Forecasting the Production of Wheat in Different States of India Using ARIMAX Models

Subhodip Chakraborty

Department of Statistics, University of Kalyani

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

India is the 2nd largest producer of wheat in the world and it is the second most produced staple food crop of the country. The daily production of wheat in 14 different states of India, during 2000–2022 are used here to predict the same for 2023-2025.The dataset, comprising of daily data on wheat yield, temperature, rainfall, reservoir levels and some other variables, was first converted into monthly data by taking the arithmetic means of the values available for each month and then converted into annual data by adding those means. Further, it was decomposed to check the presence of trend and seasonality. Before fitting the model, it was necessary to know whether the data was stationary or not. In this regard, statistical tests for stationarity were applied, and for the states with non-stationary data, differencing were applied to make them stationary. Due to the presence of exogenous variables, such as temperature in the state, amount of rainfall, capacity of reservoir, and amount of water available, in the dataset, ARIMAX model was used. Also, the validity and readability of the ARIMAX models were further enhanced by various model evaluation techniques such as residual analysis and model selection criteria like Akaike Information Criterion (AIC). Best predictions were found for the states Rajasthan, Andhra Pradesh and Madhya Pradesh where ARIMAX (0,2,1), ARIMAX (0,1,1) and ARIMAX (0,2,1) were fitted respectively.

2. Introduction

India is the second largest producer of Wheat (*Triticum Spp.*) in the world [1]. After Rice, Wheat is the second most produced staple food crop of the country [2]. A large section of population of India relies primarily on Wheat. For millions of people, especially in northern and central India, Wheat is the primary source of sustenance. Wheat is one of the essential crops to ensure rural livelihood, economic stability and food security of the country.

Forecasting is a process of using a time series model to predict future values with the help of historical, time-ordered data. Various work has been done regarding forecasting of wheat production in India. Devi et al [3] used Box-Jenkins ARIMA model and Artificial Neural Network to forecast production of wheat in Haryana. Bali and Singla [4] developed a deep learning based RNN model to predict wheat yield in Punjab. Although such models have high predictive accuracy, they often lack interpretability and demand substantial computational resources. In contrast, ARIMAX framework can be useful to balance interpretability, statistical rigor and forecasting accuracy. Priya and Kaushalya [5] had employed both ARIMA and ARIMAX model to assess the area, production and productivity of wheat in India. They found that ARIMAX (0,1,1) with temperature as exogeneous variable is most suitable for predicting wheat production. Banakara et al. [6] has employed ARIMAX model along with regression and time delay neural network for wheat yield forecasting with the help of weather parameters at Junagarh, Gujarat. Apart from wheat, ARIMAX has been used for predicting yield of various other crops as well. Using percentage of area under irrigation as an exogeneous variable for ARIMAX, Pandit et. al [7] coupled it with deep learning classifiers to predict yield of various Rabi crops in India. Verma [8] used fortnightly weather data for temperature and rainfall as exogeneous variables for fitting ARIMAX model to forecast yield of sugarcane in Karnal, Kurukshetra and Ambala regions of Haryana.

During autumn internship at IDEAS, ISI Kolkata, we have received training on many important topics, which are presented in **Table 2.1.** Using knowledge of those topics along with various statistical concepts, we have attempted to forecast yield of Wheat for the years 2023 - 2025 in 14 different states of India applying ARIMAX model with numerous exogenous variables such as temperature of states, amount of rainfall, amount of water available etc, using the data on these variables during 2000 – 2022.

Rest of the report is arranged as follows. Section 3 describes the objective of our work. Section 4 presents the detailed methodology followed during the project. Results are discussed in Section 5. Finally, we conclude the report in Section 6.

Table 2.1: Topics covered in training during first two weeks of internship

WEEK	DAY	TOPIC
WEEK 1	Day 1	Introduction to Data Science
	Day 2	Basic Statistics for Data Science
	Day 3	Data visualization
	Day 4	Introduction to GitHub and cloud computing
	Day 5	Streamlit-1 (Widget handling)
WEEK 2	Day 1	Streamlit-2 (Database integration)
	Day 2	Machine learning 1 (Regression)
	Day 3	Machine learning 2 (Classification)
	Day 4	LLM fundamentals
	Day 5	Communication skills

3. Project Objective

The objective behind this work is to forecast production of wheat in the presence of various exogeneous variables such as the temperature of the state (both maximum and minimum), the amount of rainfall in the state, full reservoir level (FRL), Live capacity of the FRL, observed water level and the amount of water available for the year 2023-2025 in 14 Indian states applying ARIMAX model.

