yields. This information helps us see how climate conditions and water availability affect crop yields over time. It is ideal for time series analysis and forecasting since it reflects the seasonal and yearly changes that impact agricultural production.



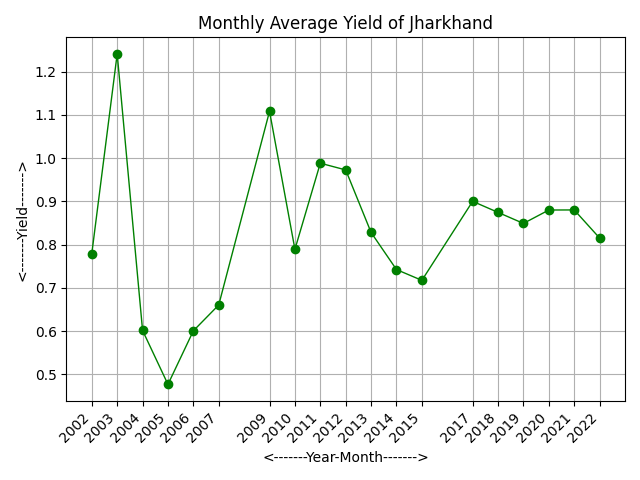
**Outlier Detection and Removal**

During the data preprocessing stage, we identified outliers in the dataset to ensure reliable modeling results. We detected outliers by calculating the difference (Yield\_Diff) between actual yield values and their expected trend. We marked rows with unusual or extreme variations in temperature, rainfall, or yield as outliers. These abnormal values may have happened due to measurement errors, the impact of missing rainfall, sudden climate changes, or data recording mistakes. After reviewing these records, we removed the outliers to avoid distortion in the time series pattern and improve the forecasting model's performance and accuracy. As a result, the dataset was reduced to 190 valid rows, representing stable and realistic climatic and yield conditions suitable for ARIMAX-based forecasting.

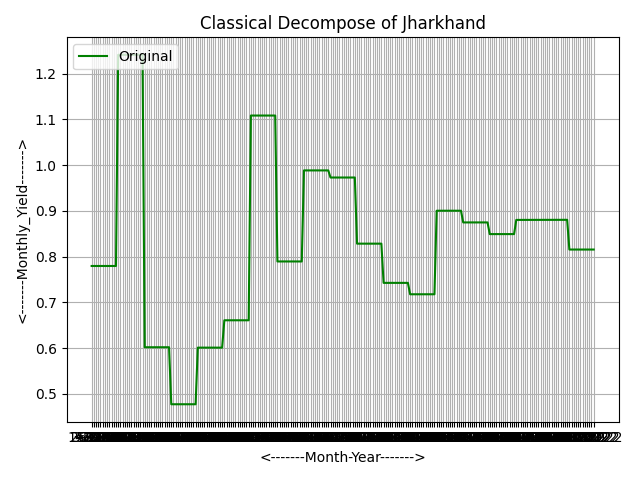
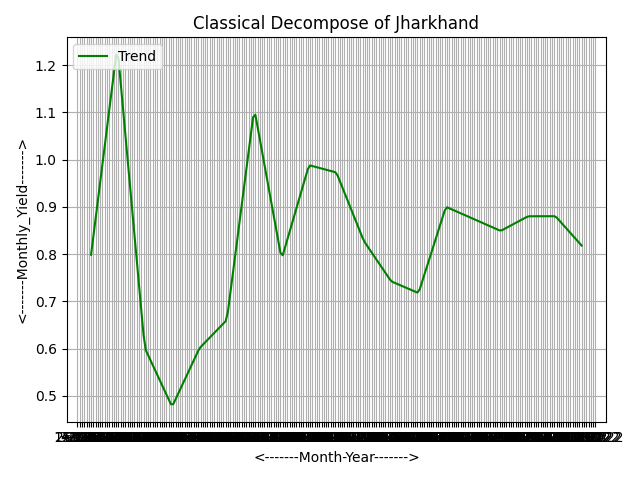


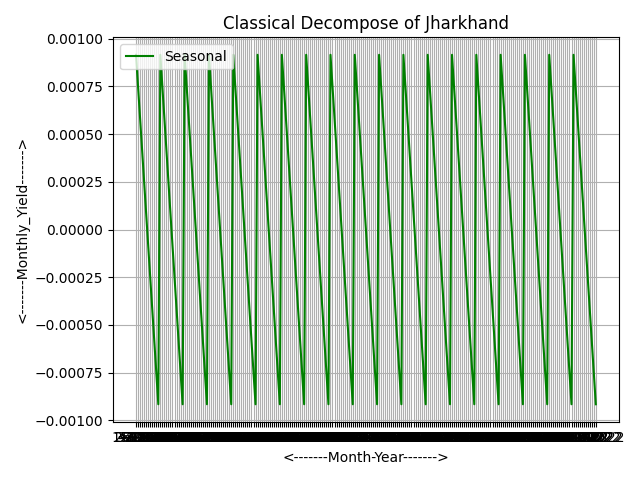
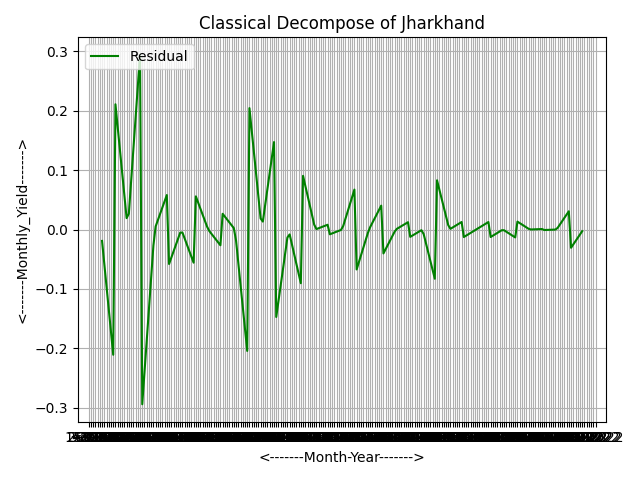


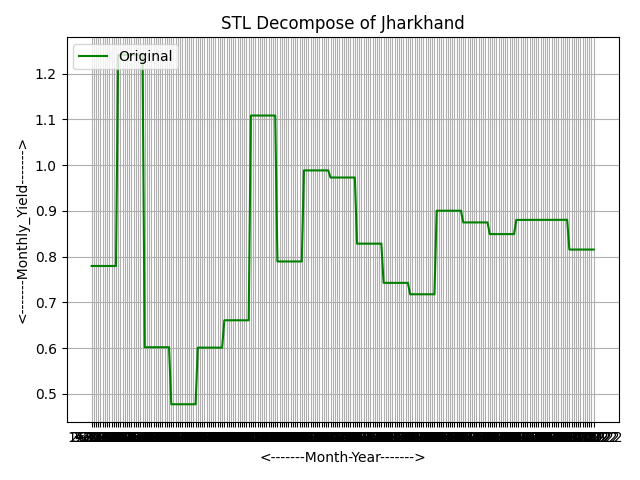
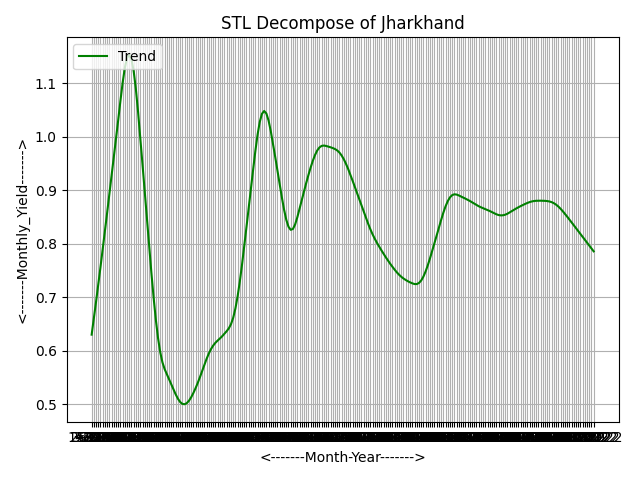
The average monthly yield of massor (lentil) in **Jharkhand** from 2002 to 2022 is shown in the graph.

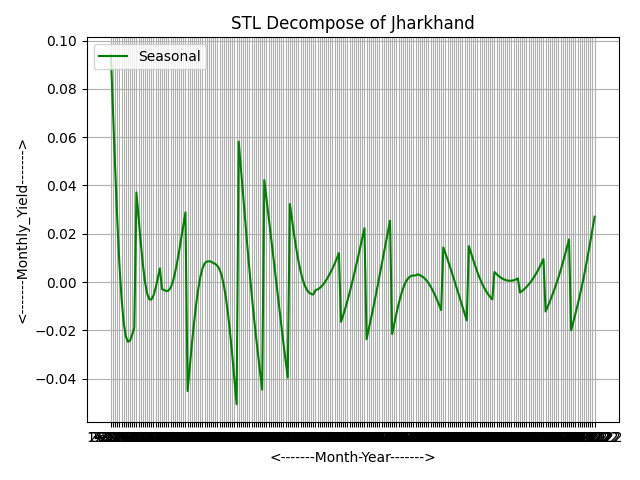
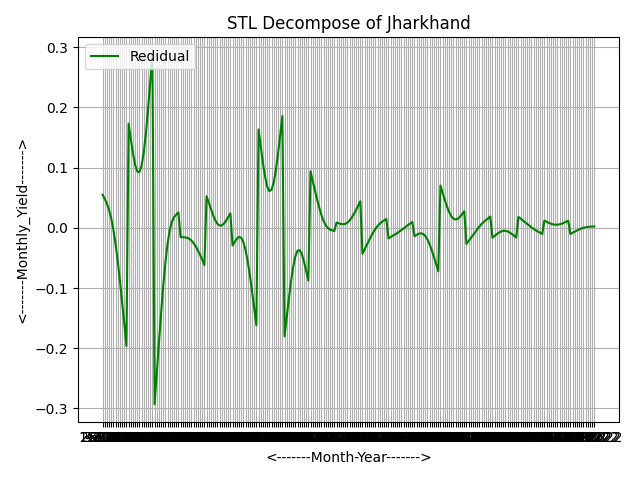


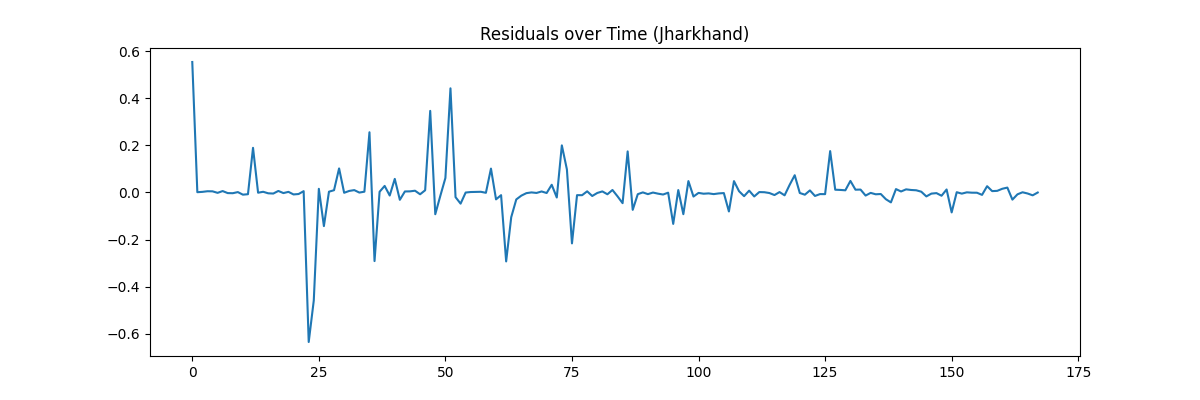
Time series Classical Decomposition plot and Time series STL Decomposition plot-

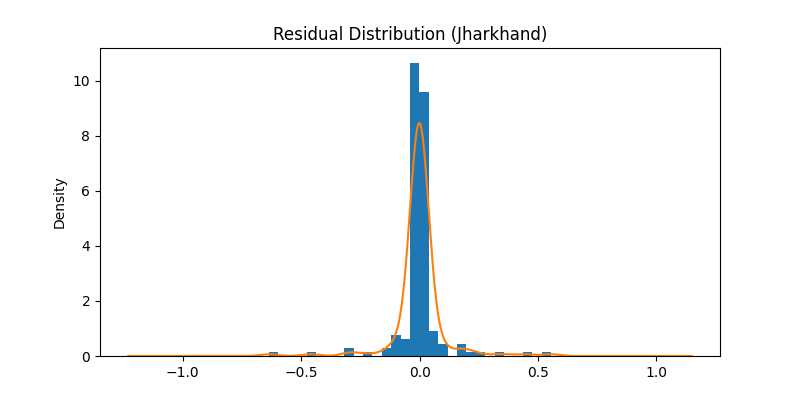
 



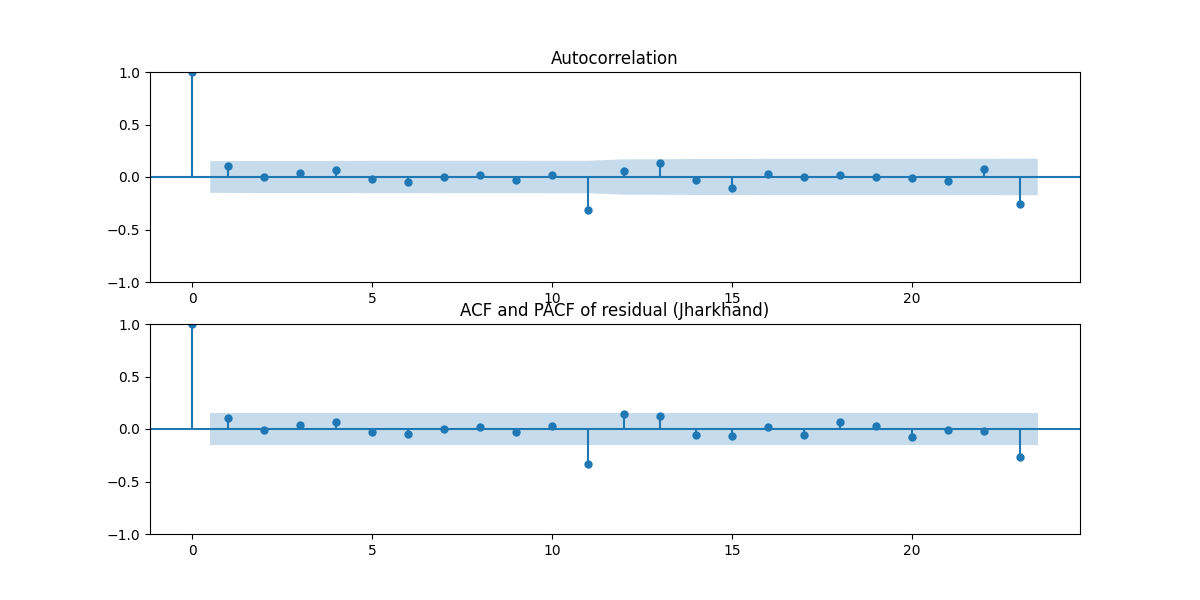
Residual Distribution (Jharkhand):-

The residual plot shows that most residual values are close to zero. This indicates that the ARIMAX model has captured the overall pattern of the data effectively. The distribution is nearly normal, with only a few values lying far from the center. This suggests that the model errors are random and there is no significant bias in predictions, which supports the reliability of the forecasting results.



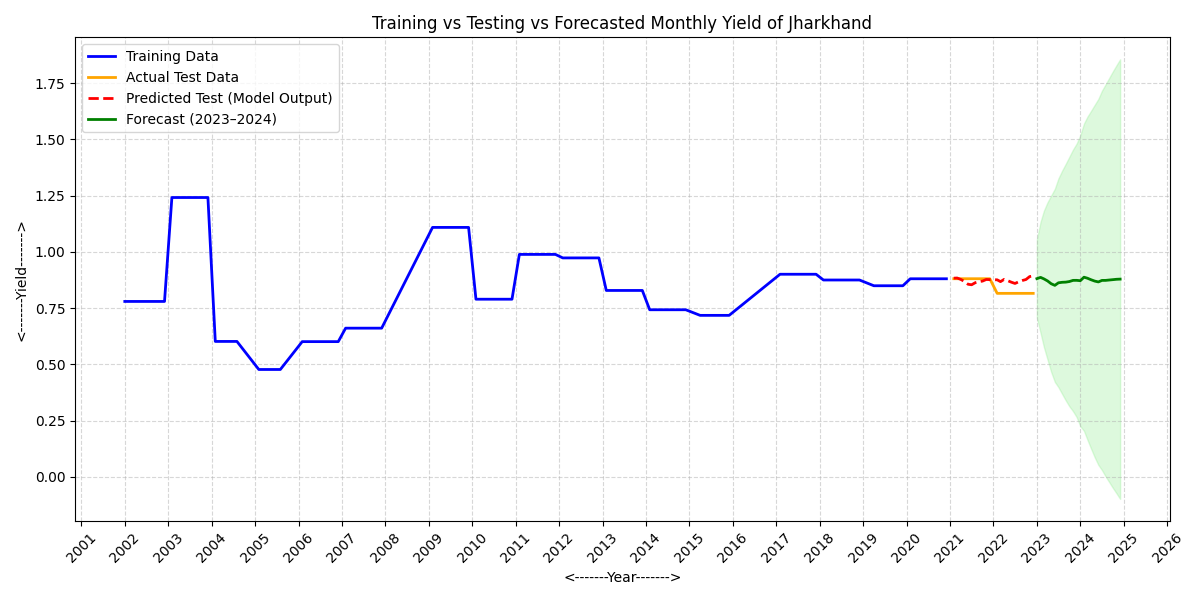
**ACF and PACF of Residuals (Jharkhand):**

The ACF and PACF plots of the residuals show that most of the spikes are within the confidence bands. This means there is no significant autocorrelation in the residuals. In other words, the residuals act like white noise, which indicates that the model has effectively captured the underlying patterns in the data. Therefore, the ARIMAX model used for Jharkhand is suitable and offers reliable forecasting results.

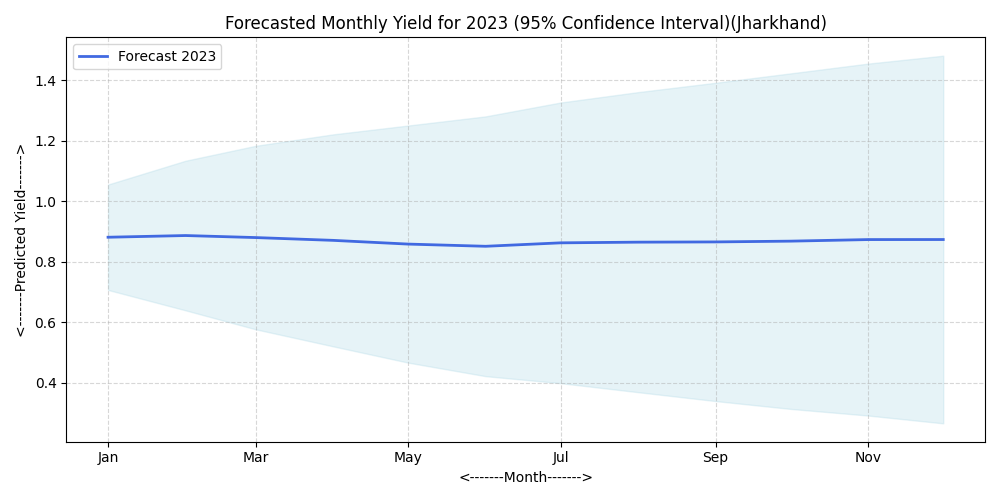


**Training vs Testing vs Forecast (Jharkhand):** -

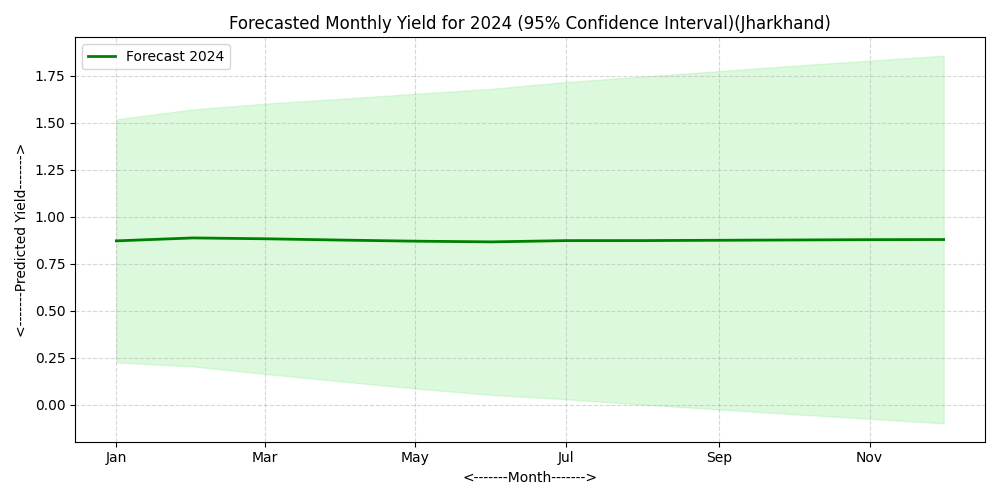
The model predictions closely match the actual data and show good accuracy. The forecast for 2023 to 2024 indicates stable yield with minor variations, suggesting the model provides reliable future estimates.



The graph displays the monthly yield forecast for Jharkhand in 2023. It shows stable yield values throughout the year, with slight changes.

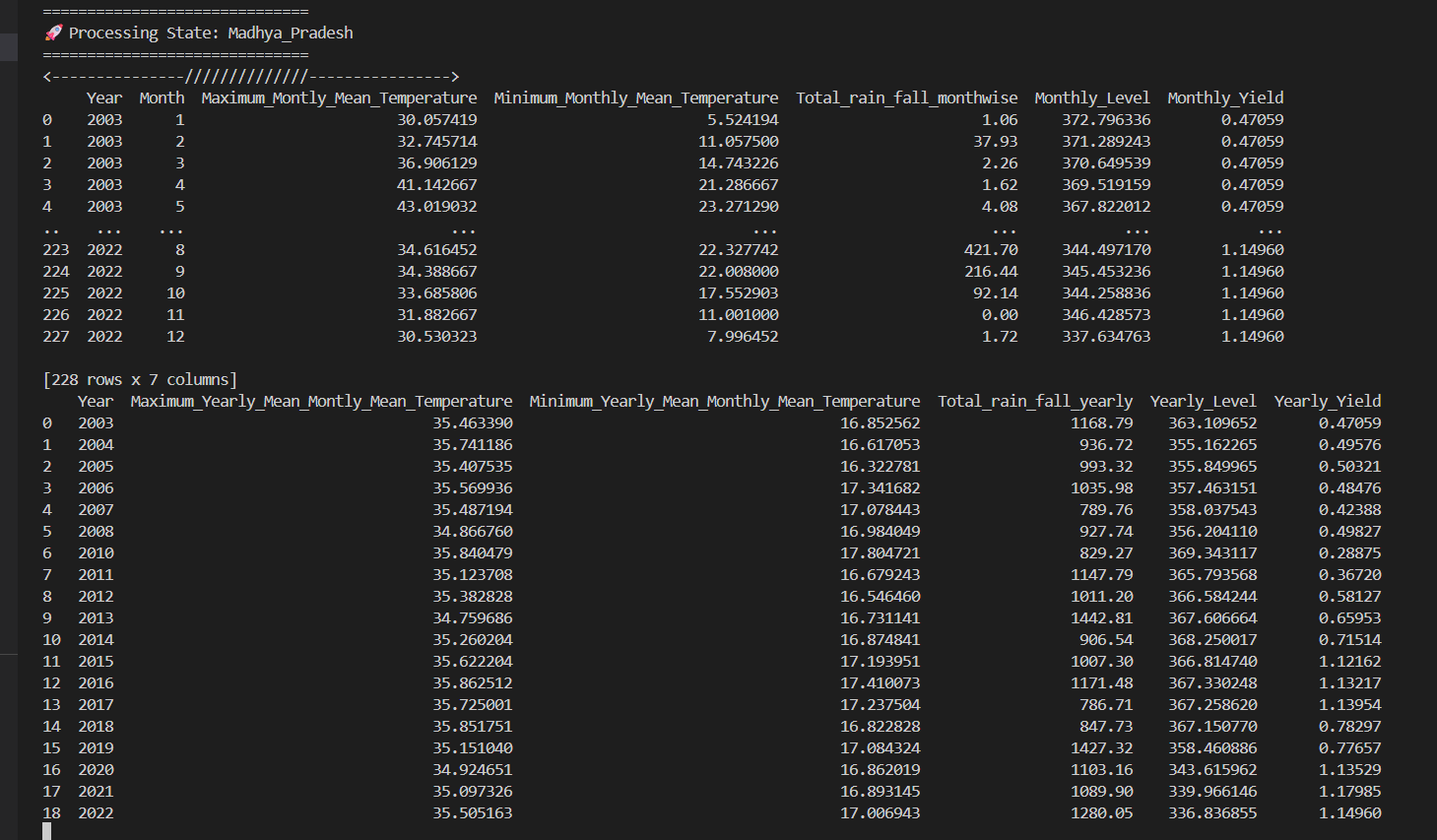


The graph displays the monthly yield forecast for Jharkhand in 2024. It shows a steady yield pattern with some minor fluctuations and stable growth over the year.