4. Methodology

4.1 Data Preparation

Our dataset included the daily yield of wheat during 2000 - 2022, for various states of India *viz.* Andhra Pradesh (AP), Chhattisgarh, Gujarat, Jharkhand, Karnataka, Madhya Pradesh (MP), Maharashtra, Odisha, Rajasthan, Tamil Nadu (TN), Telangana, Uttarakhand, Uttar Pradesh (UP) and West Bengal (WB). Apart from the yield of wheat, the temperature of the state (both maximum and minimum), the amount of rainfall in the state, Full Reservoir Level (FRL), Live capacity of the FRL, observed water level and the amount of water available were also recorded for each day. Details for each state are tabulated in **Table 4.1.1**. For each state, the daily data of each variable were first converted into monthly means, which were subsequently summed across months to obtain a single annual measure. The data recorded for the states Odisha and TN were less than the minimum requirement of data for 5 years, so these states were not considered for our work.

Table 4.1.1 First and last date of recorded data and total days recorded for each state

State	First day recorded	Last day recorded	No of days recorded
AP	01/01/2000	31/12/2022	7670
Chhattisgarh[1]	01/01/2000	31/12/2022	8401
Gujarat	01/01/2000	31/12/2022	8401
Jharkhand	01/01/2002	31/12/2022	7670
Karnataka	01/01/2000	31/12/2022	8401
MP	01/01/2000	31/12/2022	8036
Maharashtra	01/01/2000	31/12/2022	8401
Odisha	01/01/2000	31/12/2003	1461
Rajasthan	01/01/2000	31/12/2022	8401
TN	01/01/2000	31/12/2000	366
Telangana[1]	01/01/2000	31/12/2022	8401
Uttarakhand[1]	01/01/2000	31/12/2022	8401
UP	01/01/2000	31/12/2022	8401
WB	01/01/2000	31/12/2022	8401

Suppose W_d be the amount of yield of wheat for any state at d^{th} day of m^{th} month in a year t during 2000 – 2022. We first estimated monthly yields by calculating their arithmetic mean. If X_m be the average wheat yield for month m having data for D days, then we have

$$X_m = \frac{1}{D}\sum_{d=1}^{D} W_d$$

Now to estimate the Annual wheat yield, say Y_t, of the year t, we summed the values of X_m for all the values of m available for that year. If the data is available for M months of the year t

$$Y_t = \sum_{m=1}^{M} X_m$$

For every state, the values of wheat yield and other exogenous variables are thus estimated for each year available during 2000 – 2022.

4.2 An Overview of the Time Series data:

Time series is a set of ordered observations of a quantitative variable taken at successive points in time. It gives a bivariate distribution between time t and the value of the phenomenon Y_t, at different points of time. Let us define T = {2000, 2001, ...,2021,2022}. Then the yield of wheat for each state becomes a timeseries $\{Y_t: t \in T\}$

4.2.1. The Components of time series data: The value of time series $\{Y_t: t \in T\}$ at any year point t is considered as the resultant of the combined impact of four major forces, called "components". These components are as follows

[1] For states formed after 01/01/2000 (*viz,* Chhattisgarh, Telangana, and Uttarakhand), pre-formation data have been retained as provided in the official dataset. These records likely correspond to the geographic regions later designated as the new states.

4.2.1.1 Trend: The general tendency of the data to increase or decrease over a long period of time is known as Trend.

4.2.1.2 Seasonal variations: The rhythmic forces operating in a regular and periodic manner year with period less than a year leads to "Seasonal variations" in a time series data.

4.2.1.3 Cyclic variations: The oscillatory movements in a time series with a period of oscillations more than a year is known as "Cyclical variations"

4.2.1.4 Random movements: Apart from the regular components of a Time series mentioned above, a time series data generally contains fluctuations which are purely random, erratic, unforeseen, and unpredictable. These fluctuations are known as random movements.

4.2.2 Decomposition of time series:

Suppose at a year t, the value of a time series is Y_t, and that of trend, seasonal variations, cyclical variations and random movements are T_t, S_t, C_t and R_t respectively. Additive model of decomposition was used for decomposing the time series as the seasonal and random components are consistent over time.

4.2.2.1 Additive model: Here, a time series is expressed as

$$Y_t = T_t + S_t + C_t + R_t \qquad (1)$$

The additive model assumes that all the components of a time series operate independently of each other.

4.3 Stationarity of Time Series data:

The mean function of the time series $\{Y_t: t \in T\}$ is defined as $\mu(t) = E(Y_t)$ and autocovariance function is defined as $\gamma\,(s, t) = E\,(Y_s, Y_t)$.

4.3.1 Stationary time series:

The timeseries $\{Y_t: t \in T\}$ is said to be stationary or weakly stationary or covariance stationary if $E|Y_t|^2 < \infty$, $\mu(t) = \mu$ and $\gamma(t + u, t) = \gamma(u,0)$ for all t and u. If $\{Y_t: t \in T\}$ is a stationary time series. The auto-covariance function (ACVF) of $\{Y_t\}$at lag u can be denoted by $\gamma(u)$ and defined as

$$\gamma\,(u) = \mathrm{Cov}(\,Y_{t+u}\,, Y_t) \qquad (2)$$

4.3.2 Auto-Correlation and Partial Auto-Correlation function:

The auto correlation function (ACF) for the time series $\{Y_t\}$ can similarly be defined as

$$\rho(u) = \frac{\gamma(u)}{\gamma(0)} = \mathrm{Corr}\,(Y_{t+u}, Y_t) \qquad (3)$$

Given a stretch of time series values

$$\ldots, Y_{t-u}, Y_{t-u+1}, \ldots\ldots, Y_{t-1}, Y_t, \ldots\ldots\ldots$$

the partial auto correlation of Y_t and Y_{t-u} is the correlation between these two random variables which is not conveyed through intervening values. If Y_t 's are normally distributed then the partial auto correlation function (PACF) between Y_t and Y_{t-u} can be expressed as

$$\varphi(u) = \mathrm{Corr}\,(Y_t\,, Y_{t-u} \mid Y_{t-u+1}\,, \ldots\ldots, Y_{t-1}\,) \qquad (4)$$

Values of ACF and PACF helps us in identifying patterns, checking stationarity and identifying the values of parameters p, d and q of an ARIMA(p,d,q) model.

4.3.3 Tests for Stationarity: We have applied Augmented Dickey-Fuller test (ADF) and Kwiatkowski–Phillips–Schmidt–Shin test (KPSS) for checking stationarity. In ADF, our null hypothesis is the series is non-stationary, whereas the null hypothesis for KPSS test is the series being stationary. We have used both the tests for each state and computed the p-value, based on which we have concluded whether the data for a particular state is stationary or not. For a state, if the p-value of either ADF test or KPSS test or both is not significant, we conclude the series is non-stationary. Furthermore, we also plotted the ACF and PACF values for each state. If the series is non-stationary, the ACF values decays slowly over lags and PACF values cuts off after very few lags.

4.3.4 Dealing with Non-stationarity: For the states having non-stationary data, we had applied differencing to make them stationary. After applying differencing of first order on the series $\{Y_t: t \in T\}$ we get a new series $\{Y_t': t \in T\}$ where $Y_t' = (Y_t - Y_{t-1})$, which is also a time series. We then again applied the tests for stationarity in the new series and plotted the ACF and PACF values. If the series still didn't become stationary, we repeated the process till it became one.

4.4 Forecasting using ARIMAX Model:

Forecasts for future time points were produced using the ARIMAX (Auto Regressive Integrated Moving Average with eXogenous variables) model, which combines the exogenous, moving average, and autoregressive components and can handle a wide range of data types including time series data with trends, seasonality and other patterns.

4.4.1 Exogeneous variables: Exogeneous variable are the variables that are determined not by the economic model but by external factors. These variables are sometimes called predetermined variables and are usually independent of error terms in the model. In our model. The variables *viz.* maximum temperature of state, minimum temperature of state, amount of rainfall in the state, FRL, live capacity of FRL, observed water level and the amount of water available acted as exogenous variables.