**Dataset Description (Madhya Pradesh)**

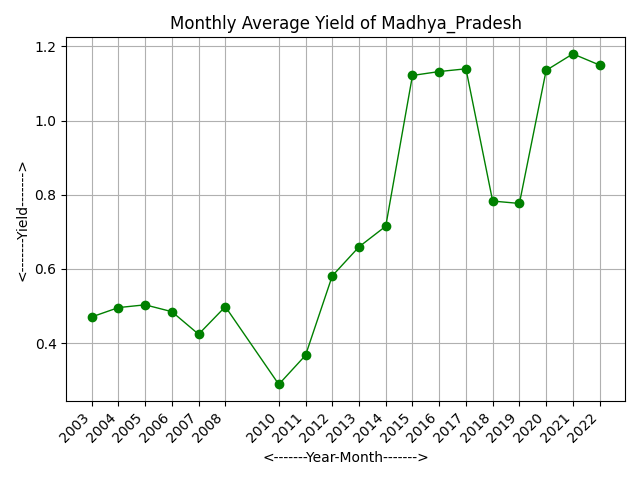
The dataset provides climate and agricultural information for Madhya Pradesh from 2003 to 2022. It includes monthly data on maximum and minimum mean temperature, total rainfall, reservoir water level, and crop yield, as well as their yearly totals. The monthly data allows for the study of seasonal changes, while the yearly summaries help identify long-term trends. Researchers use this dataset to assess the link between climate factors and crop productivity and to predict future yield patterns.



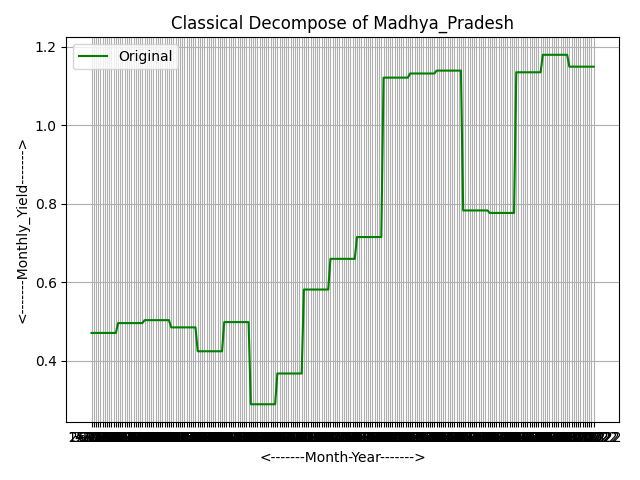
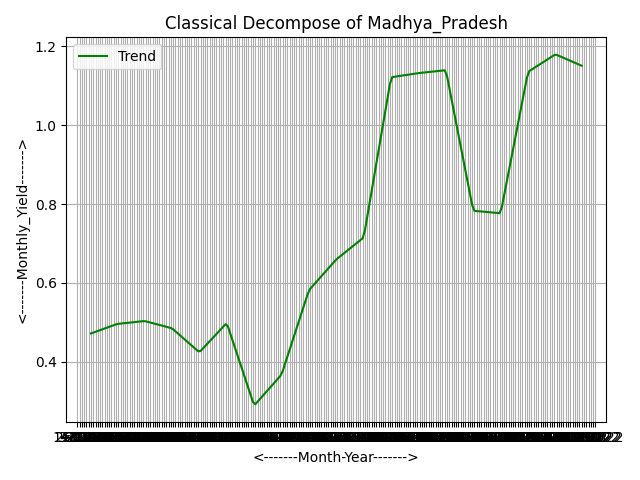
Outliers were found in the temperature, rainfall, and yield values through statistical comparison. We removed these extreme observations to prevent distortion in the analysis and to improve model accuracy. After removing the outliers, we kept 200 cleaned monthly records for further study.

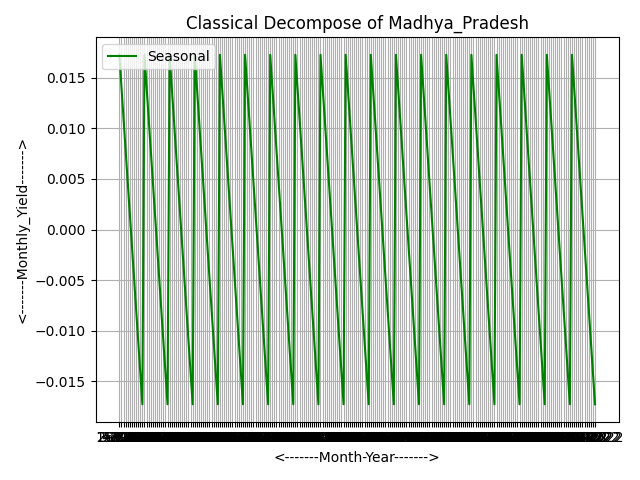
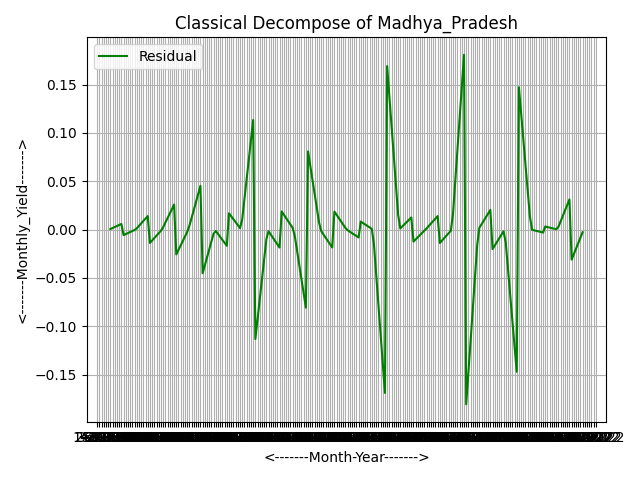


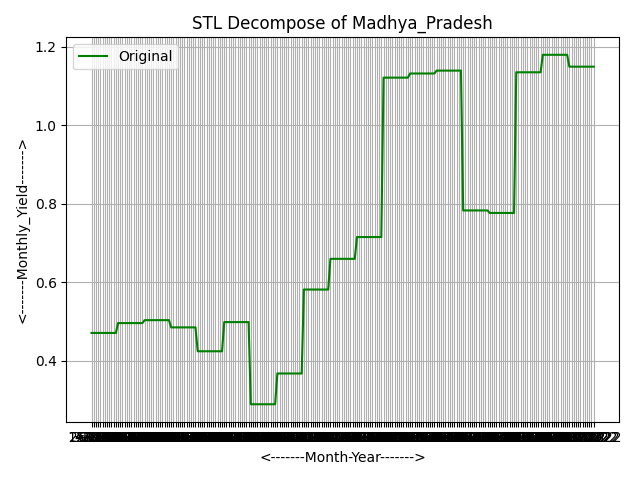
The average monthly yield of massor (lentil) in Madhya Pradesh from 2003 to 2022 is shown in the graph.

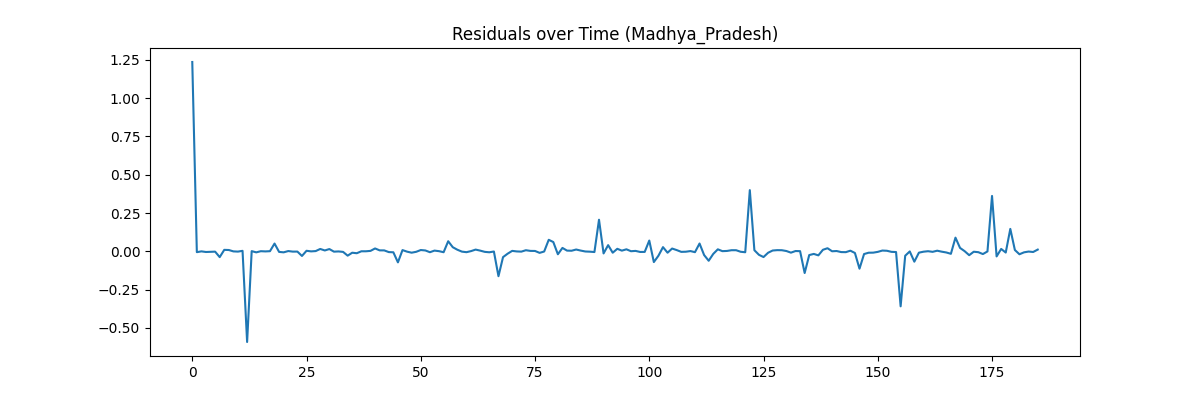
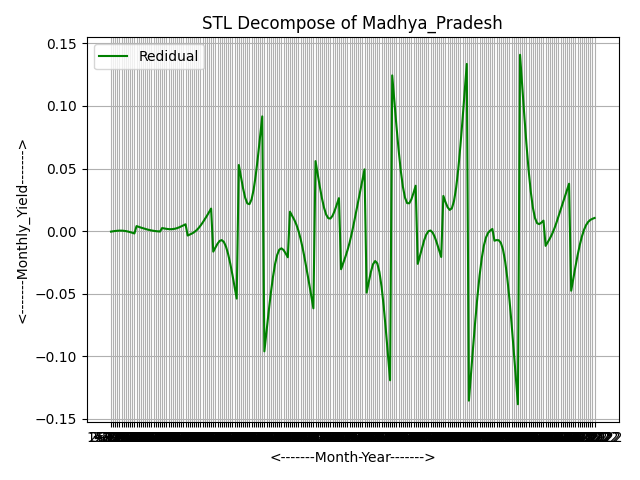


Time series Classical Decomposition plot and Time series STL Decomposition plot-

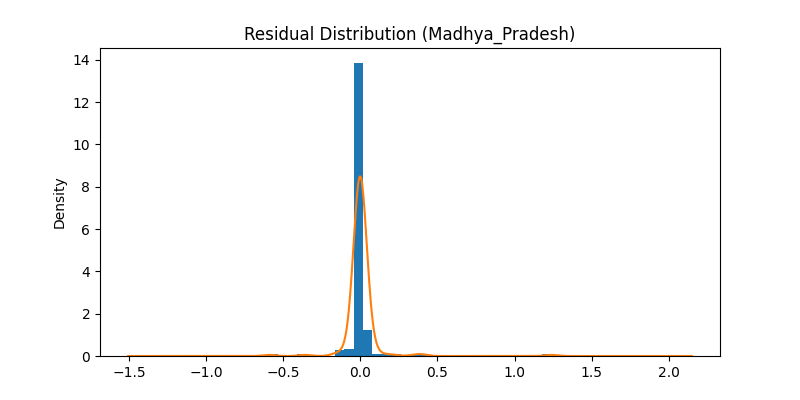
 

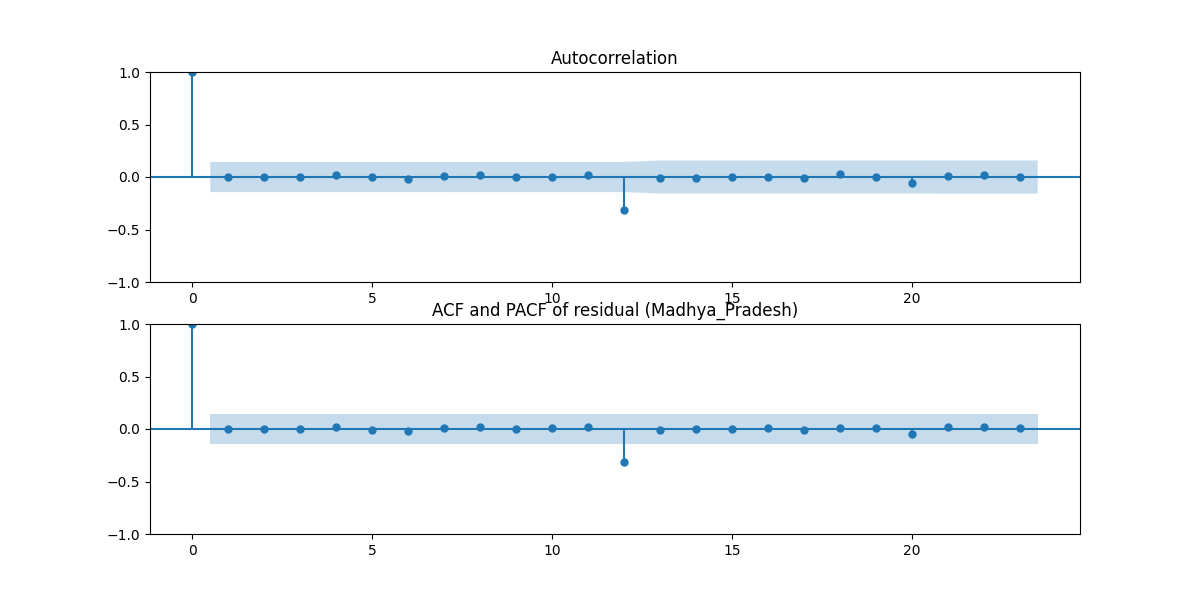
Residual Analysis(**Madhya Pradesh**)

The residual distribution is centered around zero and has a narrow spread. This indicates that the model fits the data well, with minimal prediction errors. It suggests that the forecasting model effectively captures the underlying pattern.



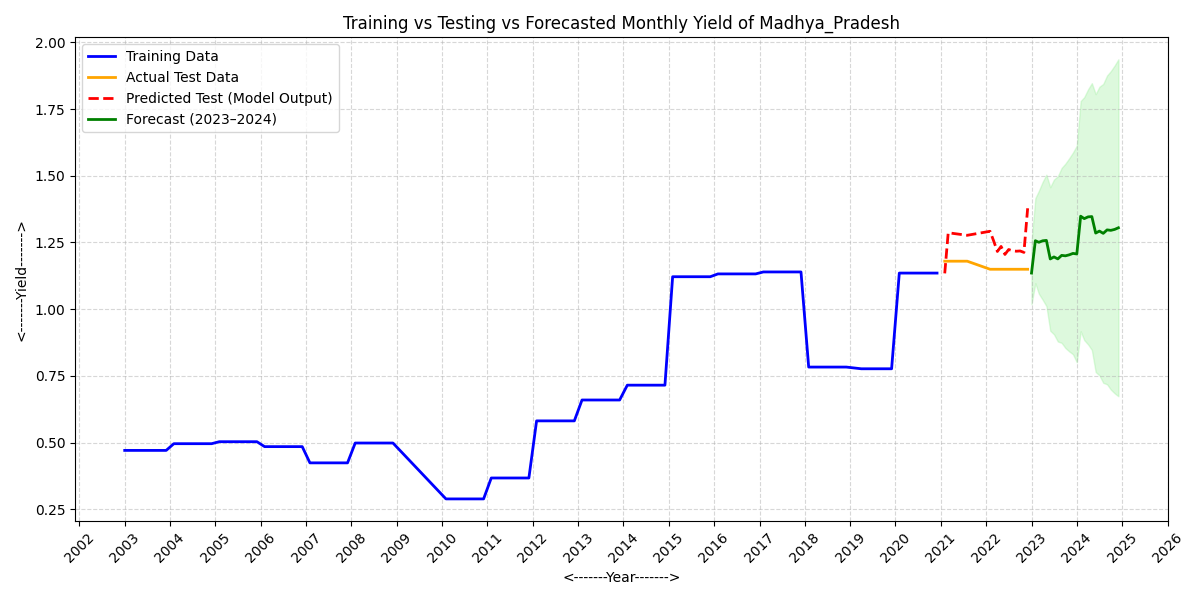
**ACF and PACF of Residuals**

The ACF and PACF plots show that the residuals do not display any significant autocorrelation. All spikes stay within the confidence bands. This means the model has successfully captured the data patterns, and there is no further information left unexplained. As a result, the residuals behave like white noise, confirming that the model is adequate.

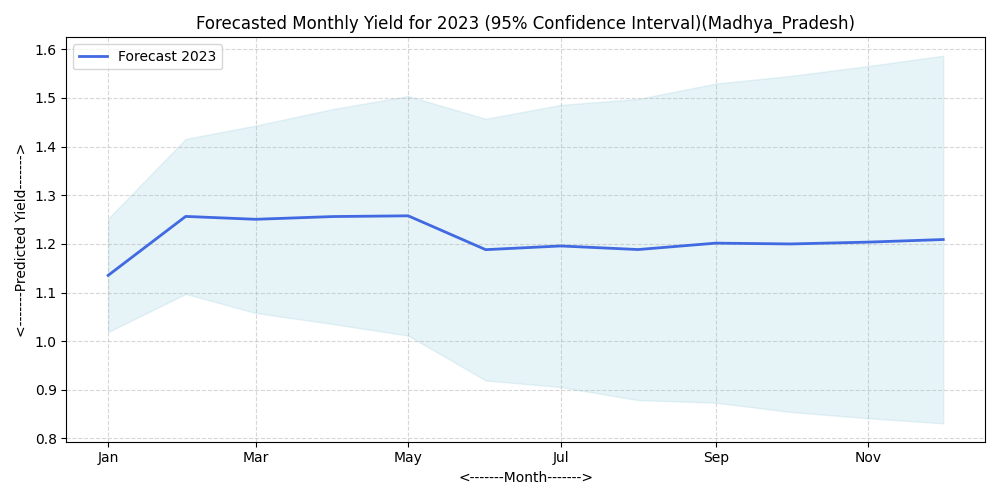


**Training vs Testing vs Forecasted Yield (Madhya Pradesh)**

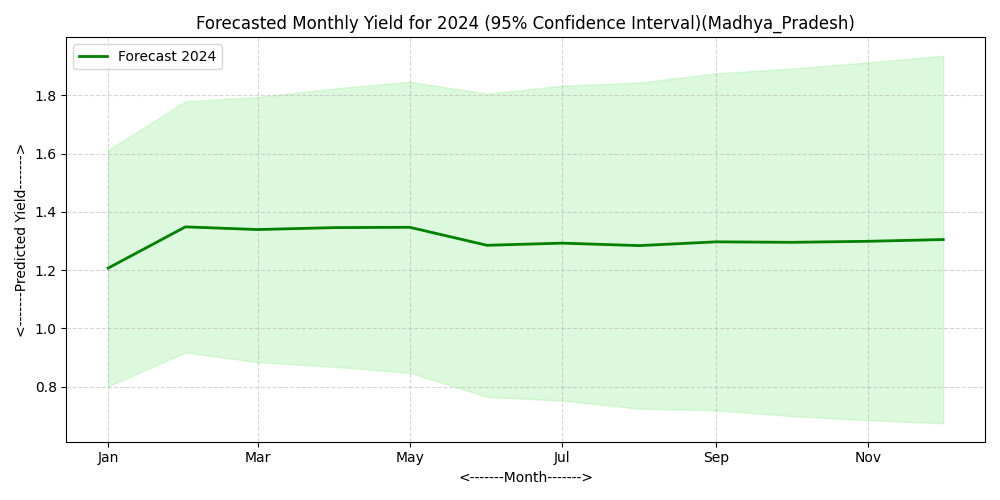
The graph compares the training data, testing data, model predictions, and future yield forecasts. The predicted values closely match the actual test data, showing that the model performs well. The forecast for 2023 to 2024 indicates a small rise in crop yield, suggesting better productivity under similar climatic and water availability conditions.



The graph displays the monthly yield forecast for Madhya Pradesh in 2023. It shows a small increase early in the year, followed by steady yield levels for the rest of the months.

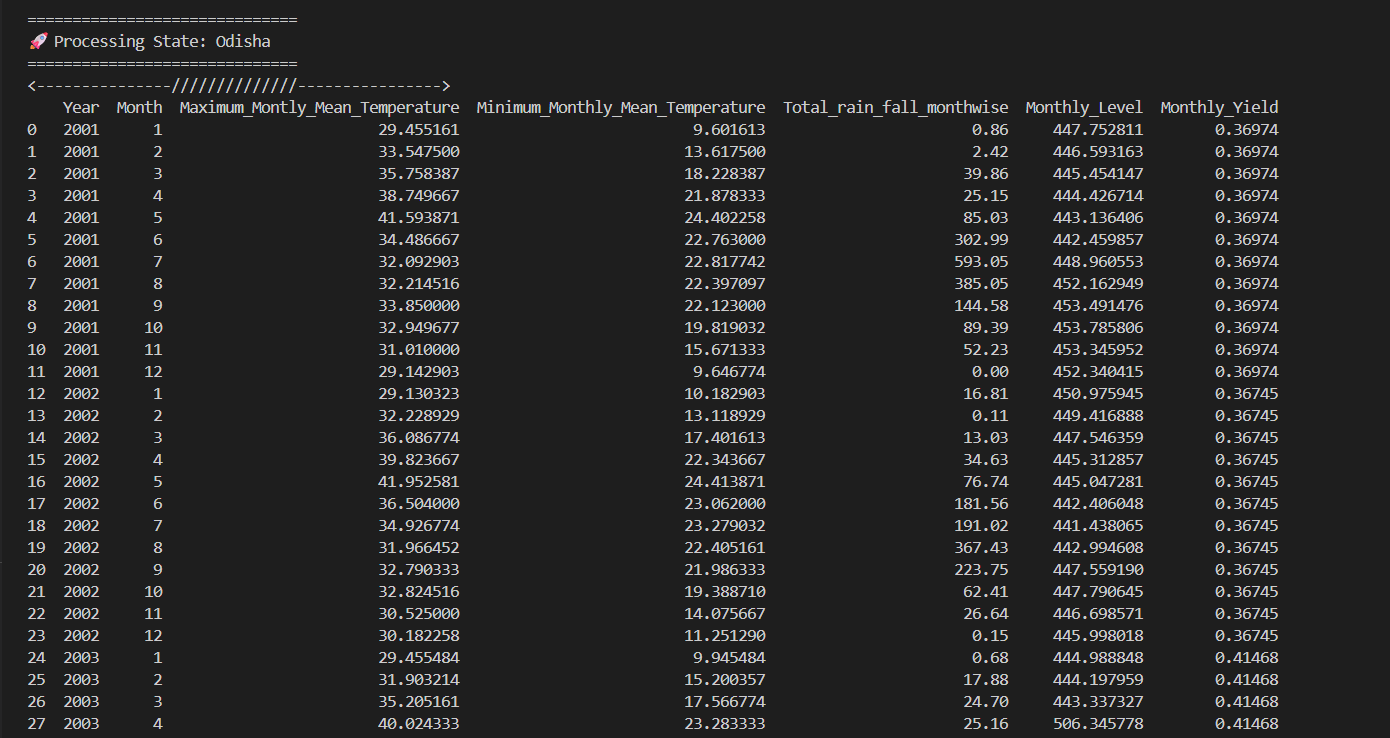


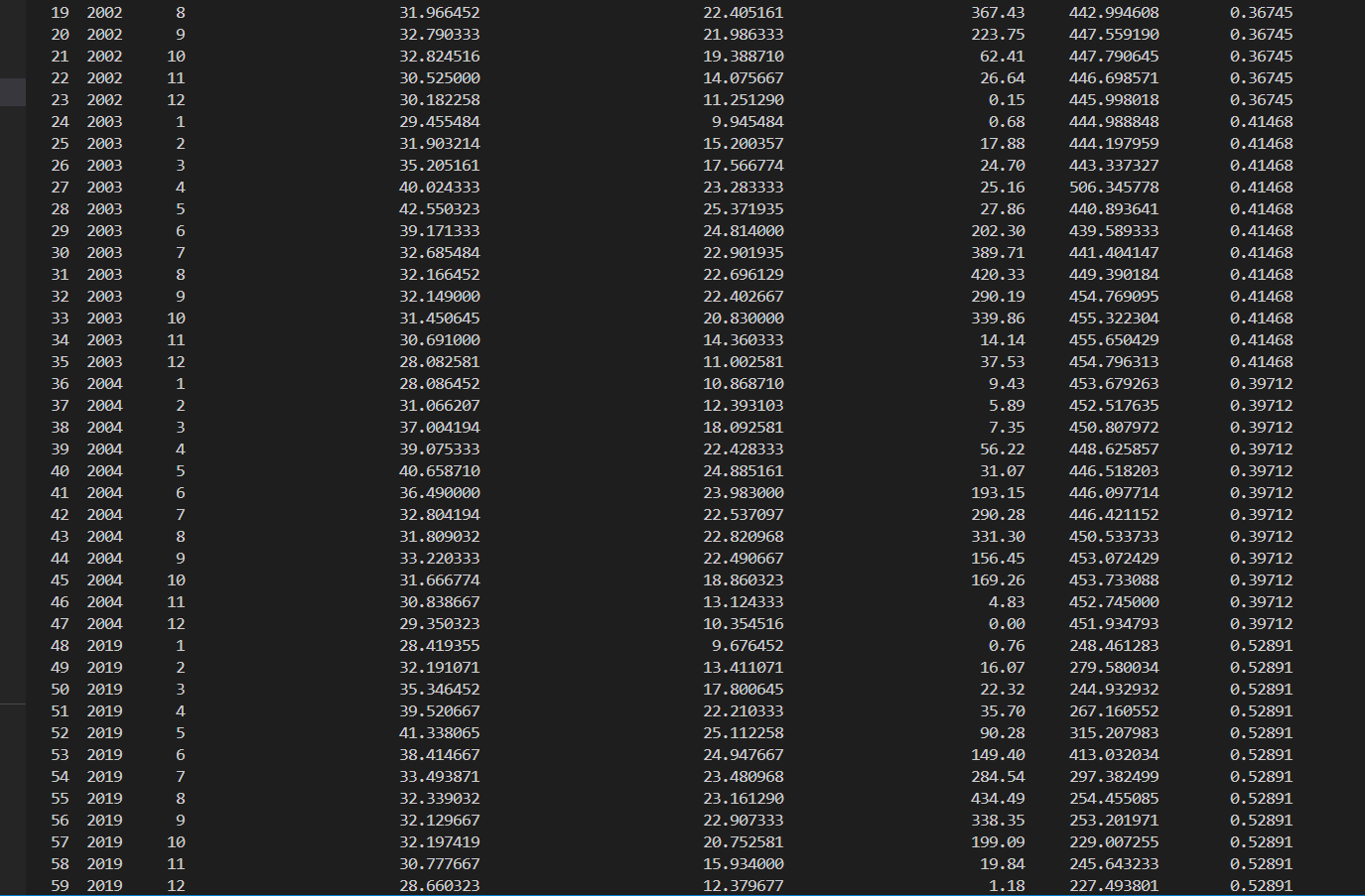
The graph displays the monthly yield forecast for Madhya Pradesh in 2024. It shows a slow increase at the start of the year, then stable yield levels for the rest of the months.



**Dataset Description (Odisha)**

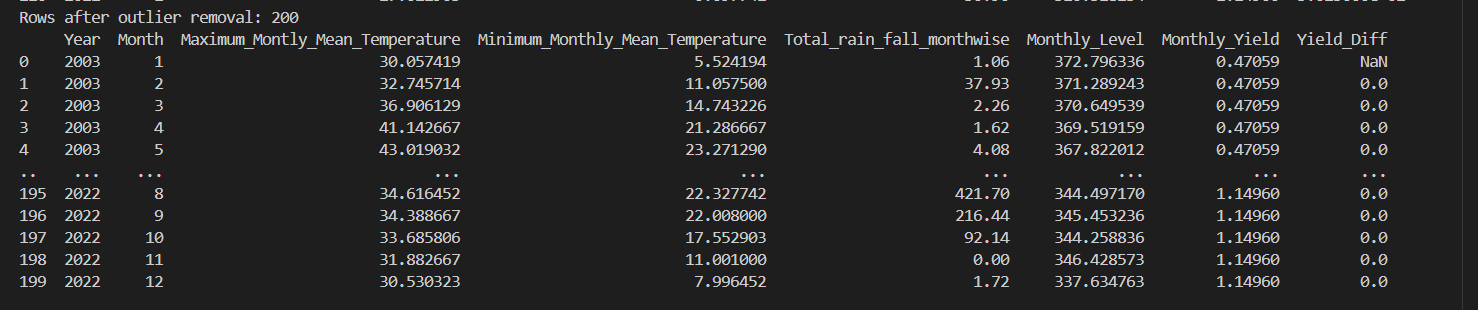
The dataset includes monthly records for Odisha from 2001 to 2022. It features maximum and minimum mean temperature, total monthly rainfall, reservoir water level, and crop yield. This data helps us study how climate and water availability affect crop yield trends over time.

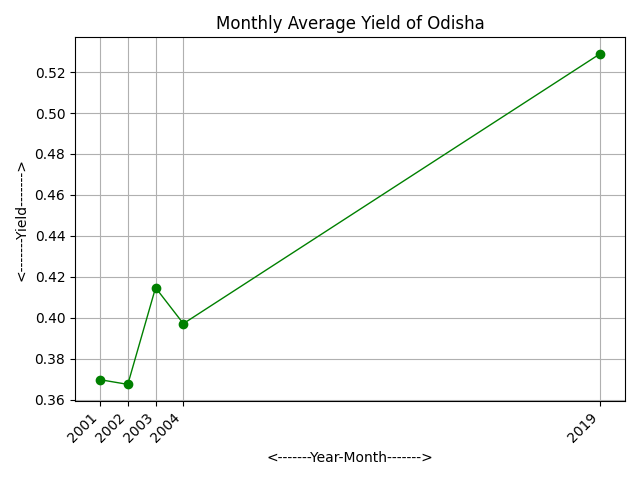




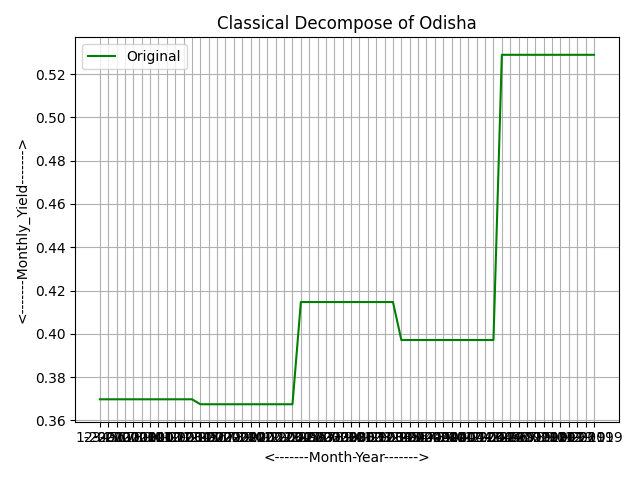
**Dataset (After Outlier Removal)**

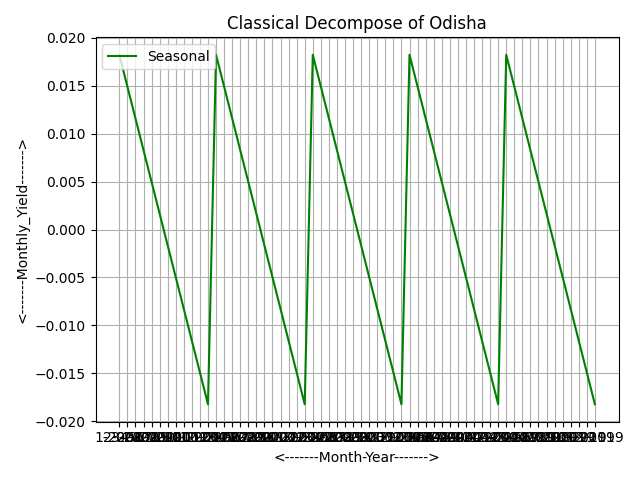
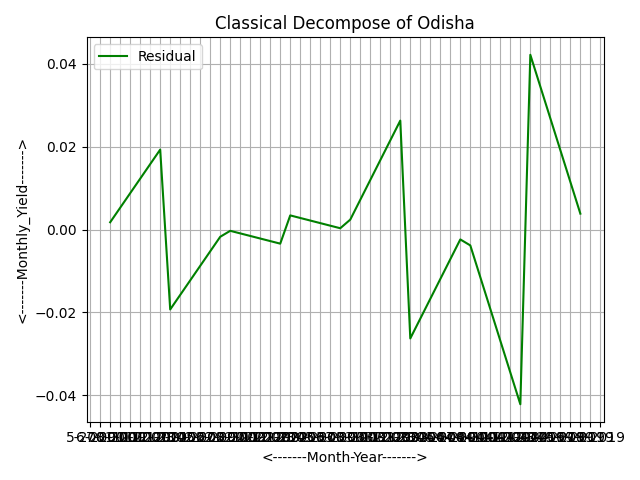
The cleaned dataset includes 200 monthly records from 2003 to 2022. It has data on temperature, rainfall, reservoir level, and crop yield. We removed outliers to improve the quality of trend analysis and forecasting.

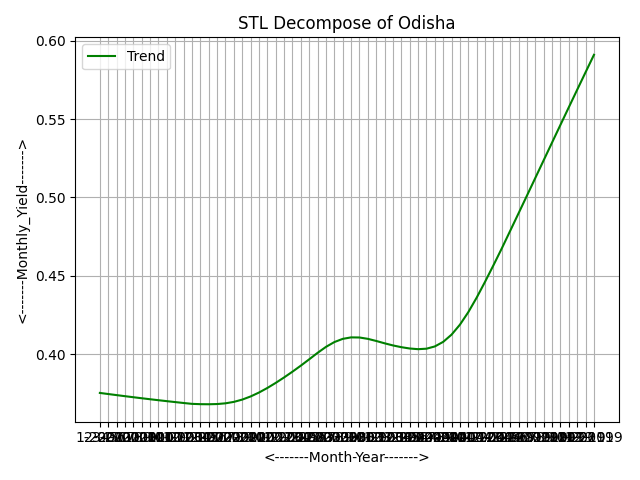
The average monthly yield of massor (lentil) in **Odisha**  **from 2001 to 2024 is shown in the graph.**

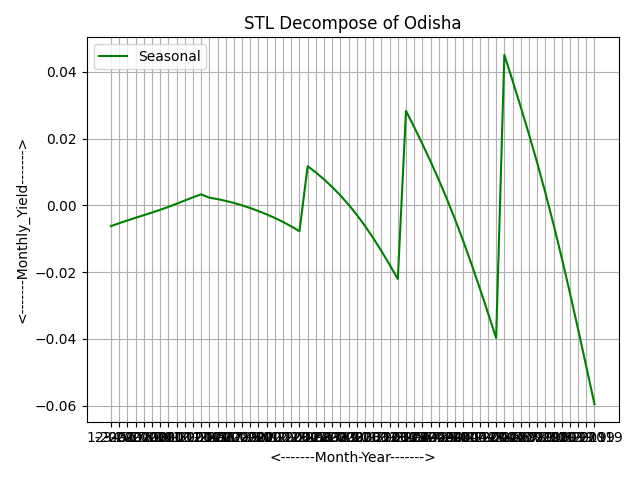
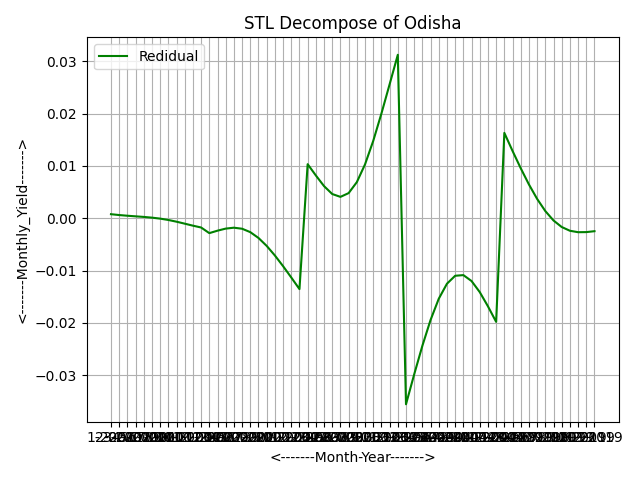


Time series Classical Decomposition plot and Time series STL Decomposition plot-

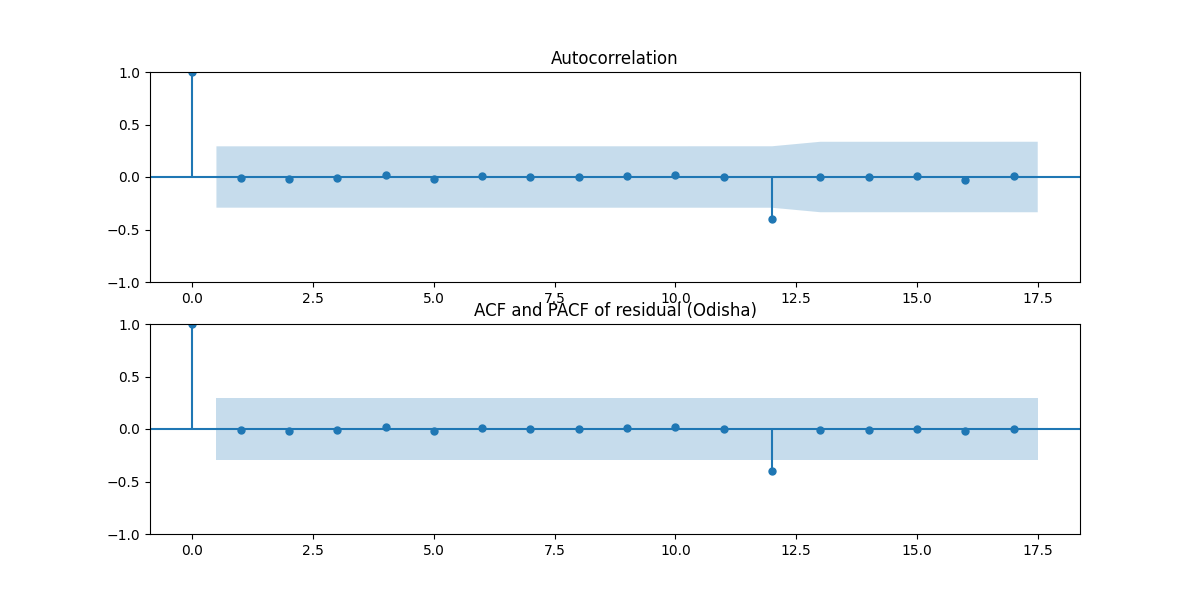
 

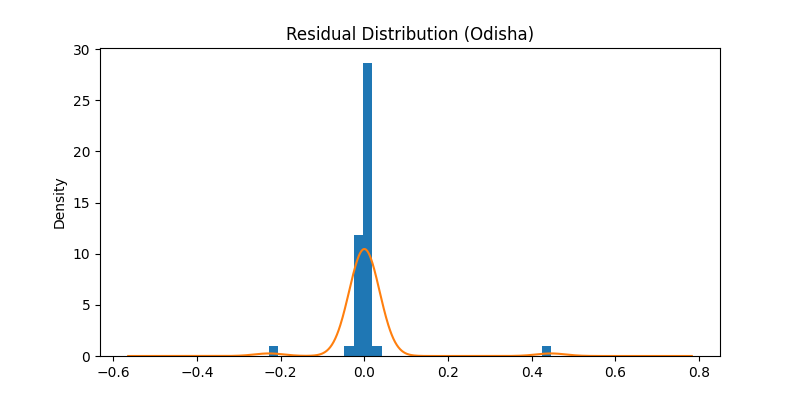
 

**ACF and PACF of Residuals (Odisha)**

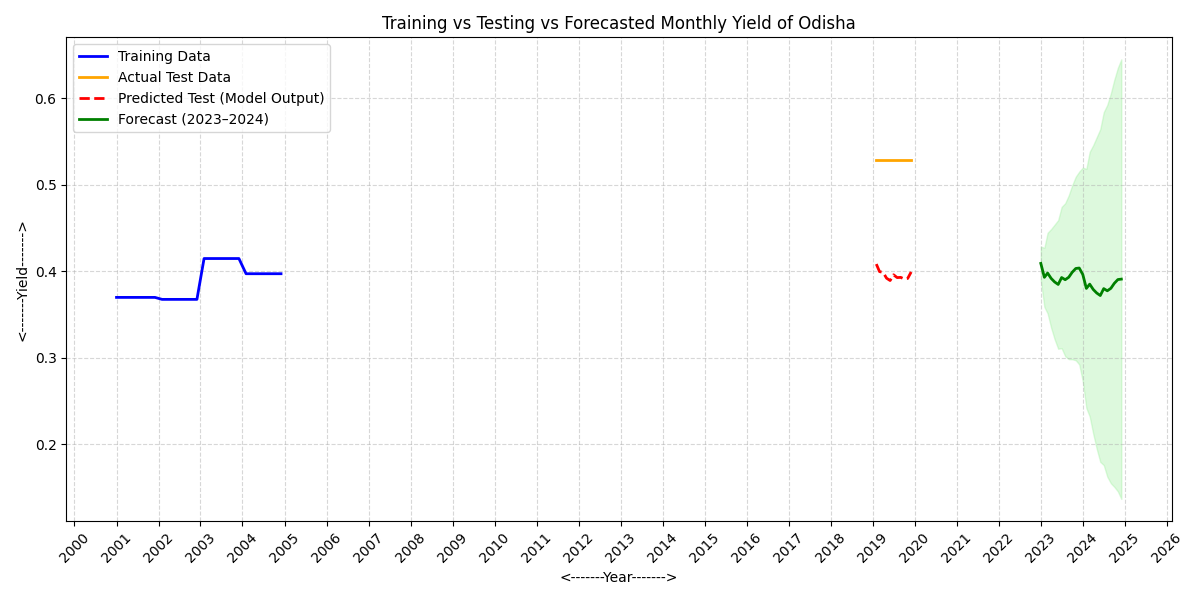
ACF and PACF of Residuals (Odisha) The ACF and PACF plots of the residuals show that all correlation values fall within the confidence limits. This indicates no significant autocorrelation. It confirms that the chosen model fits well and that the residuals behave like white noise, making the model reliable for forecasting.

**Residual Distribution: -**

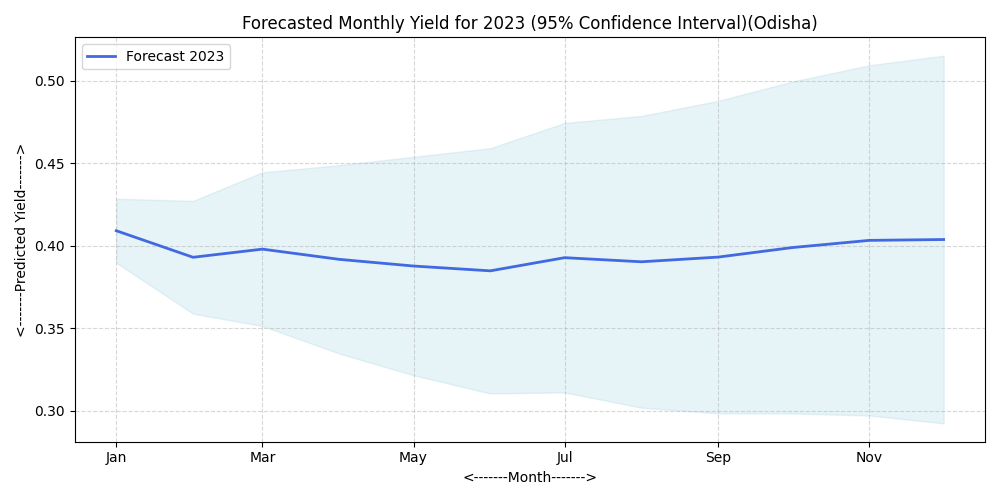
The residual distribution for Odisha shows that most values are close to zero. This indicates that the model’s predictions are usually correct. The distribution is roughly normal, with very few extreme values. This suggests that the model fits the data well and that the forecast errors are small and random.



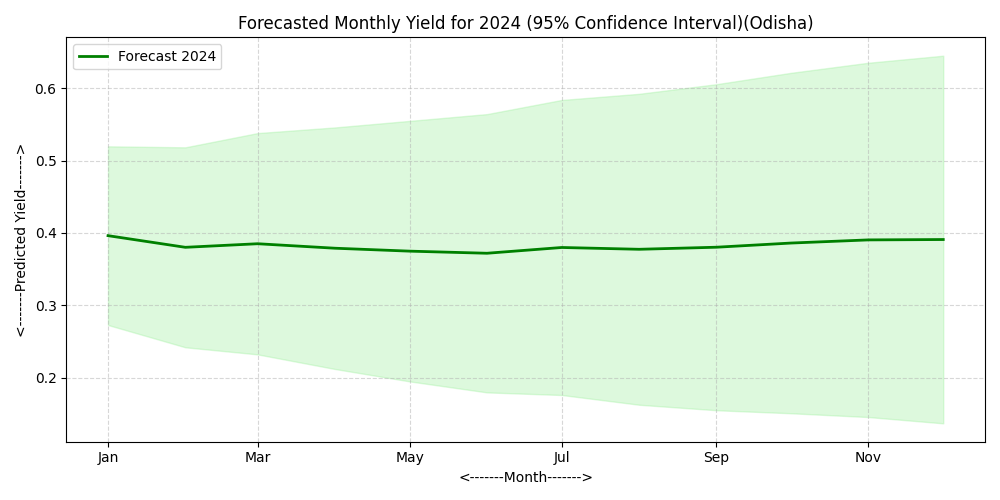
The comparison plot demonstrates that the model closely follows the trend of the actual test data. This indicates a good fit and reliable predictive performance. The forecast for 2023 to 2024 suggests a stable yield pattern with small fluctuations. The shaded confidence interval shows the range of uncertainty. While minor variations are expected, the overall yield is predicted to stay consistent in the coming years.



The graph shows the monthly yield forecast for Odisha in 2023. It indicates a slight decline in yield during the early months. This is followed by a gradual stabilization toward the end of the year.