4.4.2 ARIMAX model: The ARIMAX model is an extension of ARIMA (Auto Regressive Integrated Moving Averages) models that incorporates one or more exogenous variables (X) that are thought to have an impact on dependent variables (Y). Mathematically an ARIMAX model can be expressed as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{k=1}^{r} \beta_k X_{t-k} + \varepsilon_t \qquad (5)$$

Where c is a constant term,

Y_t is the value of the dependent variable at time t,

$\phi_1, \phi_2, \ldots, \phi_p$ are autoregressive coefficients,

$\theta_1, \theta_2, \ldots \theta_q$ are moving average coefficients,

$\beta_1, \beta_2, \ldots, \beta_r$ are coefficients of the exogeneous variables,

$X_{t-1}, X_{t-2}, \dots, X_{t-r}$ are the exogeneous variables, and

ε_t is the error term at time t

If differencing is applied, the model becomes ARIMAX (p,d,q), where d is the order of differencing. Various ARIMAX models, depending on the values of p, d and q had been fitted for each state and the one with lowest Akaike information criterion (AIC) value was chosen as the best model.

.**4.4.3 Goodness of fit:** After fitting, it is important to perform model diagnostics such as residual analysis and goodness of fit tests to make sure the model accurately reflects the underlying patterns. The performance of the ARIMAX models for every state was evaluated using error metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE).

4.4.3.1 MAE: MAE calculates the mean of the absolute differences between the observed values and predicted values. It is defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

Where n is the number of points, y_i is the actual value of ith data point and $\hat{y}_i$ is the predicted value of the ith data point.

4.4.3.2 MSE: MSE, defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (7)$$

calculates the average of the squared value of the deviations between the observed values and the predicted values.

4.4.3.3 MAPE: The percentage absolute difference between the expected and actual values is called MAPE and is defined as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100 \quad (8)$$

5. Data Analysis and Results

5.1 Annual Yield of wheat for each state:

Our dataset has data of wheat yield and other exogeneous variables for 12 Indian states *viz.* AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, MP, Maharashtra, Rajasthan, Telangana, Uttarakhand, UP and WB during 2000 – 2022. The year-wise yield of wheat for each state is plotted in **Fig 5.1.1.a** and **Fig 5.1.1.b.**