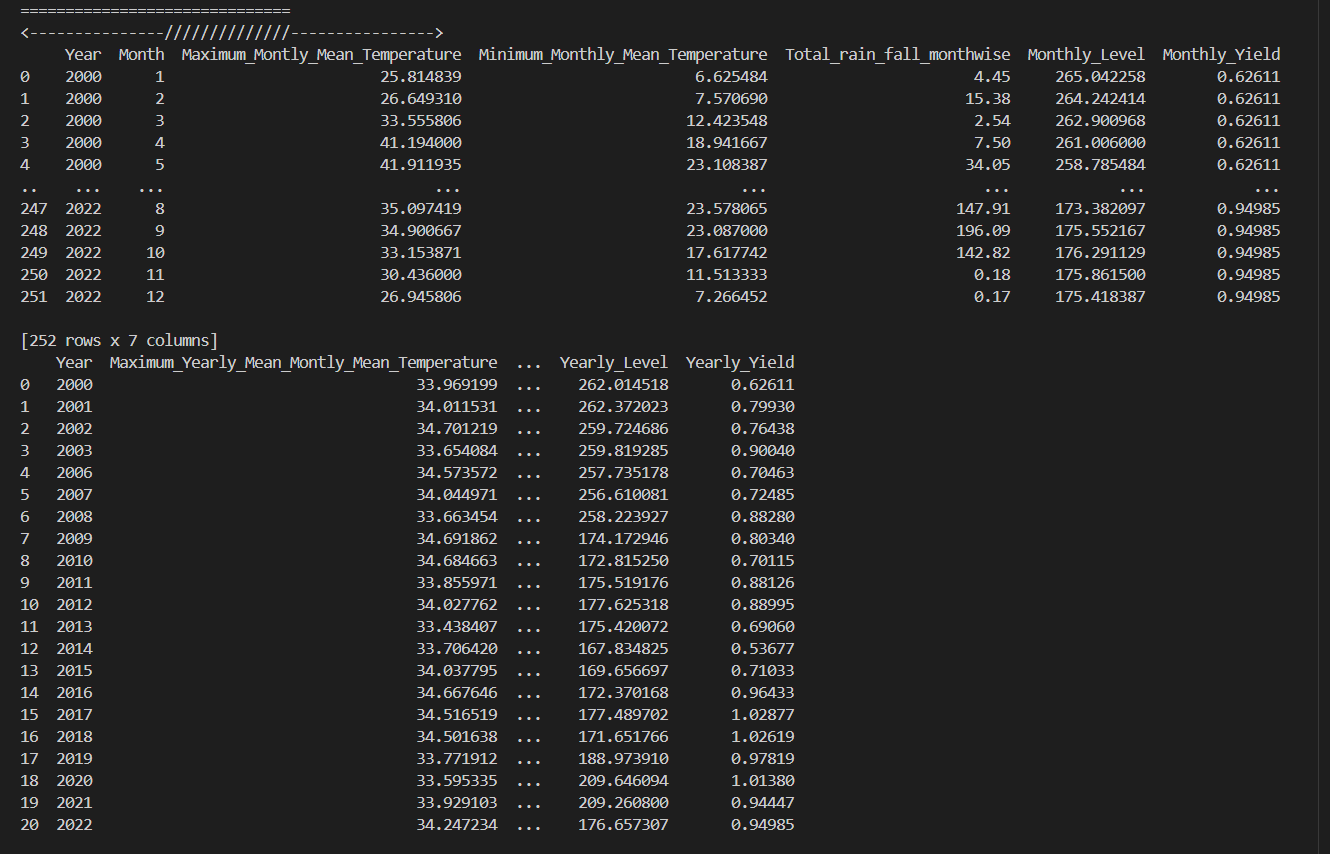


The graph displays the monthly yield forecast for Odisha in 2024. It shows a consistent yield trend with slight changes throughout the year.

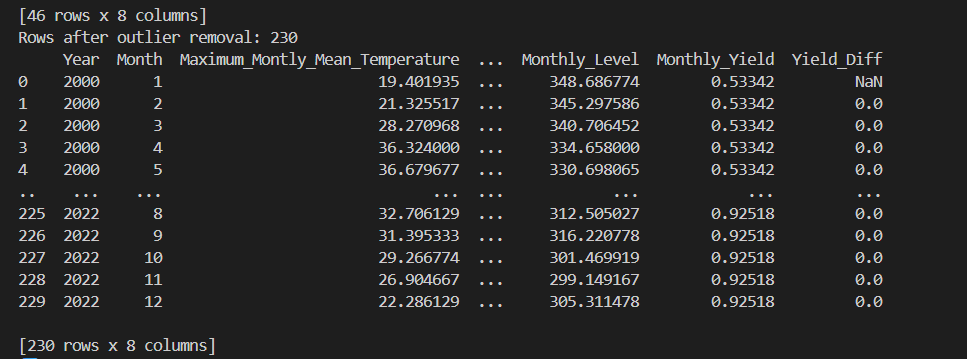


**Dataset Description (Rajasthan)**

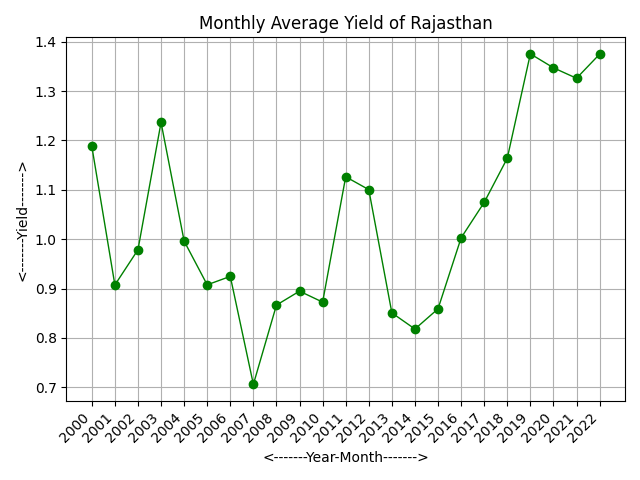
The dataset includes monthly and yearly records about climate factors and crop yield for the chosen state. The monthly data covers Maximum and Minimum Mean Temperatures, Total Rainfall, Water Reservoir Level, and Monthly Yield values from 2000 to 2022. Additionally, the dataset is summarized in yearly form, displaying Yearly Mean Temperature, Total Annual Rainfall, Average Reservoir Level, and Annual Crop Yield. These variables help us understand how climate factors and water availability affect crop yield trends over time.



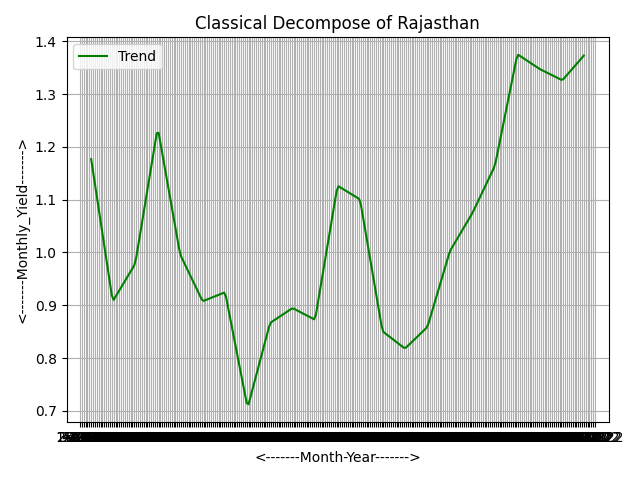
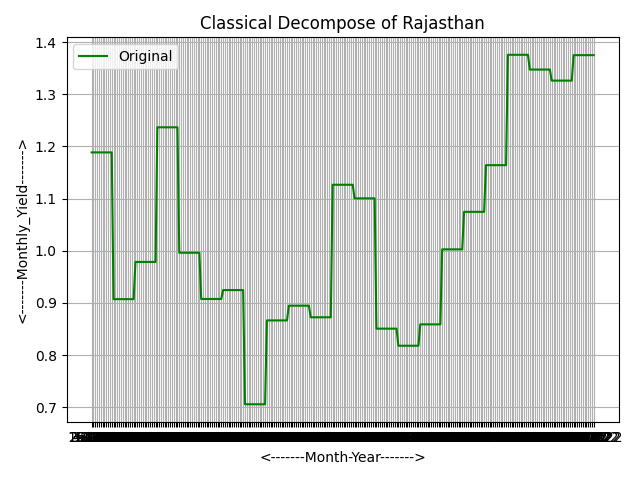
Outliers were removed to make the data more accurate. After taking out 46 unusual records, 230 clean monthly observations remained. This leads to a more reliable analysis and forecasting of yield trends.

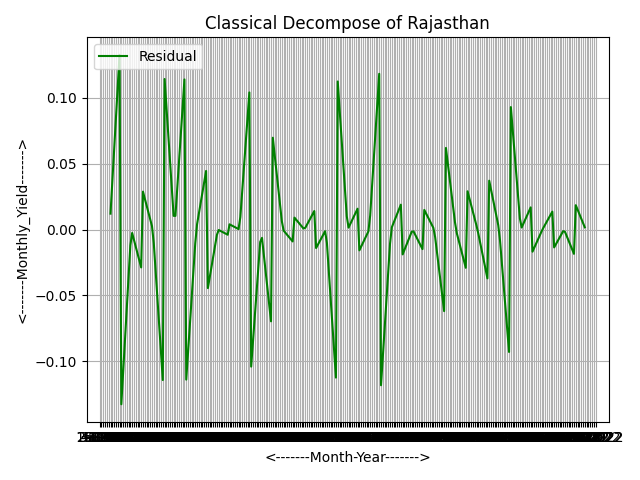
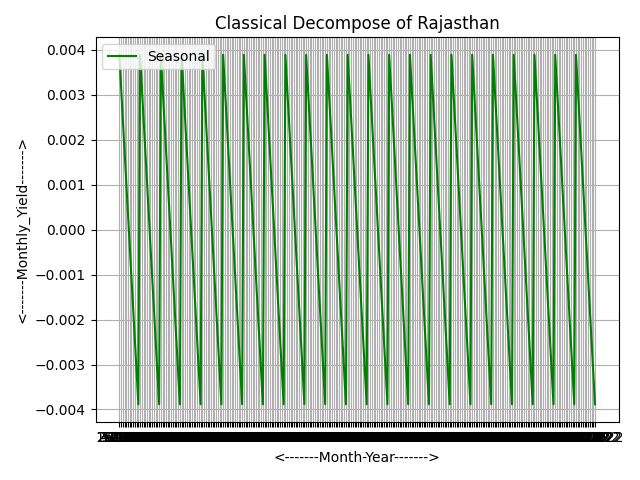


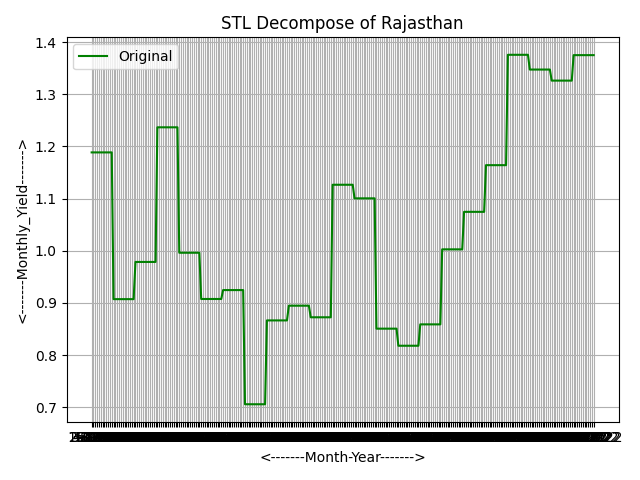
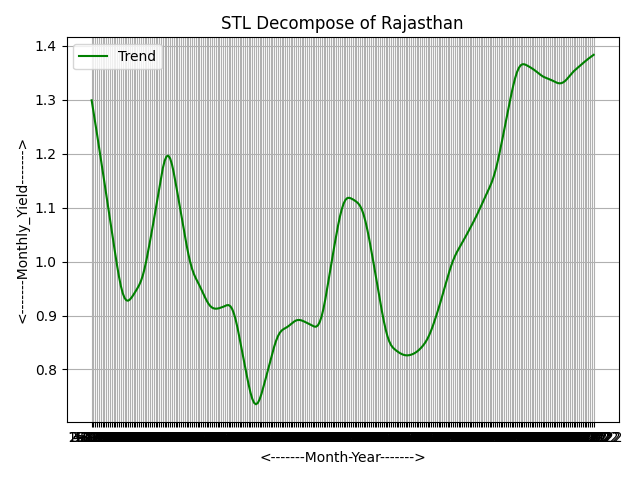
The average monthly yield of massor (lentil) in **Rajasthan**  **from 2000 to 2022 is shown in the graph.**

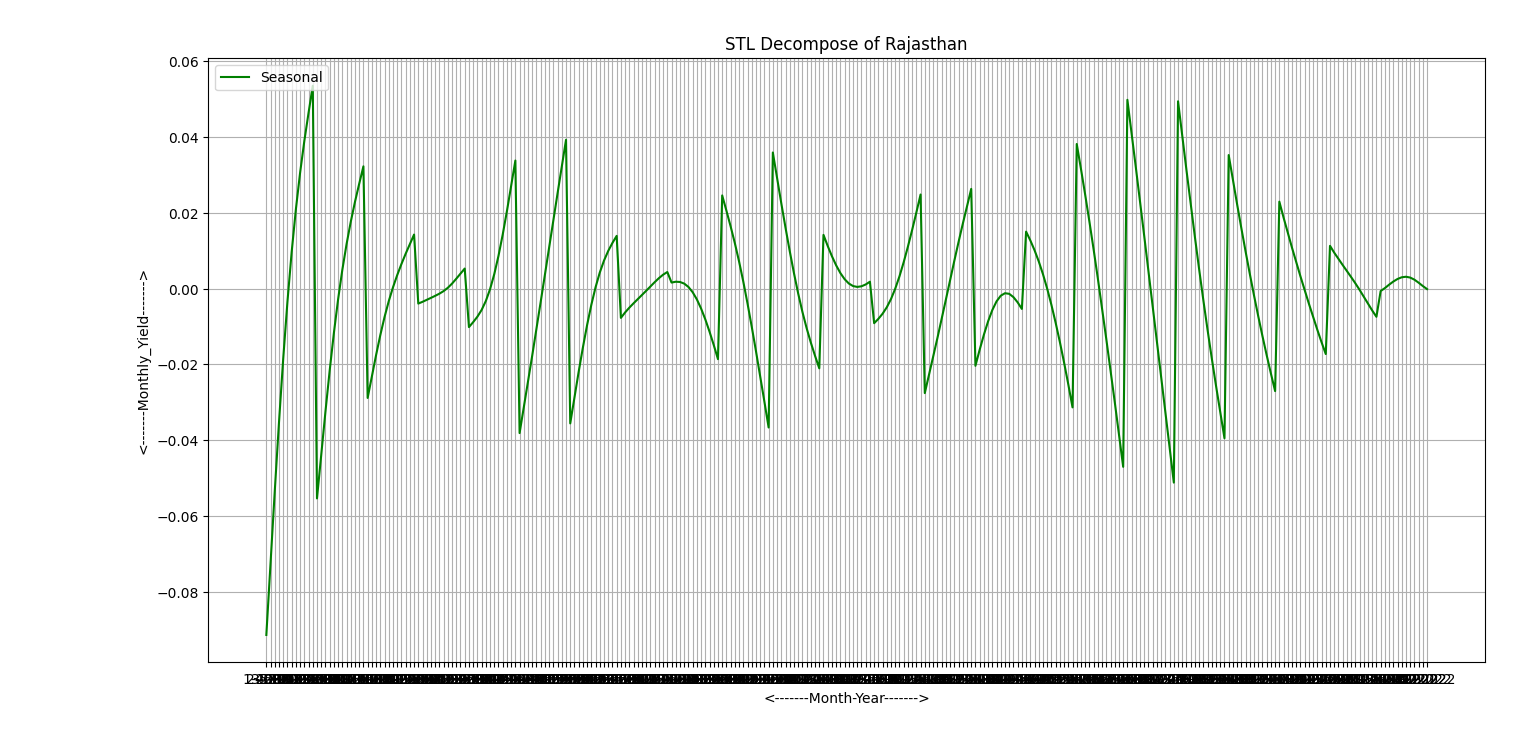
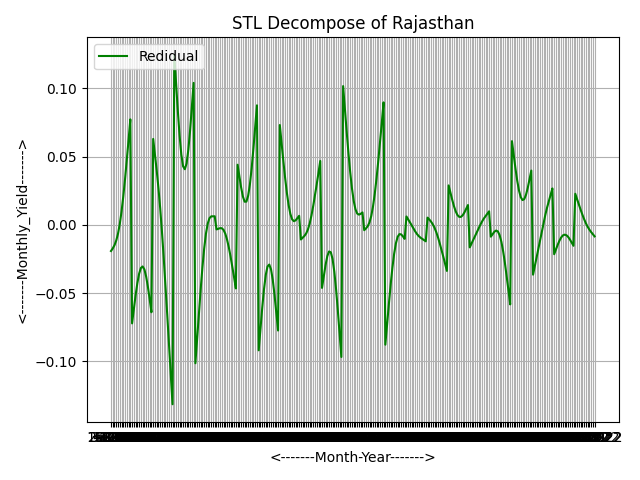


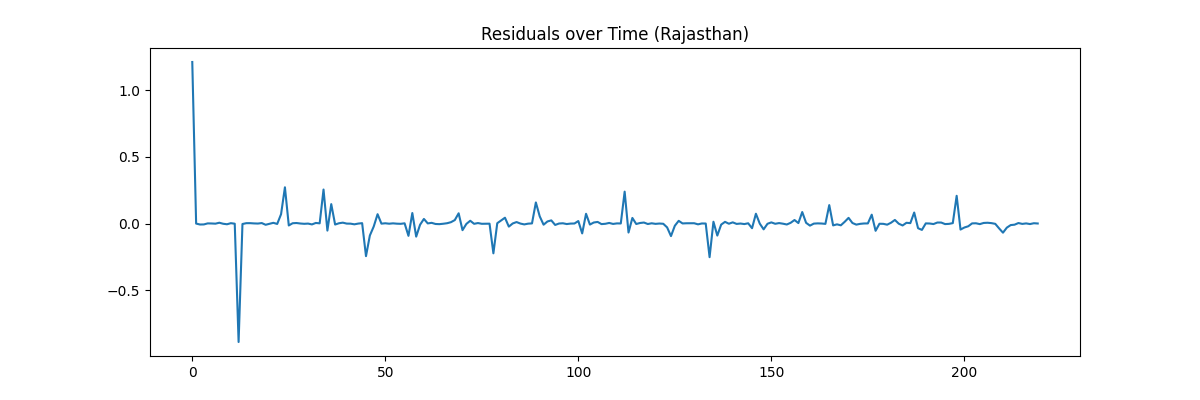
Time series Classical Decomposition plot and Time series STL Decomposition plot-





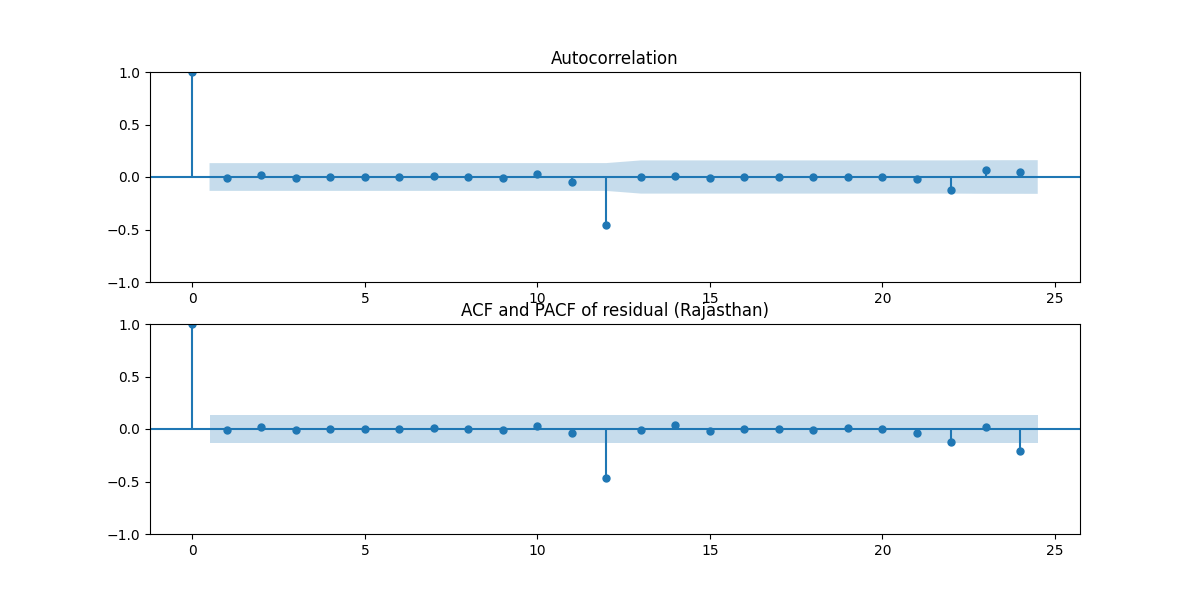
 

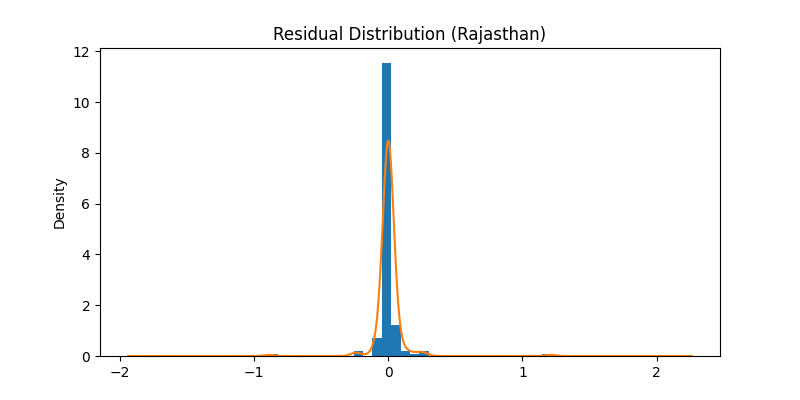


**ACF and PACF of Residuals**

The ACF and PACF plots of the model residuals show that all autocorrelation values fall within the confidence limits. This means there is no significant correlation remaining in the residuals. It suggests that the model has effectively captured the underlying patterns in the data. The residuals act like white noise. Therefore, the fitted ARIMAX model is suitable and dependable for forecasting.

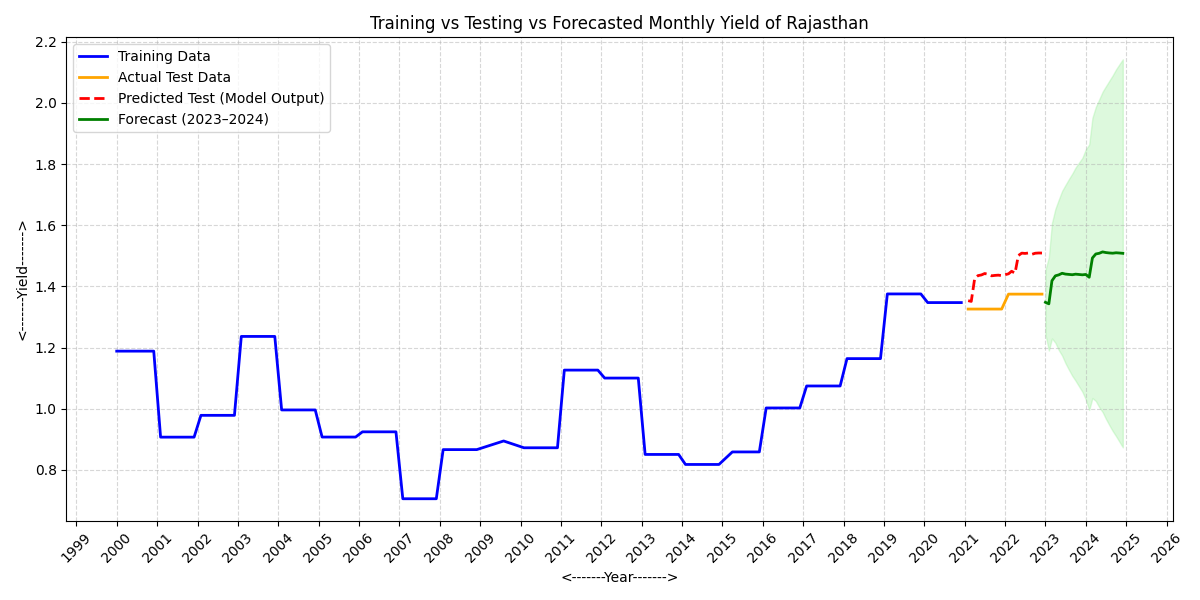


The residuals are usually distributed around zero. This means the model errors are random, and the model fits the data well.

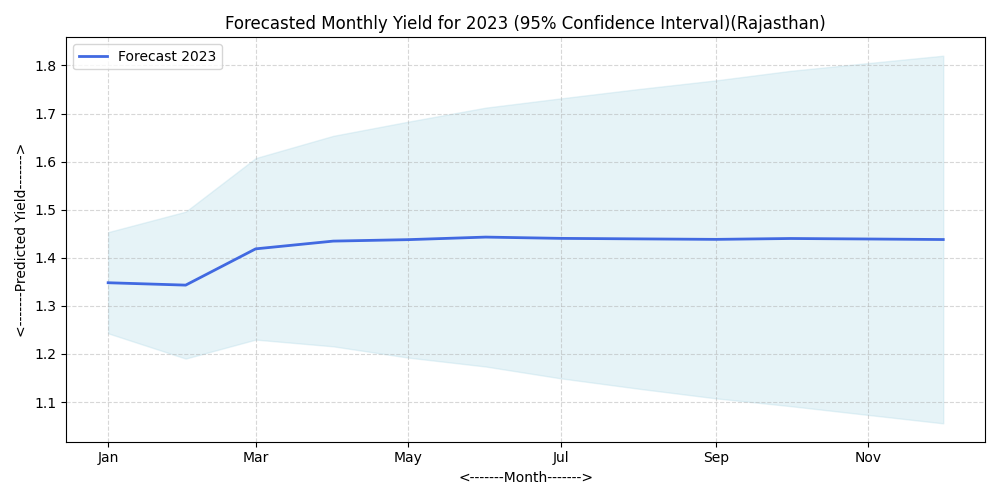


Training, Testing and Forecast Results

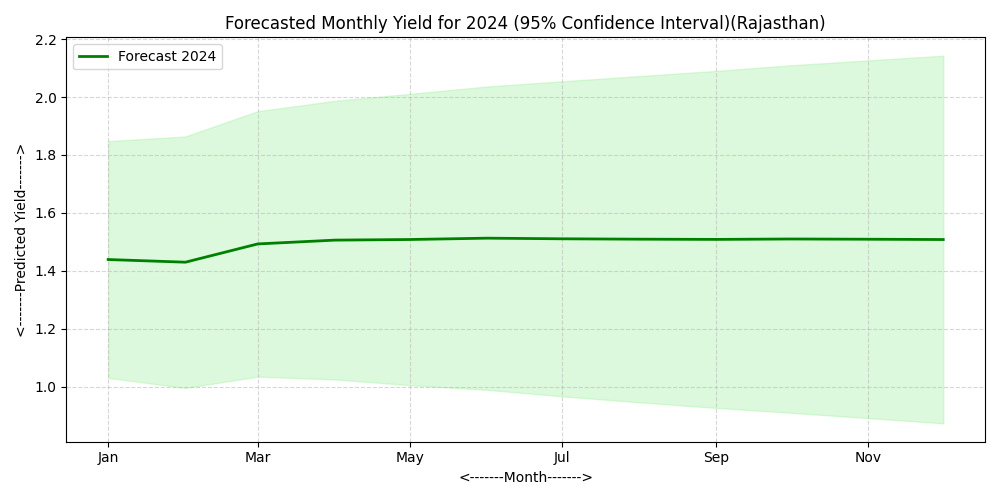
The graph displays the historical monthly yield using training data, along with the actual test values and model predictions. The predicted values closely match the actual test data, indicating strong model performance. The model is used to forecast the yield for 2023, 2024. It shows a slight upward trend in future yield, with uncertainty bands indicating the confidence range of the forecast.



The graph displays the monthly yield forecast for Rajasthan in 2023 . It shows a small increase at the beginning of the year, followed by a stable yield later on, all within a 95% confidence range.

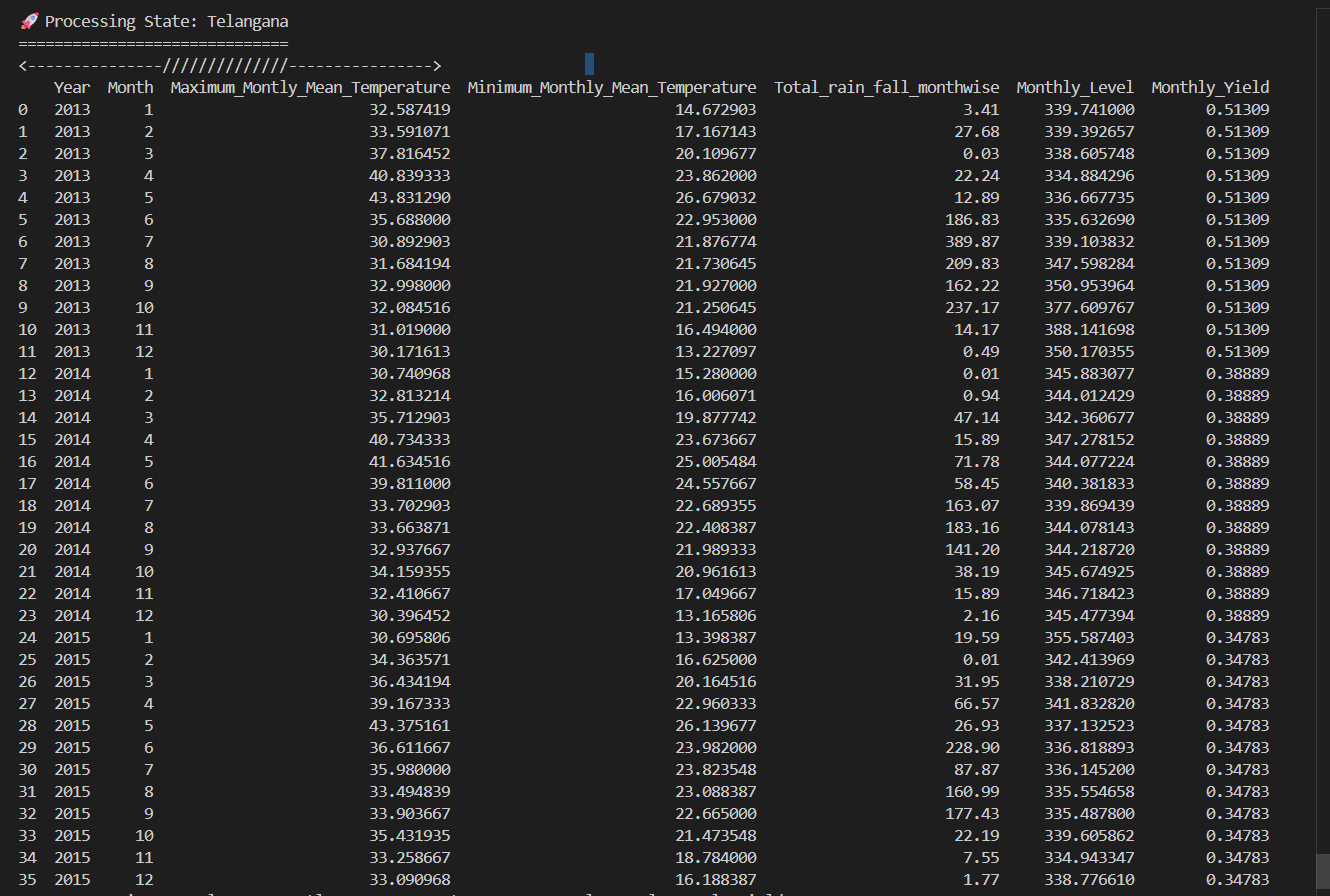


The graph displays the monthly yield forecast for Rajasthan in 2024 . It shows a small increase at the beginning of the year, followed by a stable yield later on, all within a 95% confidence range.



**Dataset Description (Telanganna)**

The Telangana dataset has monthly records of climate and crop-related variables from 2013 to 2022. It includes the highest and lowest mean temperatures, total monthly rainfall, reservoir levels, and monthly yield. These variables help us understand how climate conditions affect crop yield trends over time.

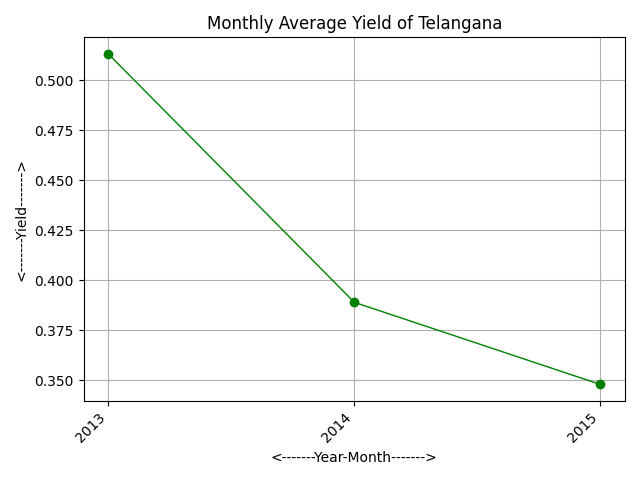


After removing outliers, we kept 34 clean monthly observations for analysis. This makes yield forecasting more reliable and precise.

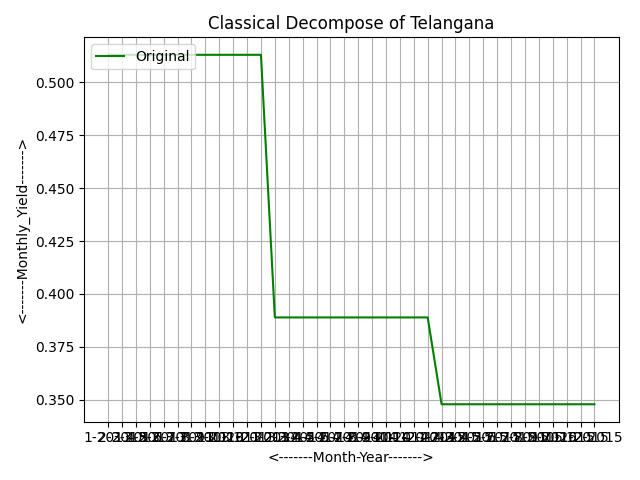
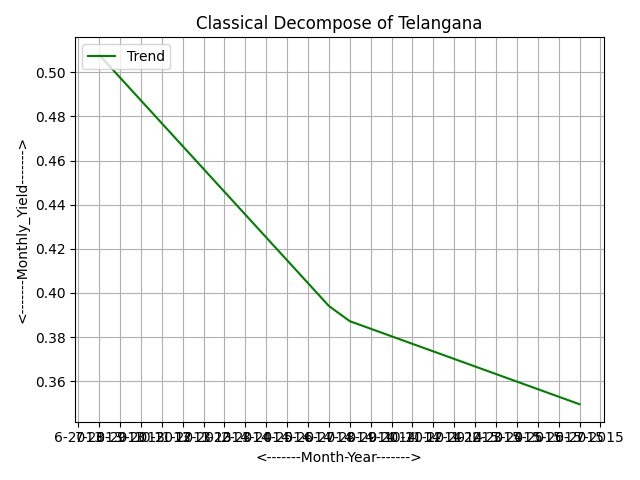
A screenshot of a computer

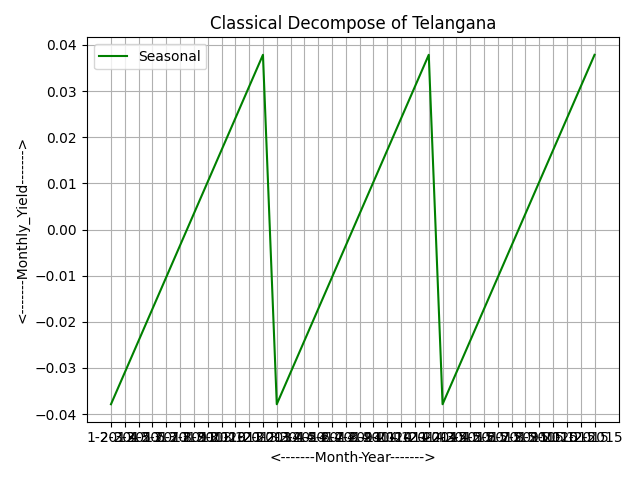
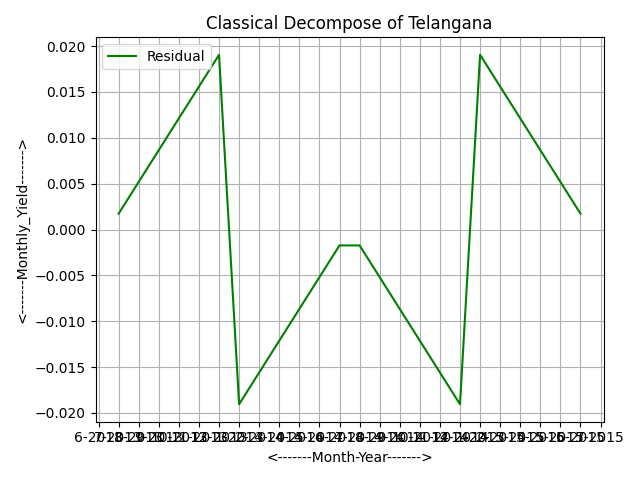
AI-generated content may be incorrect.

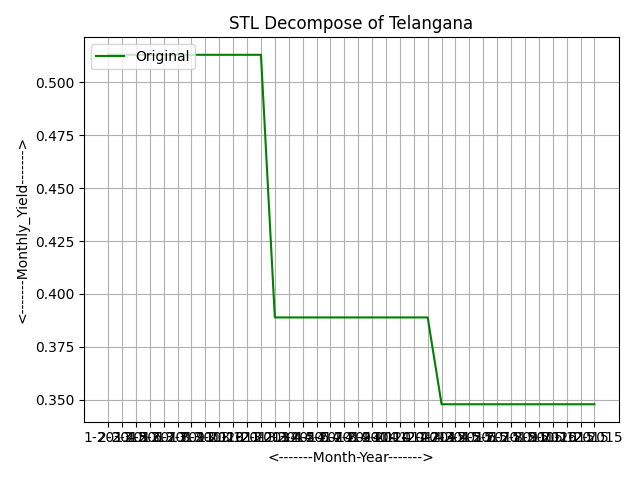
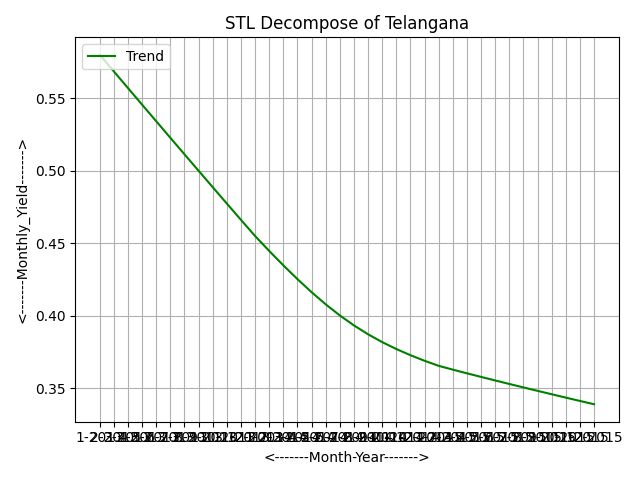
The average monthly yield of massor (lentil) in Telanganna  **from 2013 to 2015 is shown in the graph.**

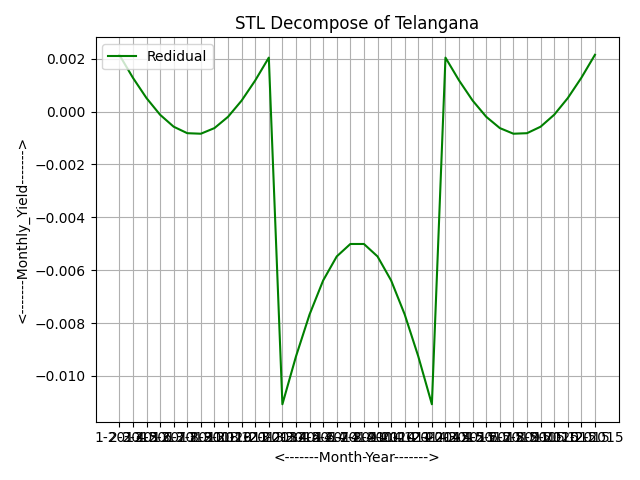


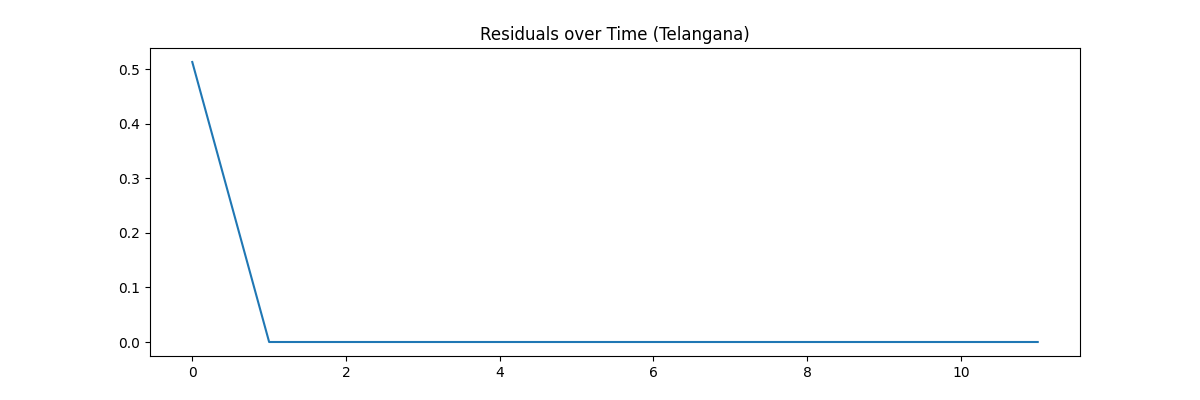
Time series Classical Decomposition plot and Time series STL Decomposition plot-

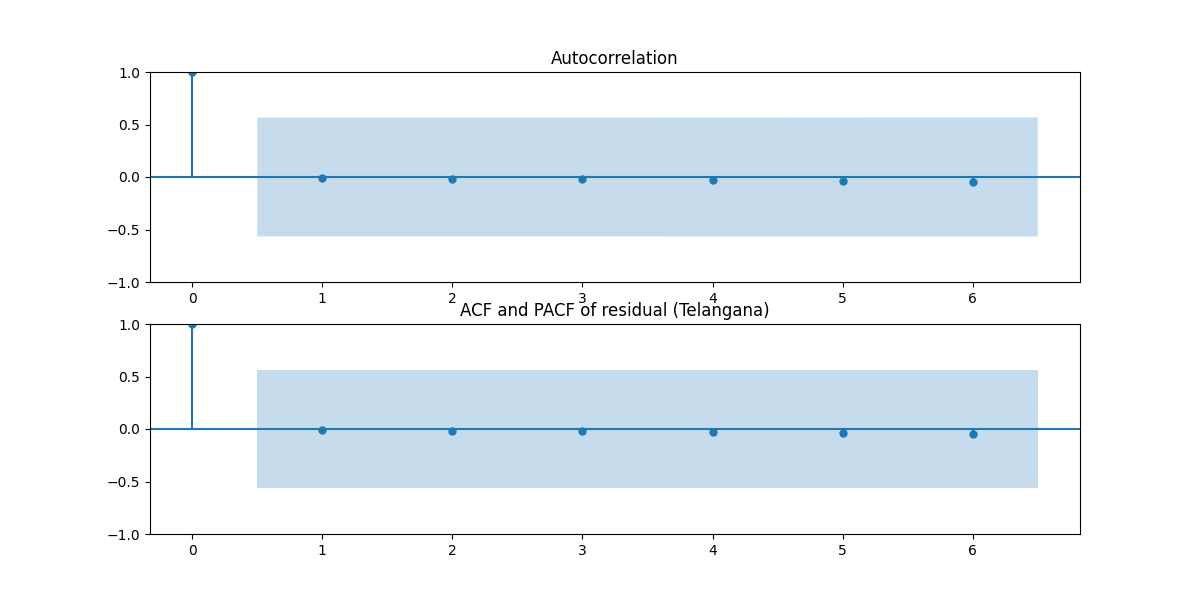
 

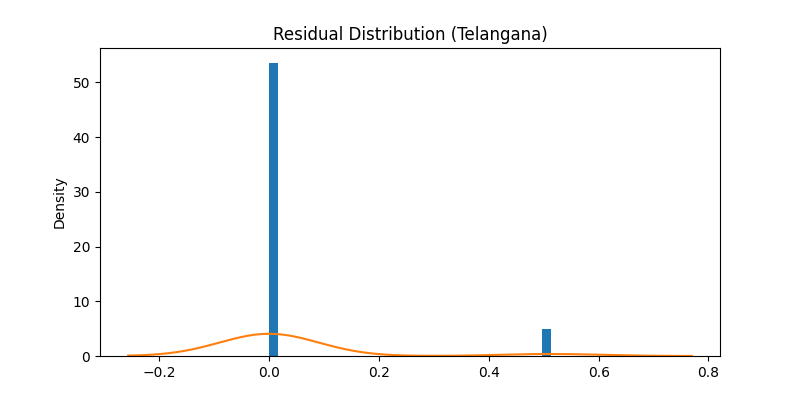
 

The ACF and PACF plots of the residuals show no significant spikes. This indicates that the residuals are white noise. This confirms that the model fits well and is suitable for forecasting.



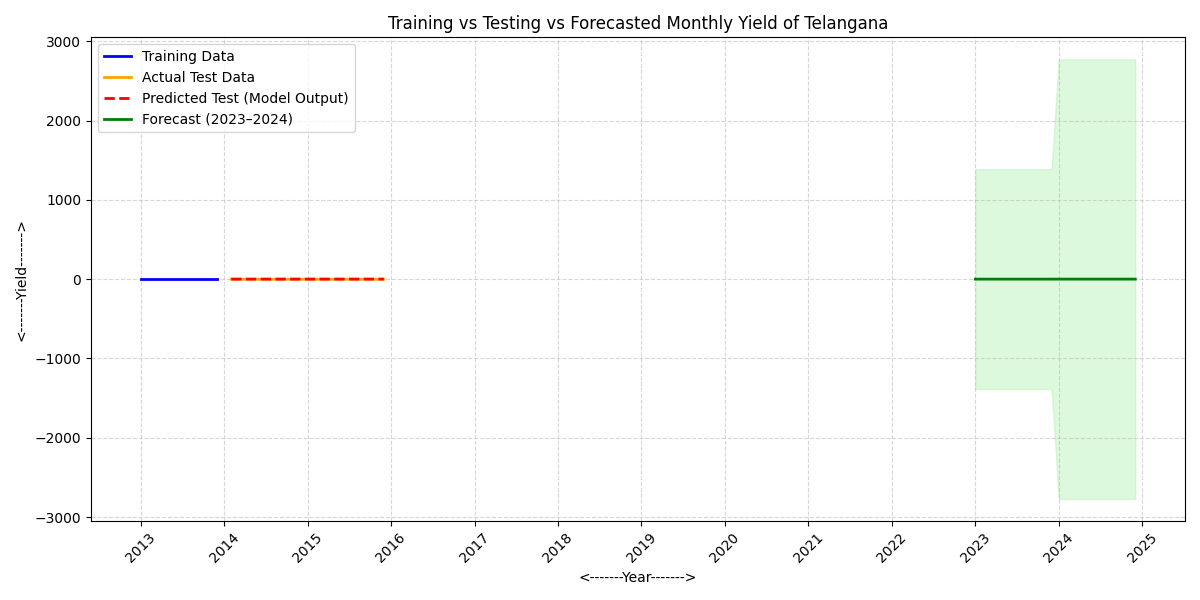
**Residual Distribution (Telangana)**

The residuals are mainly close to zero, which shows that the prediction errors are small. This indicates that the model fits the data fairly well, without significant systematic bias.