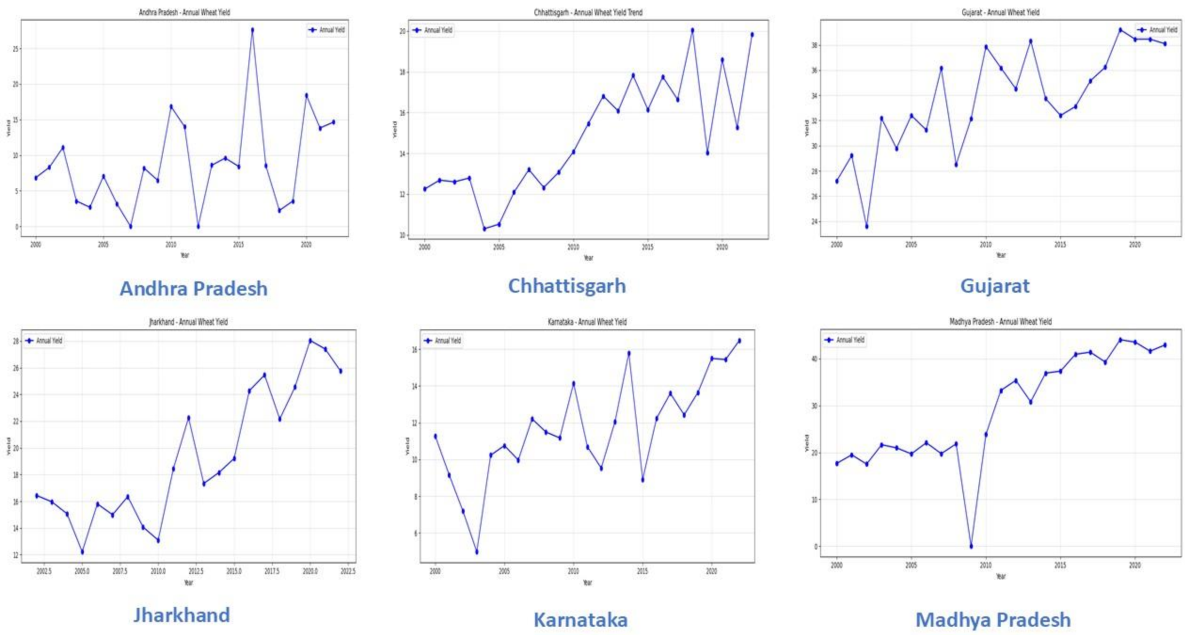

Fig 5.1.1.a Yield of wheat v/s years for the states AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, MP

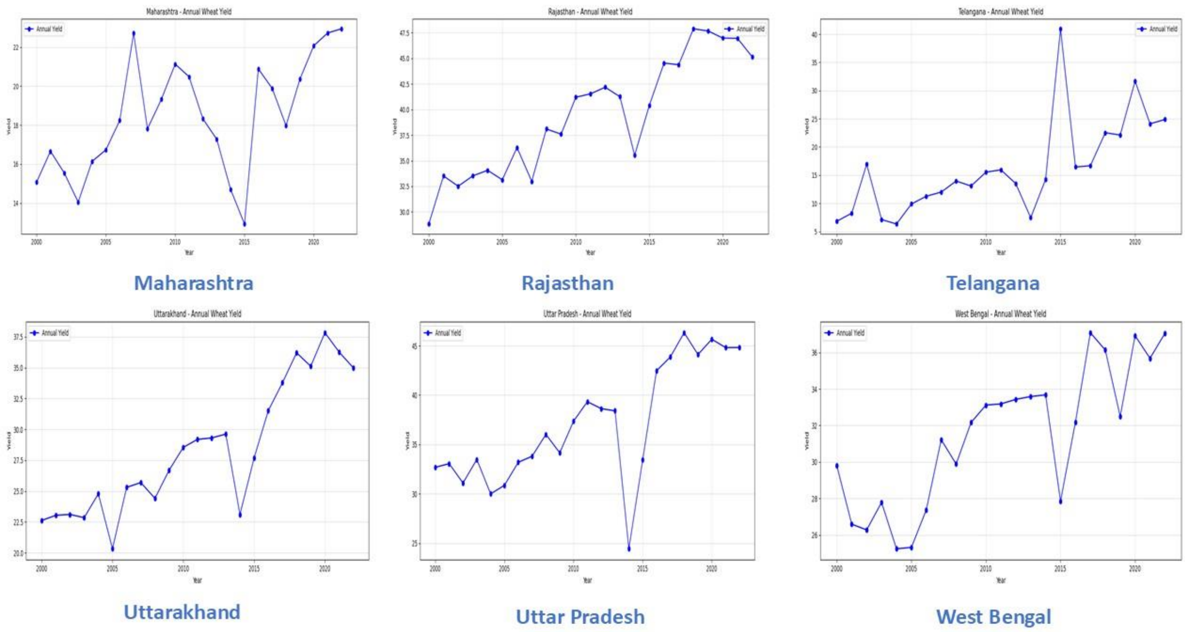

Fig 5.1.1.b Yield of wheat v/s years for the states Maharashtra, Rajasthan, Telangana, Uttarakhand, UP and WB

5.2 Decompositions of the series: After plotting the, it was time to decompose the series of wheat yield into its components *viz.* Trend, seasonality and random components. Additive model of decomposition was chosen because the components seasonality and random movements are consistent over time. **Fig 5.2.1.a** and **Fig 5.2.1.b** presents the components for different states.