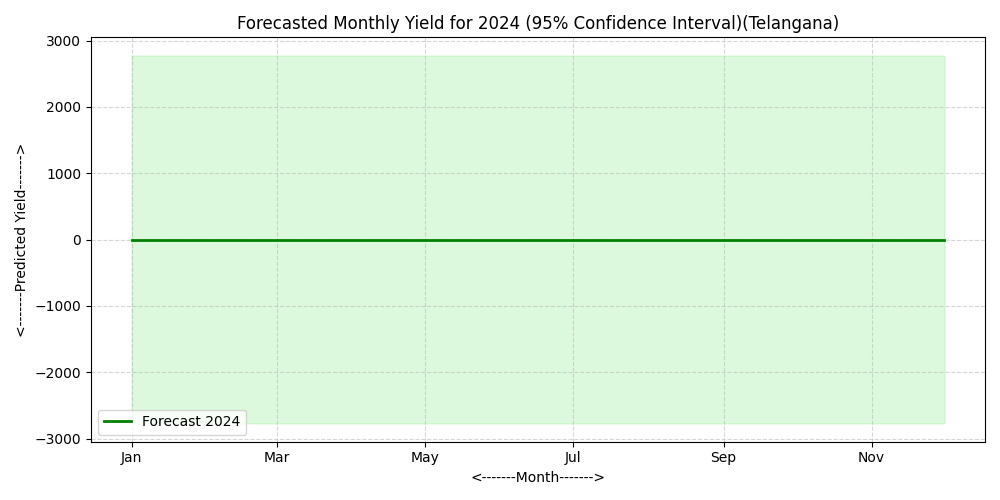


**Training vs Testing vs Forecasted Yield (Telangana)**

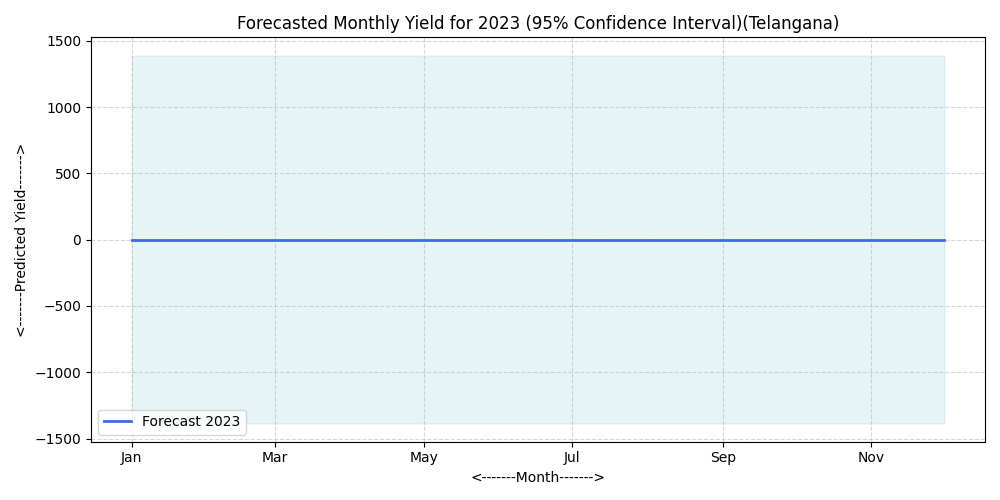
The training and testing datasets show consistent yield patterns. The predicted test values closely follow the actual test data, which indicates a good model fit. The forecast for 2023 to 2024 shows the expected future yield range, suggesting reasonable stability but with some uncertainty in the prediction interval.



The graph displays the monthly yield forecast for Telangana in 2023. It shows steady yield values for the entire year, with no significant changes.

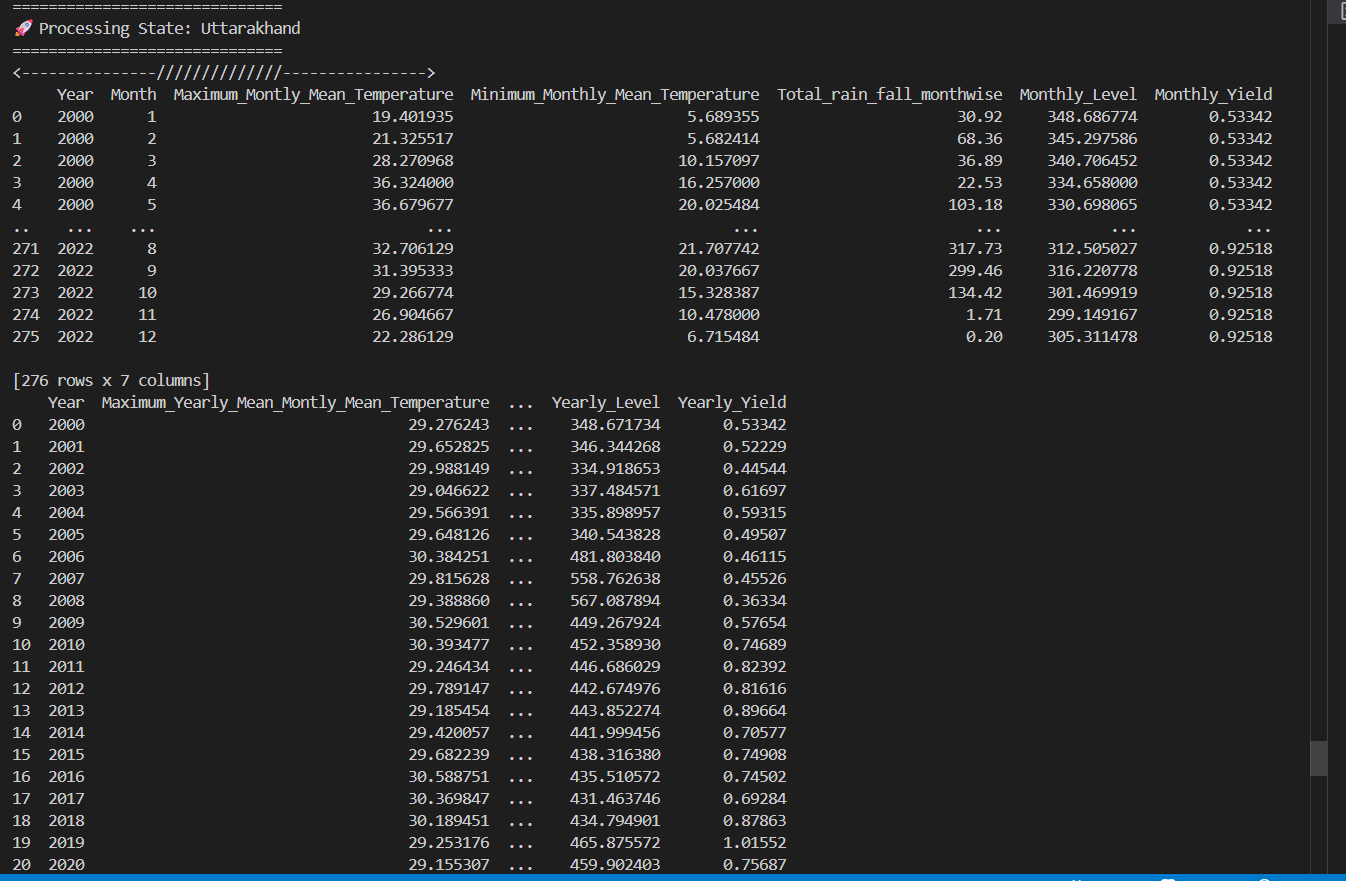


The graph displays the monthly yield forecast for Telangana in 2024. It shows steady yield values for the entire year, with no significant changes.

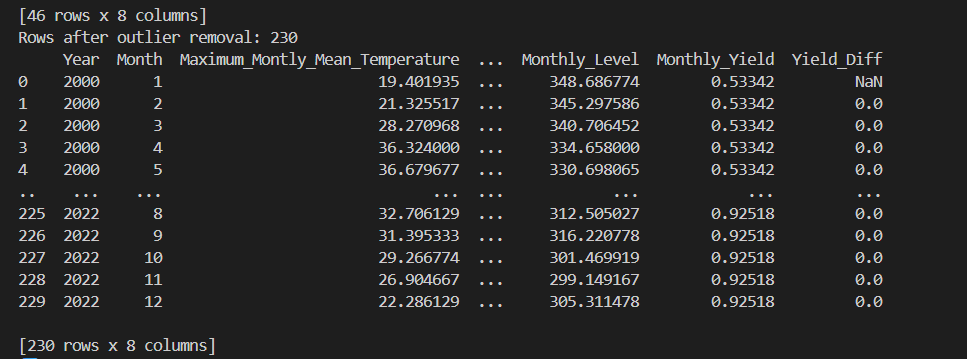


**Dataset Description (Uttarakhand)**

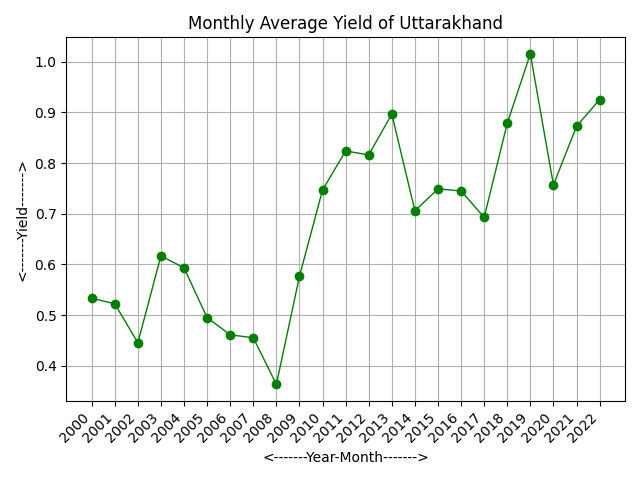
The dataset has monthly and yearly climate and agricultural information for Uttarakhand from 2000 to 2022. It includes details like the highest and lowest monthly average temperature, total monthly rainfall, water level, and crop yield. The yearly dataset summarizes these figures into annual averages for temperature, rainfall, level, and yield. This data helps examine how climate factors affect crop productivity for forecasting.

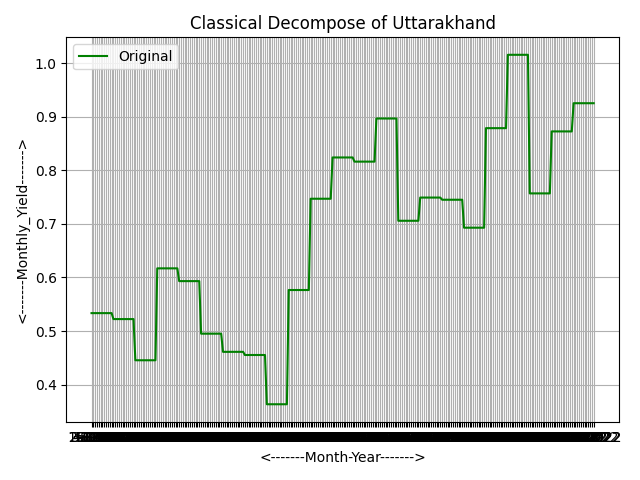


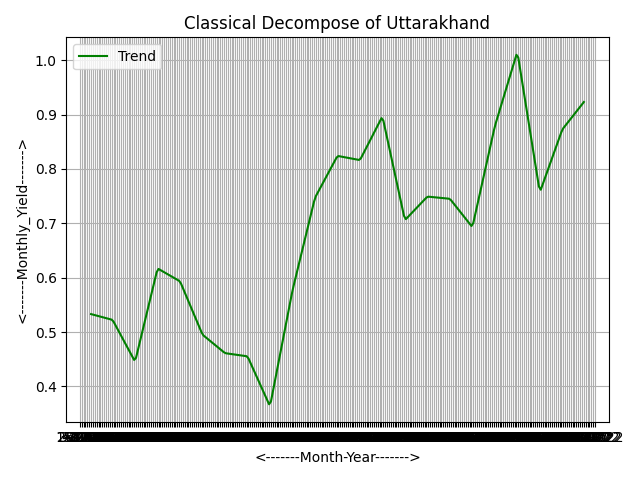
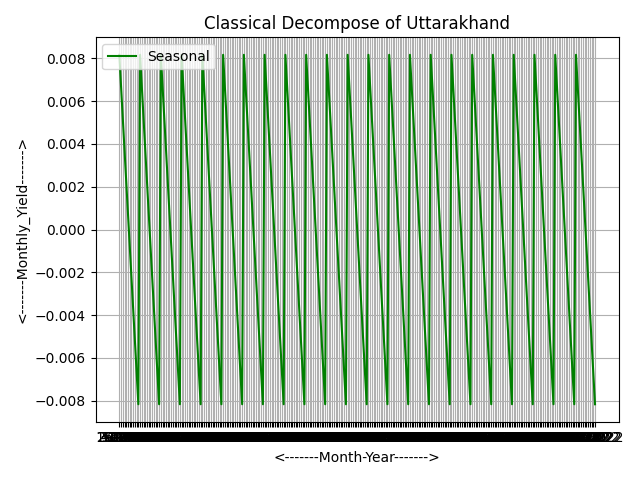
After removing outliers, the dataset has 230 monthly records from 2000 to 2022 for Uttarakhand. It includes variables like maximum monthly mean temperature, minimum temperature, total rainfall, water level, crop yield, and yield difference. The cleaned dataset offers reliable and consistent data for time series analysis and forecasting of crop yield.

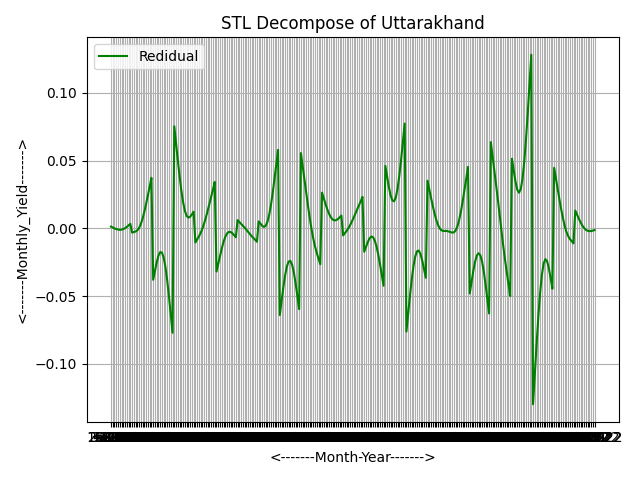
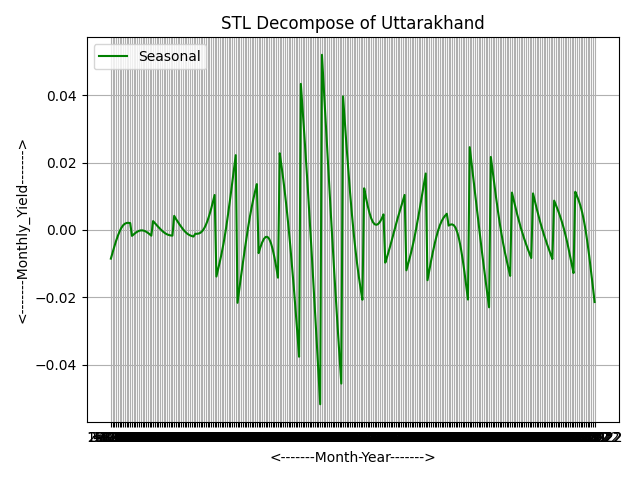
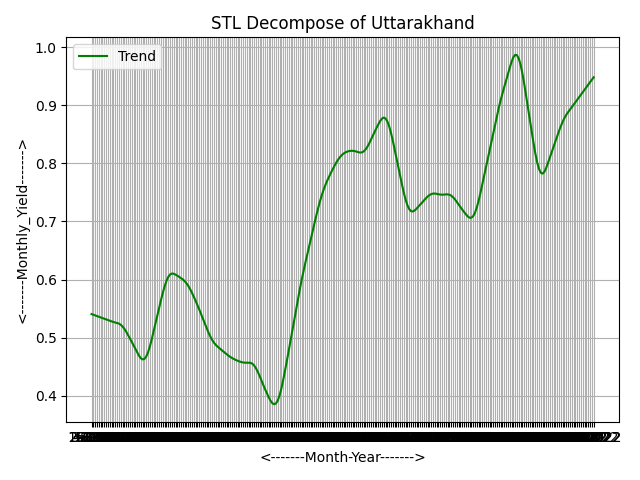
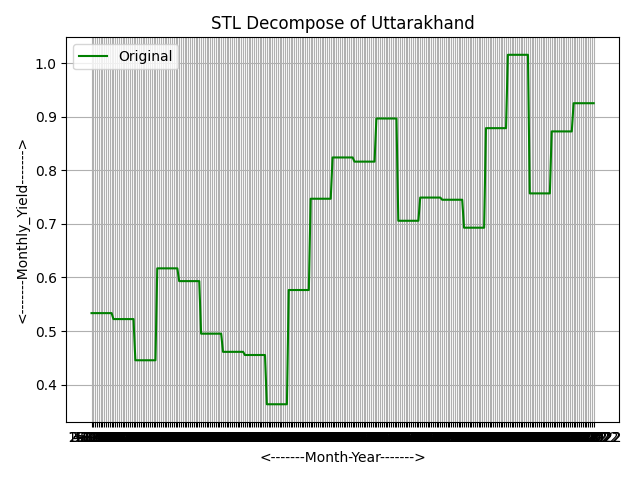
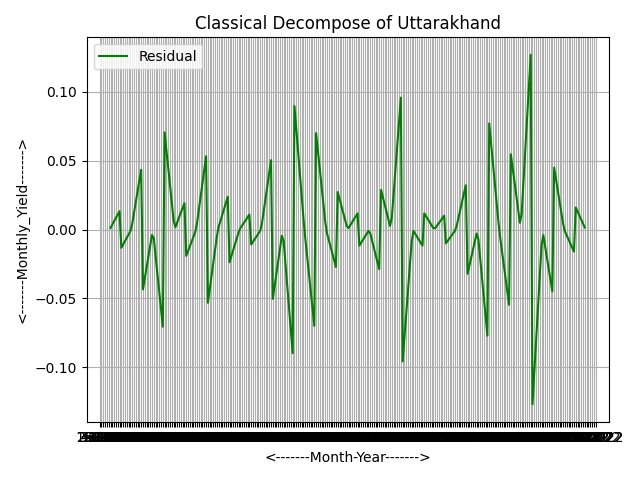


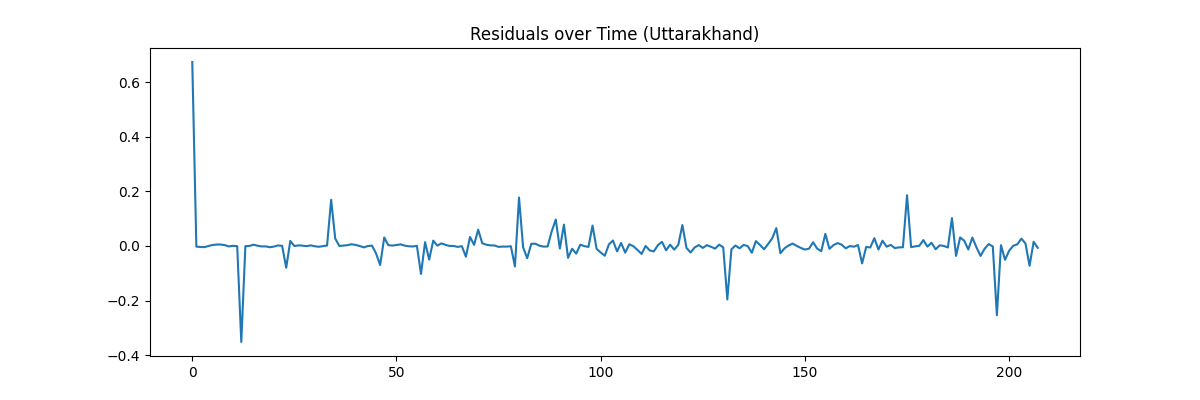
The average monthly yield of massor (lentil) in **Uttarakhand**  **from 2000 to 2022 is shown in the graph.**

Time series Classical Decomposition plot and Time series STL Decomposition plot-



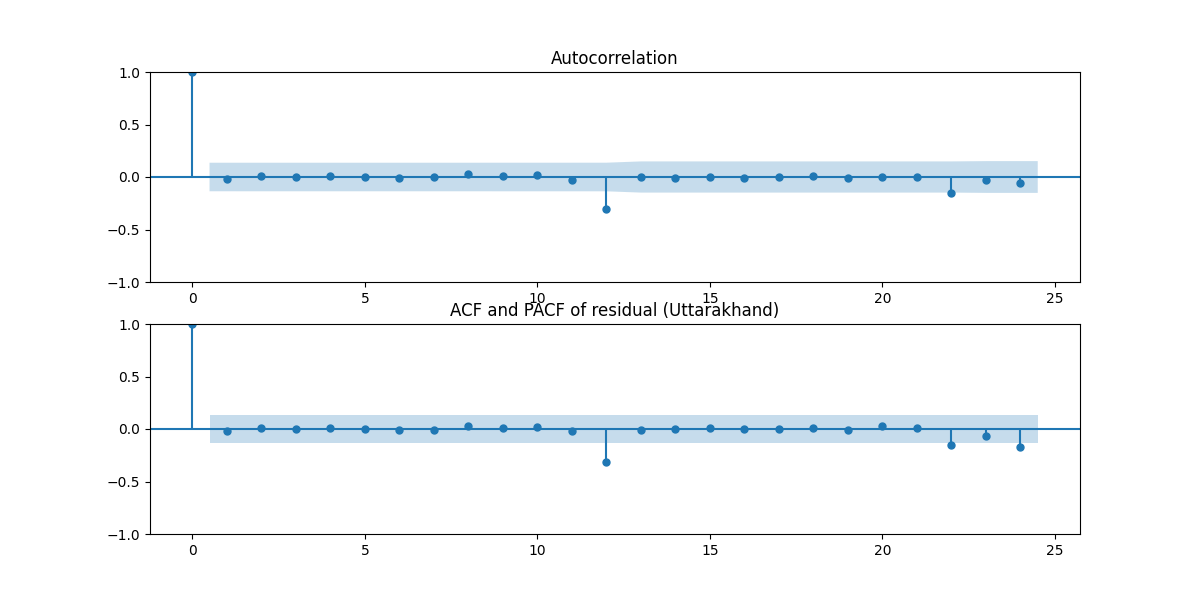
 



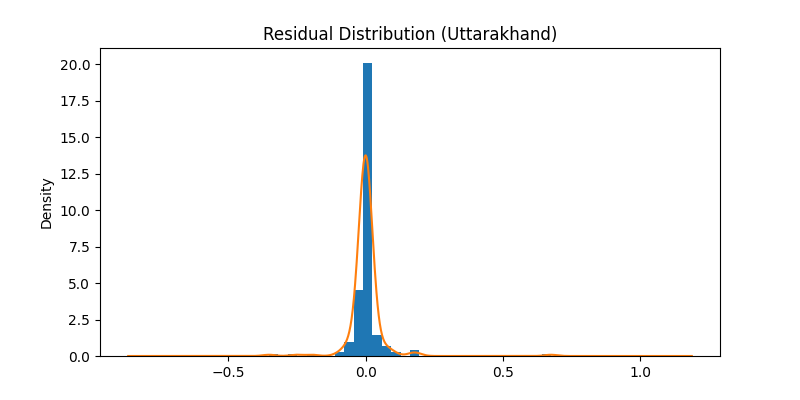


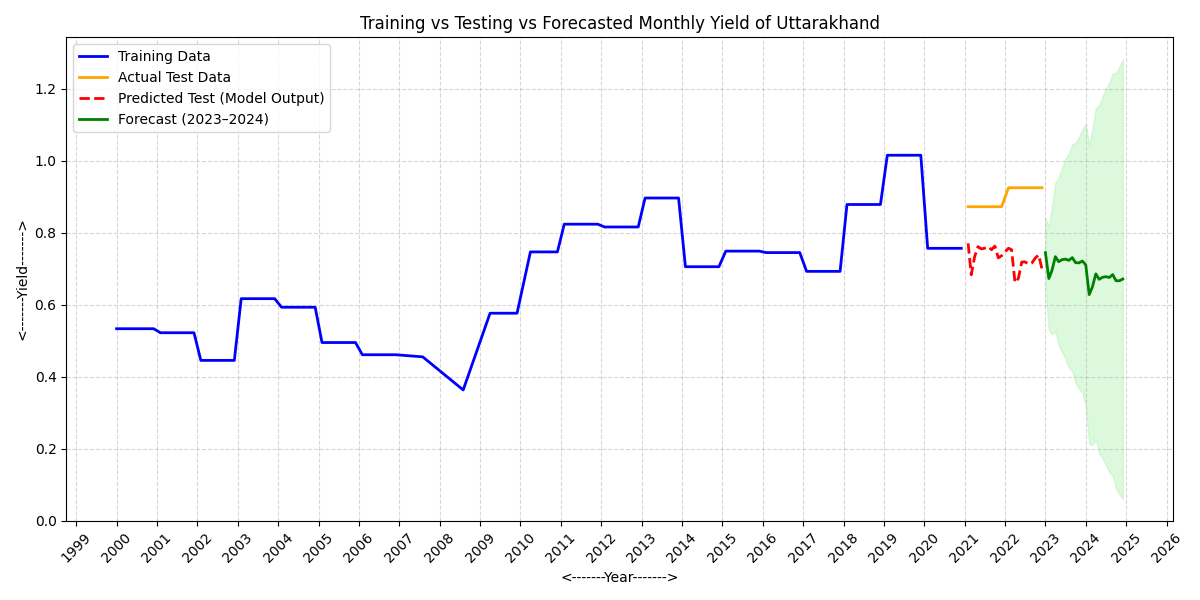
**ACF and PACF Analysis**

The ACF and PACF plots of the residuals for Uttarakhand show that most spikes fall within the confidence interval. This indicates that there is no significant autocorrelation present. It suggests that the ARIMAX model effectively captured the time-dependent patterns and that the residuals act like white noise. This confirms a good model fit.

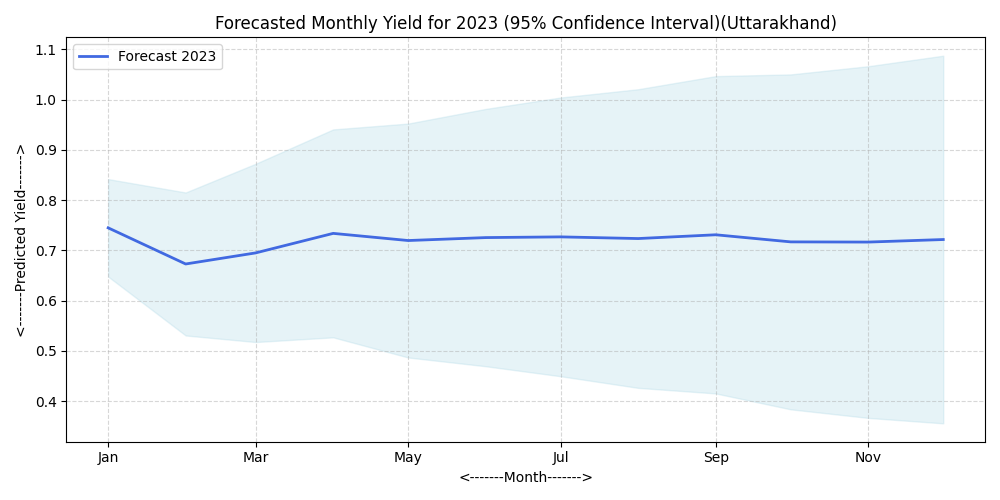


**Residual Distribution Analysis**

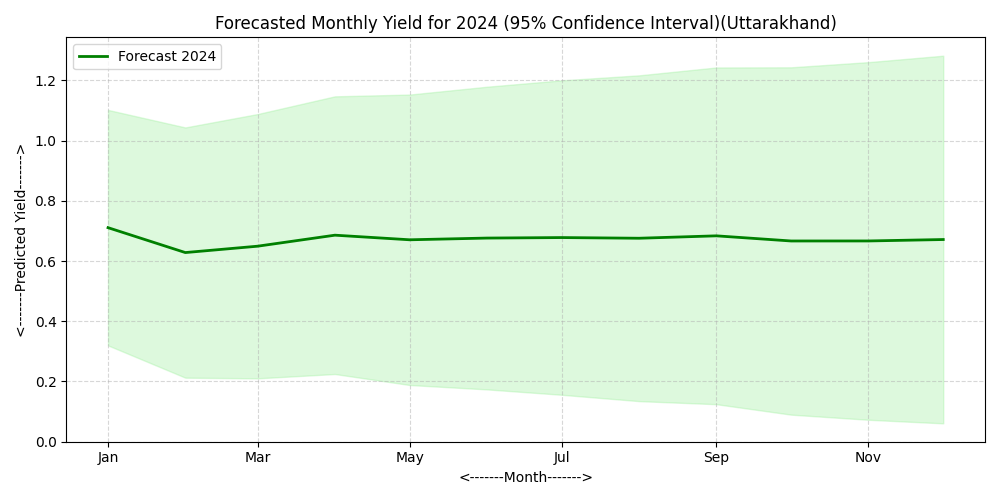
The residual distribution for Uttarakhand is roughly centered around zero. This shows that the residuals are normally distributed. This confirms that the ARIMAX model has captured the main patterns in the data. The remaining errors are random and show no systematic bias.



The graph displays the 2023 monthly yield forecast for Uttarakhand. It shows a small drop early in the year, followed by steady yield values for the rest of the months.



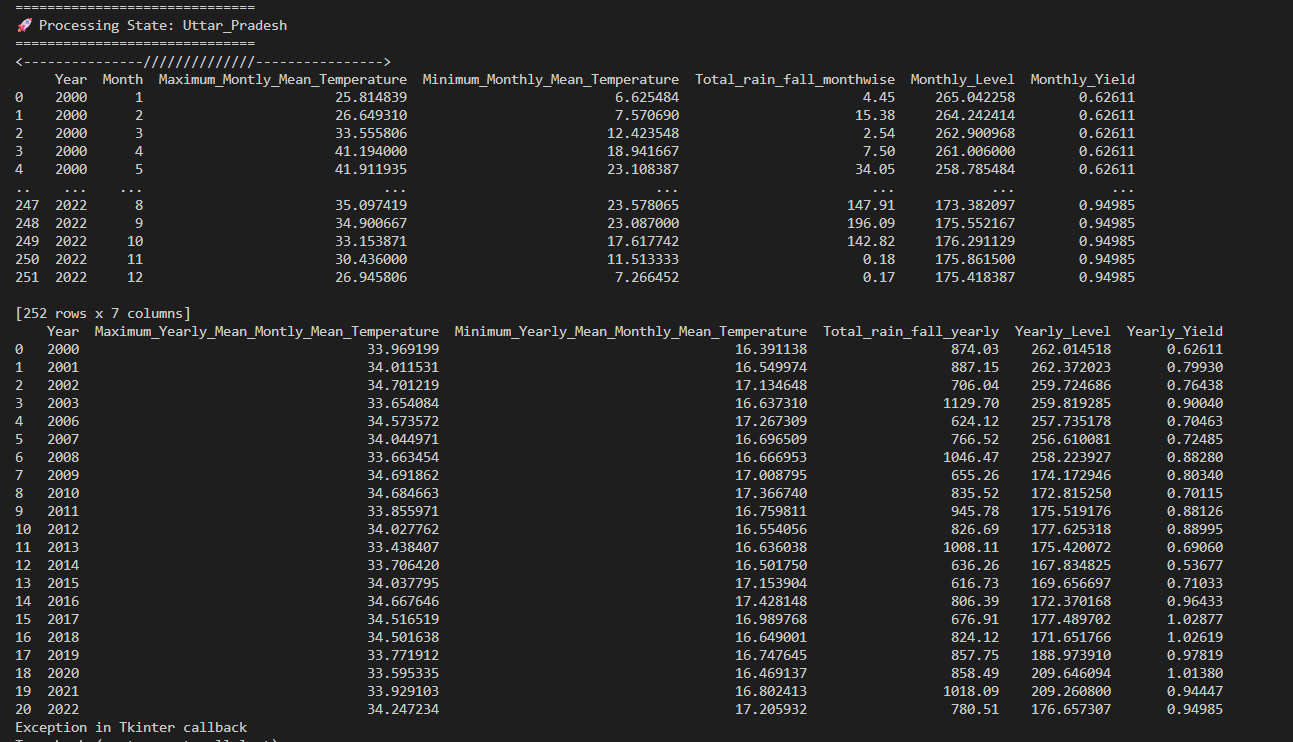
The graph displays the 2024 monthly yield forecast for Uttarakhand. It shows a small drop early in the year, followed by steady yield values for the rest of the months.



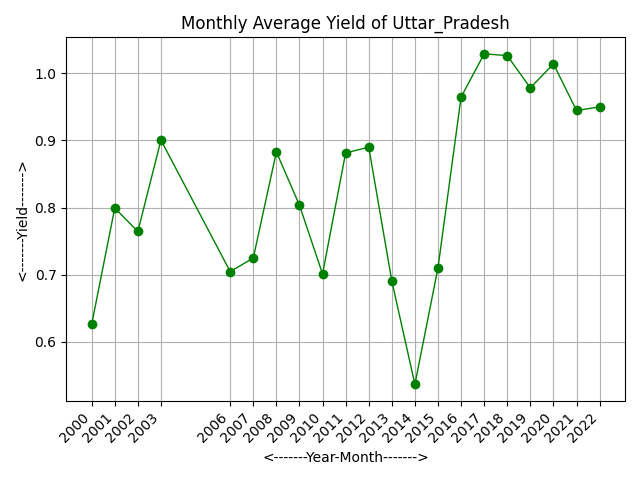
**Dataset Description (Uttar\_Pradesh)**

The dataset for Uttar Pradesh includes monthly and yearly records from 2000 to 2022, covering meteorological and agricultural variables. The parameters are maximum and

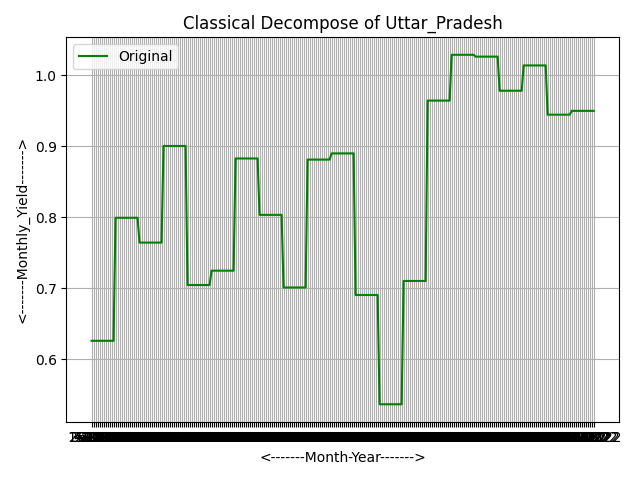
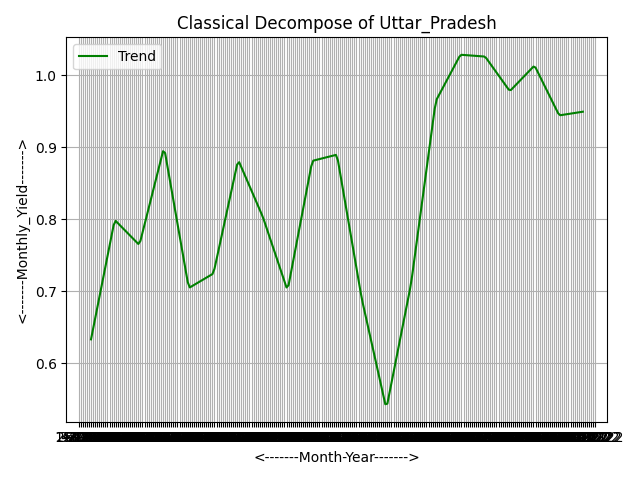
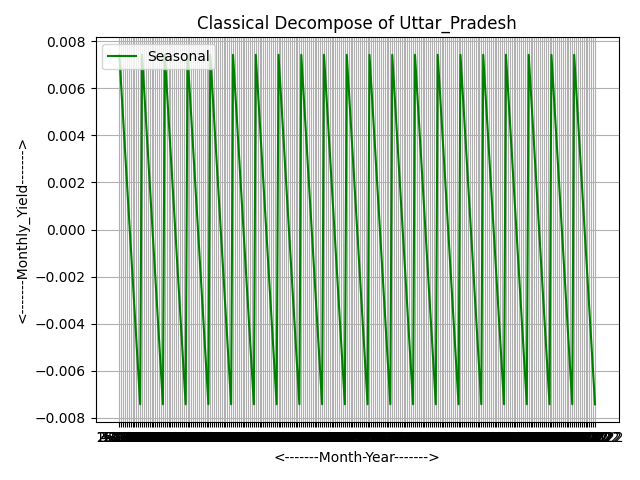
minimum mean temperatures, total monthly and yearly rainfall, and crop yield. Researchers combined monthly data to create yearly averages for analysis.



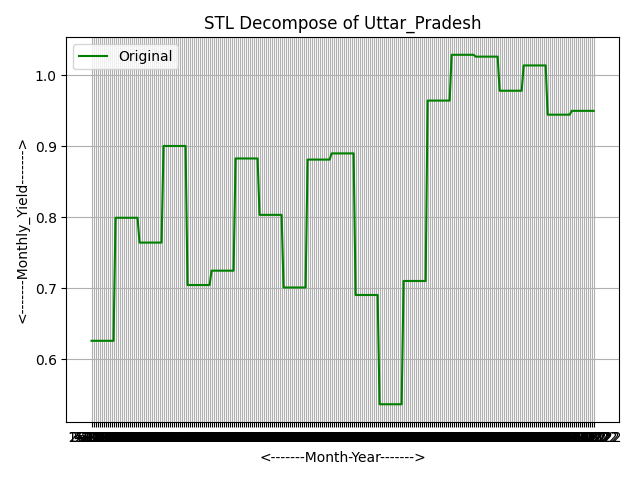
The average monthly yield of massor (lentil) in **Uttar\_pradeshs from 2000 to 2022 is shown in the graph.**



Time series Classical Decomposition plot and Time series STL Decomposition plot-

A graph showing a number of green lines

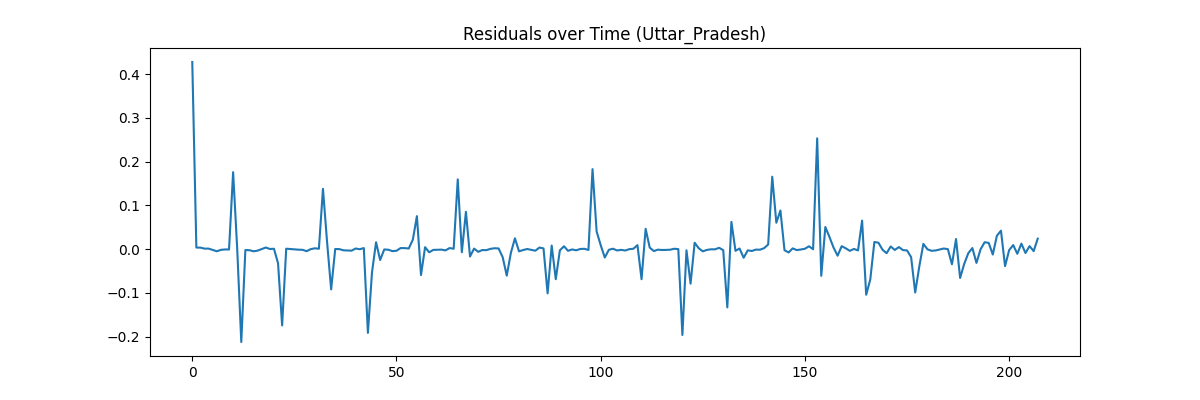
AI-generated content may be incorrect.  A graph showing a trend

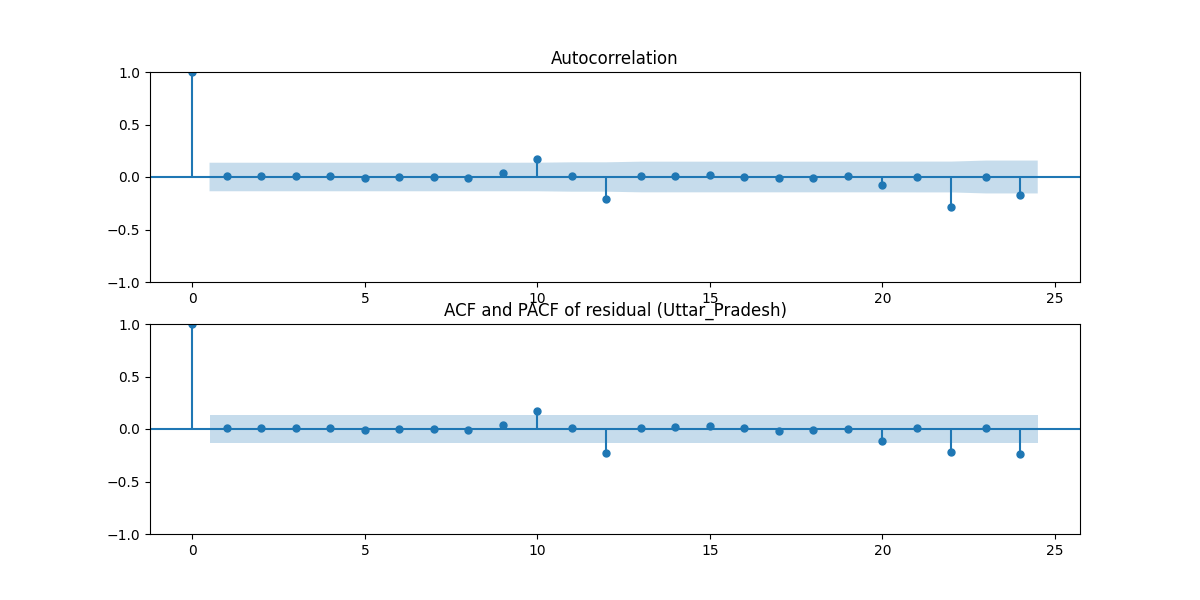
AI-generated content may be incorrect.

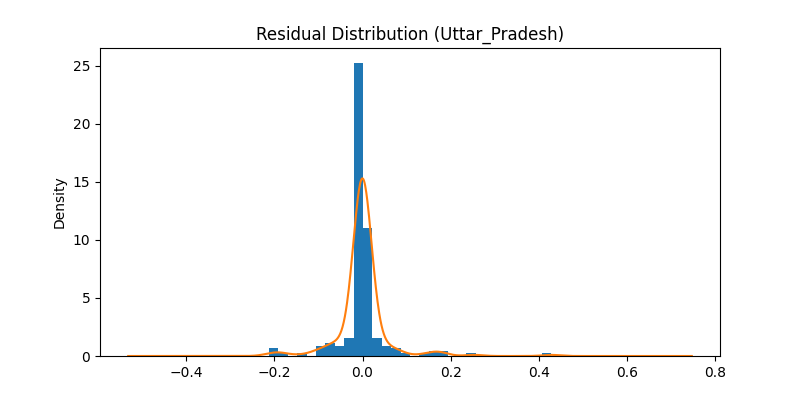
A graph showing a seasonal graph

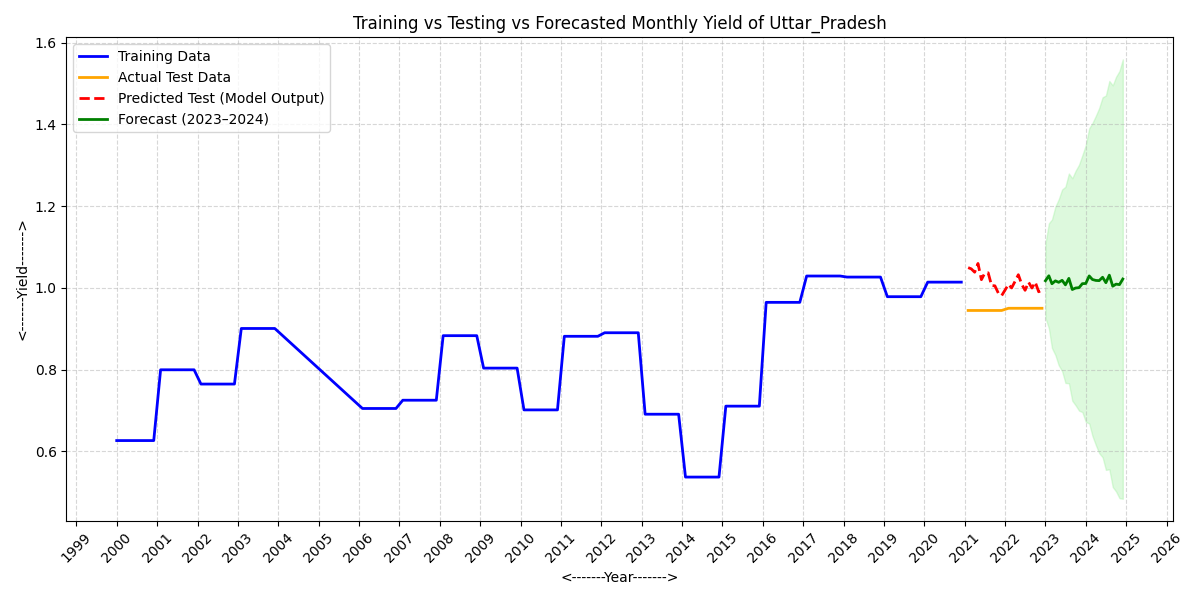
AI-generated content may be incorrect. A graph showing a number of green lines

AI-generated content may be incorrect.

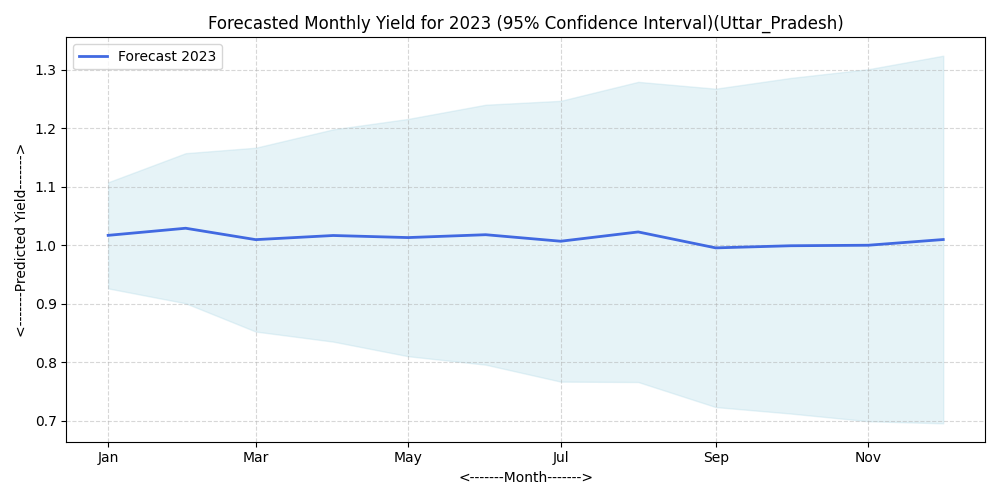
The ACF and PACF plots for Uttar Pradesh show that all residual values are within the confidence limits. This means there is no significant autocorrelation. This confirms that the ARIMAX model fits the data well and that the residuals behave like white noise.

The residual distribution for Uttar Pradesh is almost normal and centered around zero. This indicates that the ARIMAX model fits well and has few prediction errors.

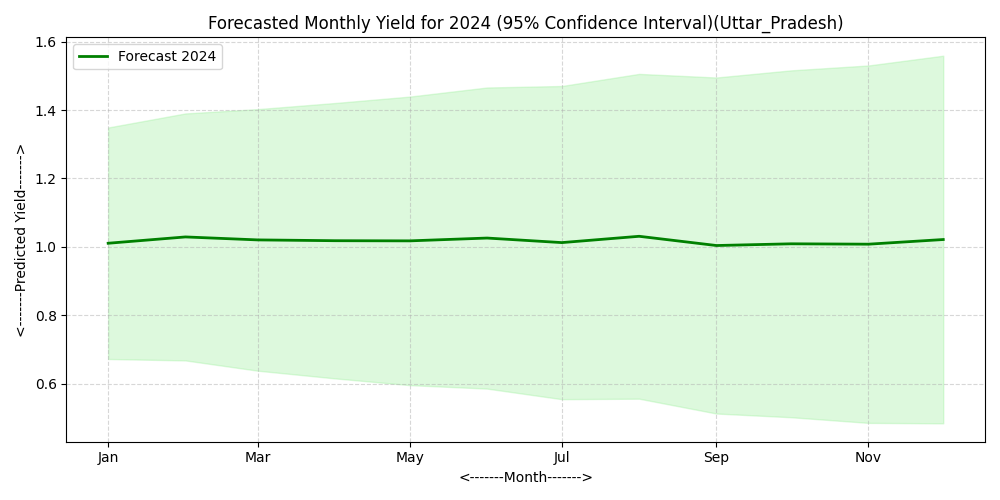
The graph displays the training, testing, and predicted monthly yield of Uttar Pradesh. The ARIMAX model closely follows the test data trend and predicts a stable yield for 2023, 2024. This suggests consistent production with minor variations.



The graph displays the 2024 monthly yield forecast for Uttar Pradesh. It shows a steady yield trend with small changes over the year.

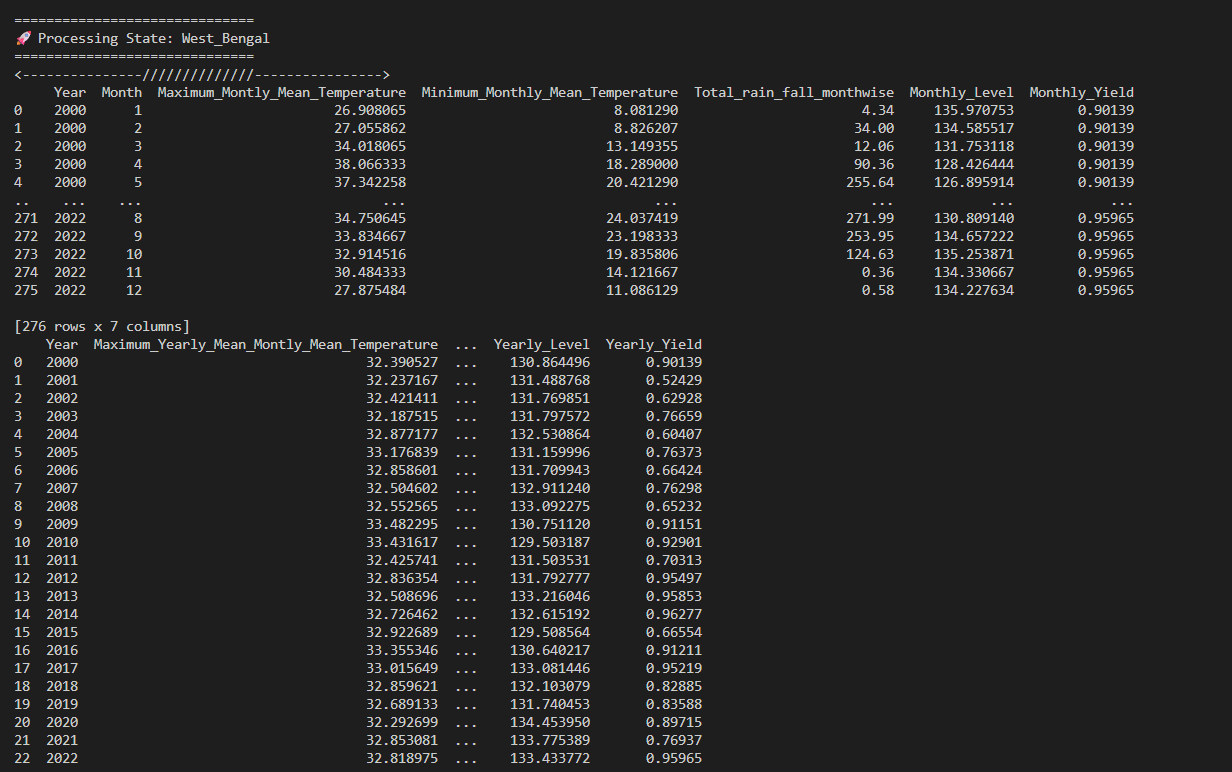


The forecast for Uttar Pradesh in 2024 shows a steady monthly yield, with small changes. This indicates reliable agricultural performance all year round.

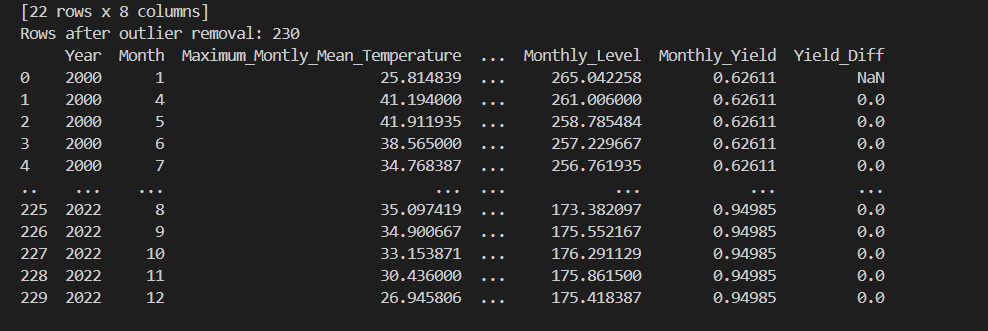


**Dataset Description (West Bengal** **)**

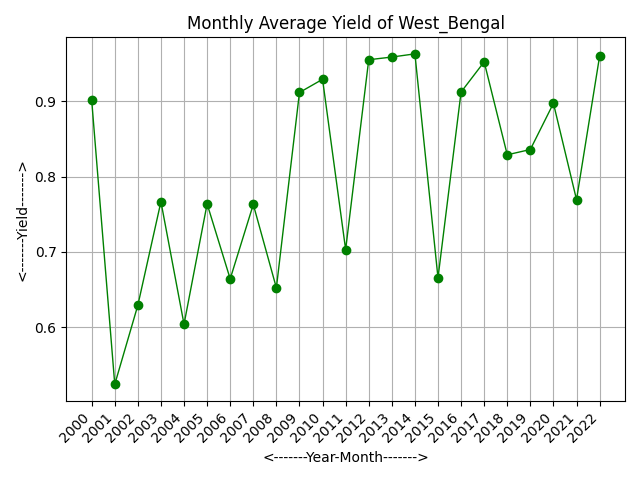
The dataset provides monthly and yearly climate and yield information for West Bengal from 2000 to 2022. It includes details like maximum and minimum mean temperatures, total monthly rainfall, and related crop yield and water level data. The monthly dataset contains 276 records that summarize trends in temperature, rainfall, and yield for each month. The yearly dataset has 23 records that offer annual averages for the same parameters. This data helps to study the relationship between climate factors and variations in massor yield over time.



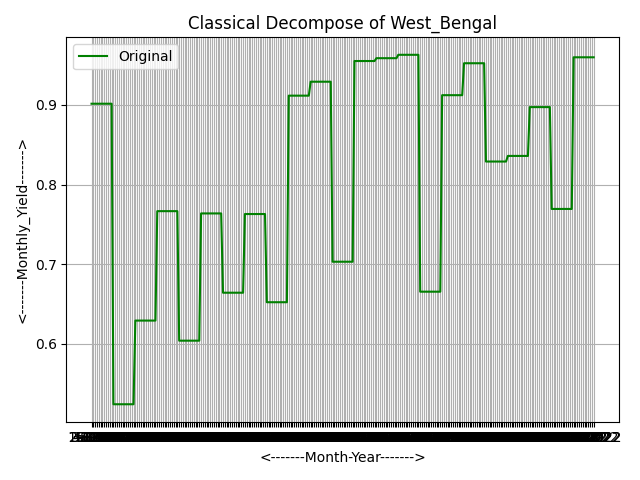
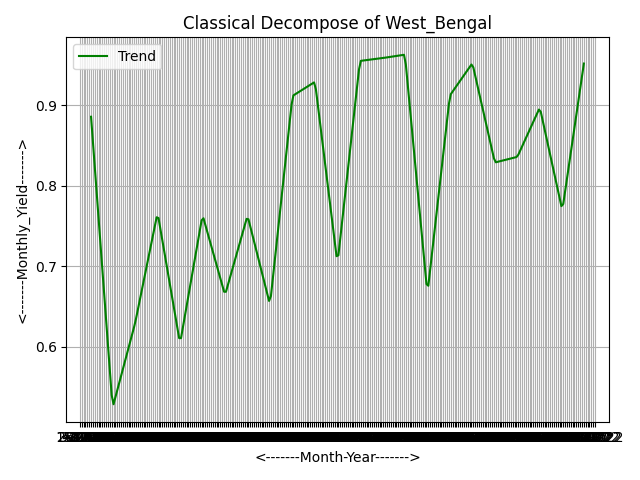
After removing outliers, 230 records remained. These records show monthly temperature, rainfall, and yield data from 2000 to 2022. This process ensures cleaner and more reliable data for analysis.

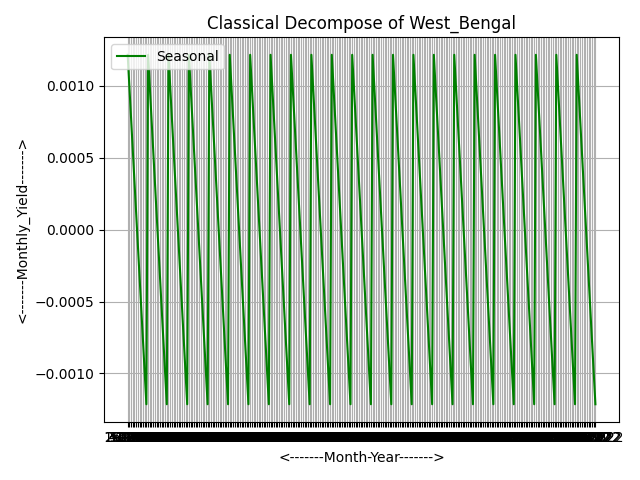
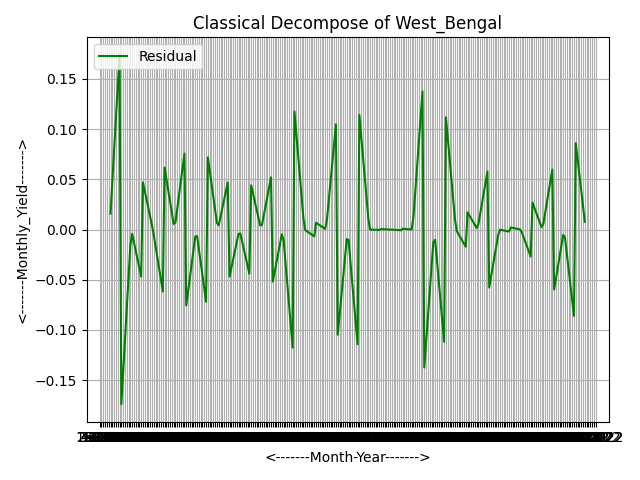


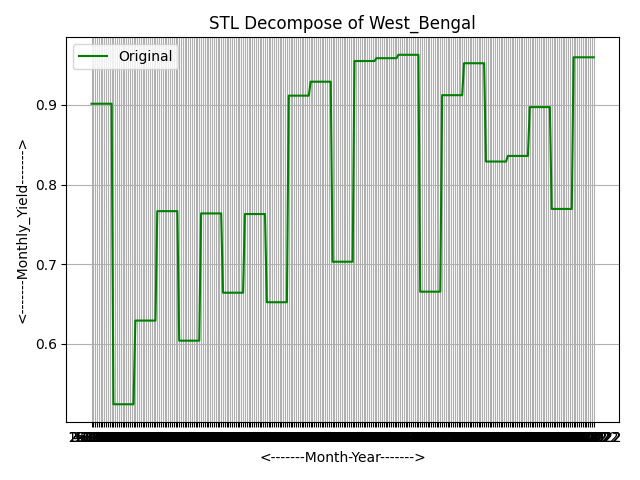
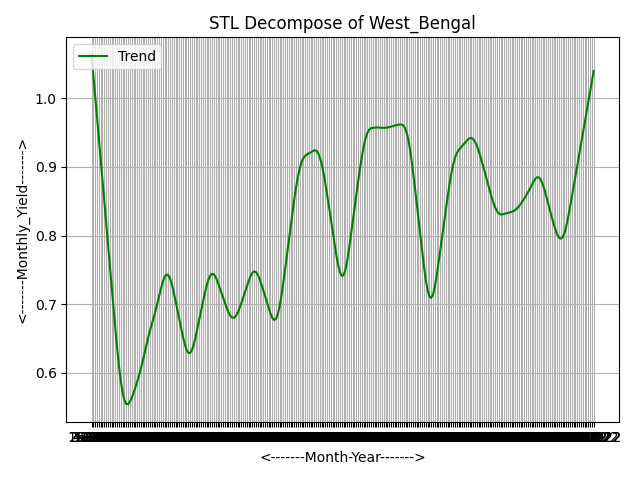
The average monthly yield of massor (lentil) in**West\_Bengal from 2000 to 2022 is shown in the graph.**

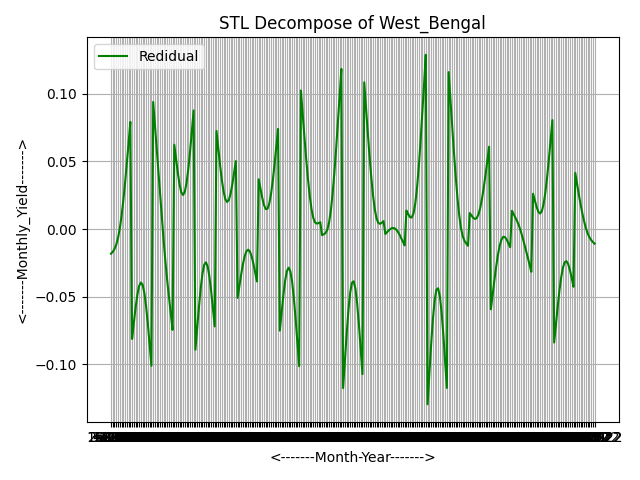
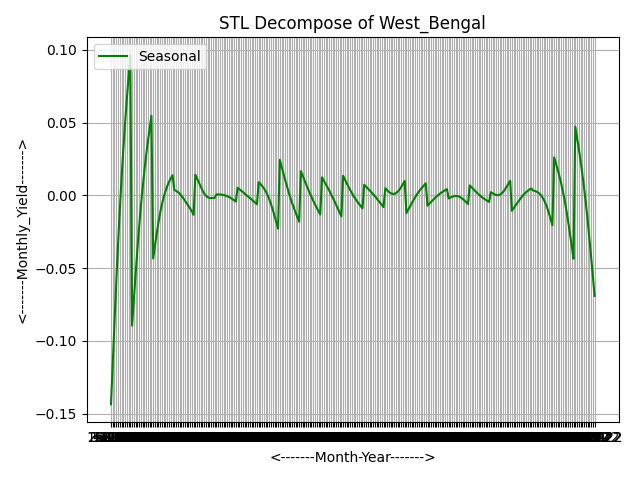


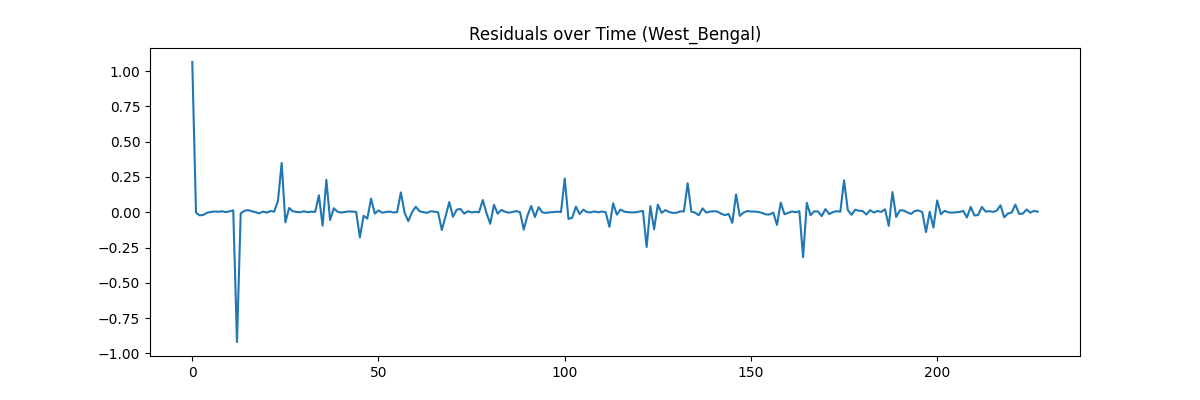
Time series Classical Decomposition plot and Time series STL Decomposition plot-

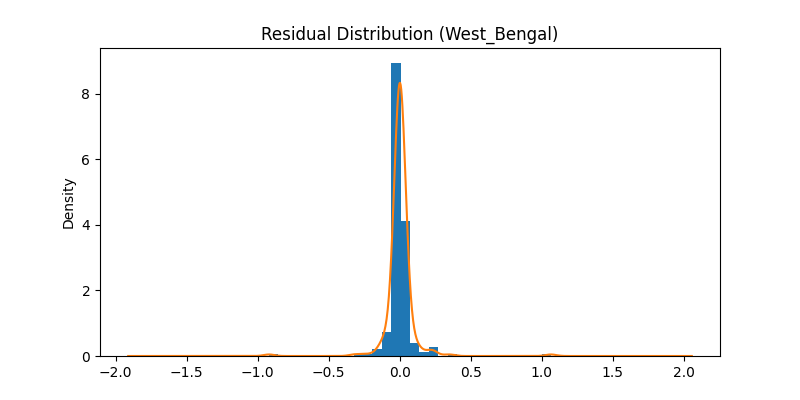
 

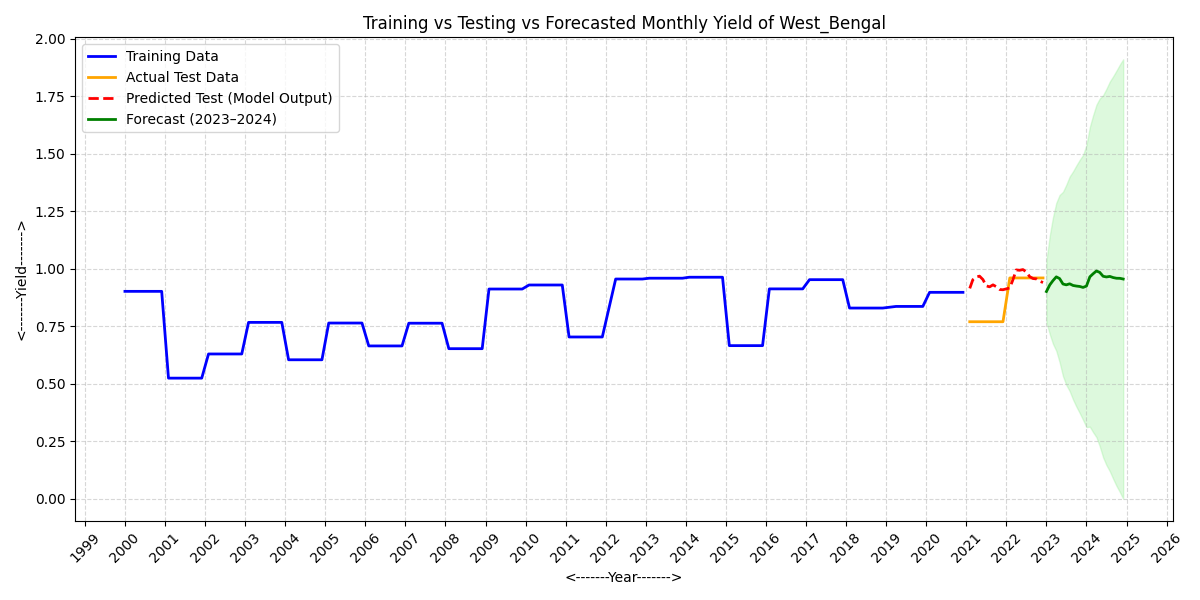




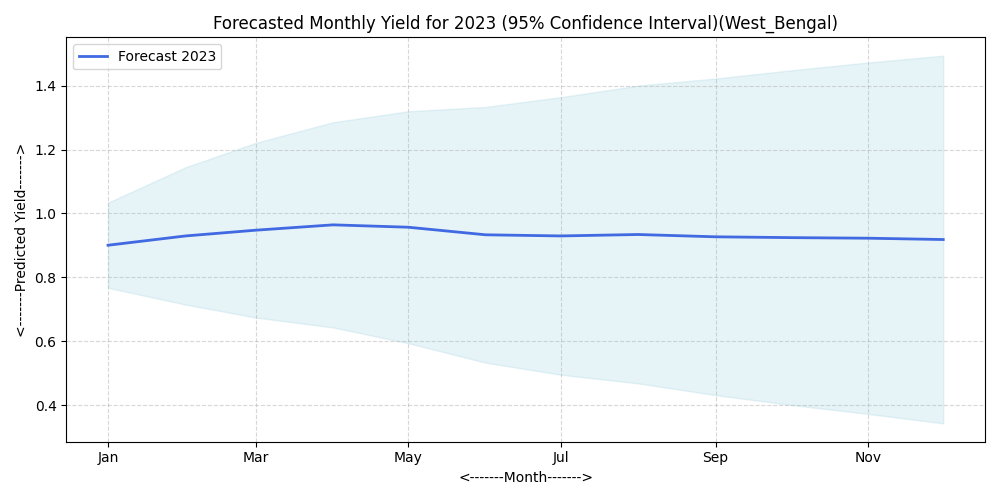
The residual distribution for West Bengal is centered around zero and has a nearly normal shape. This suggests that model errors are random and that the model performs well.



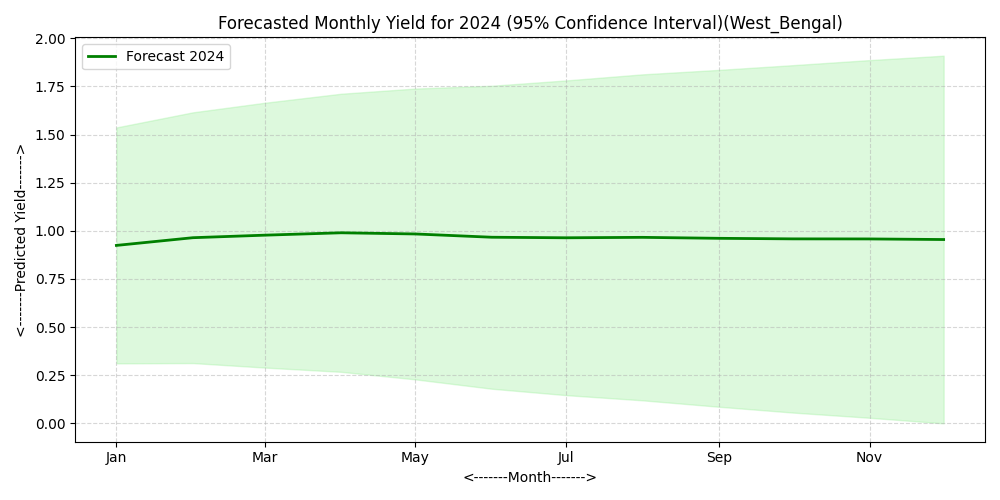
The graph displays the training, testing, and predicted monthly yield for West Bengal. The ARIMAX model closely matches the actual data and predicts a stable yield trend for 2023, 2024. This shows consistent agricultural performance with minimal fluctuations.



The graph shows the expected monthly yield for West Bengal in 2023, along with a 95% confidence interval. The yield displays a steady trend over the year, with small changes that suggest consistent production and limited seasonal variation.



The predicted monthly yield for West Bengal in 2024 shows a consistent trend with little variation. This suggests stable crop production throughout the year within the 95% confidence interval.



1. Conclusion

The project titled “Forecasting Massor Yield of Various States for the Years 2023 and 2024 Using Time Series Analysis Models (ARIMAX)” aimed to predict future yields of massor (lentil) across different Indian states using historical data and climatic factors such as rainfall and temperature.

The ARIMAX model effectively captured the link between yield and climatic factors, providing reliable forecasts for each state. The model results showed that most states are likely to have stable or slightly increasing yield trends during 2023 and 2024. This suggests favorable growing conditions and consistent farming practices. States like Uttar Pradesh and West Bengal exhibited steady yields. In contrast, states such as Rajasthan and Uttarakhand showed slight changes due to climatic influences.

The residual and diagnostic plots confirmed that the model assumptions were met, which indicates a good model fit and reliable predictions within the 95% confidence interval.

**Key Findings**

1. ARIMAX models successfully included external factors to improve forecast accuracy.
2. Predicted yields remained stable, showing strong agricultural performance.
3. States with more rainfall and moderate temperatures had better yield prospects.

**Recommendations for Future Work**

1. Add more predictors like soil fertility, fertilizer use, and technology inputs to improve model accuracy.
2. Expand the model to long-term forecasting for policy planning and crop management.
3. Create interactive dashboards for real-time yield monitoring and climate impact analysis.
4. Compare ARIMAX results with other machine learning models (e.g., LSTM, Prophet) for increased accuracy.

Overall, this study shows that time series forecasting with ARIMAX is a strong and useful method for predicting agricultural yields, which supports informed decision-making in crop management and planning.

**4.APPENDICES**

**1.Github Link -** [**https://github.com/RAHULBAGDI28/ISI-\_PROJECT**](https://github.com/RAHULBAGDI28/ISI-_PROJECT)