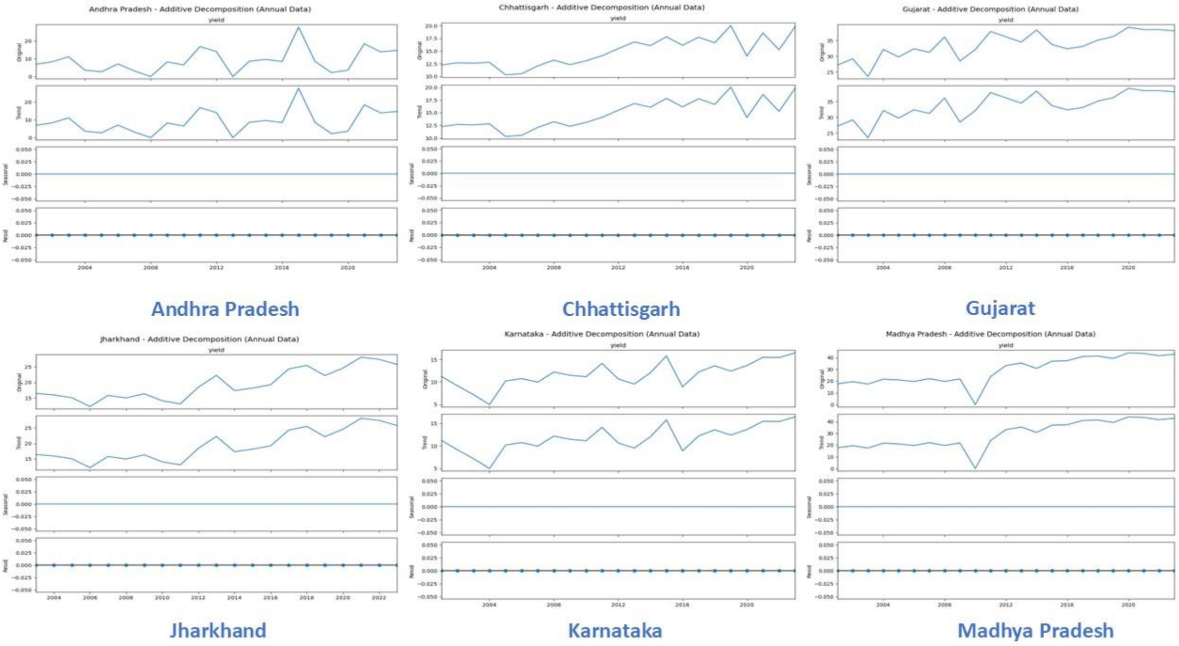

Fig 5.2.1.a Additive decomposition of yield for the states AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, and MP: (From top to bottom) Original, Trend, Seasonality, Random components

Fig 5.2.1.b Additive decomposition of yield for the states Maharashtra, Rajasthan, Telangana, Uttarakhand, UP, and WB: (From top to bottom) Original, Trend, Seasonality, Random components

5.3 Stationarity of the series: To check whether the time series data for wheat yield is stationary or not we have used two tests viz. ADF and KSS. The null hypothesis for ADF test was H_0: series is non - stationary whereas that for KPSS test was H_0: the series is stationary. Both the tests were performed for all 12 states given and values of the test statistics and corresponding p-values are tabulated in **Table 5.3.1**.

Table 5.3.1. Measures of the tests for stationarity of yield data for all states

State names	ADF test statistic	p-value	KPSS test statistic	p-value	Remark
AP	-1.0164	0.7472	0.3495	0.0989	Non-stationary
Chhattisgarh	-0.5404	0.8839	0.6057	0.0221	Non-stationary
Gujarat	-0.9882	0.7575	0.5906	0.0235	Non-stationary
Jharkhand	-0.2882	0.928	0.6997	0.0136	Non-stationary
Karnataka	-1.0245	0.7442	0.6618	0.0170	Non-stationary
MP	-6.2516	0.0000	0.5914	0.0234	Non-stationary
Maharashtra	-4.5368	0.0002	0.3889	0.0820	Stationary
Rajasthan	-1.2562	0.6490	0.6449	0.0186	Non-stationary
Telangana	2.728	0.9991	0.7270	0.0111	Non-stationary
Uttarakhand	2.3761	0.9990	0.6105	0.0217	Non-stationary
UP	0.7183	0.9902	0.5336	0.0341	Non-stationary
WB	-1.3884	0.5878	0.5858	0.0239	Non-stationary

Furthermore, the ACF and PACF values of all the states were plotted as well. Those states where the series is non-stationary, a gradual decline of ACF values is observed and the PACF values drop sharply after lag 1. For those with stationary series, both ACF and PACF cuts off quickly. The plots having ACF and PACF values for each state is given in **Fig 5.3.1.a** and **Fig 5.3.1.b.**

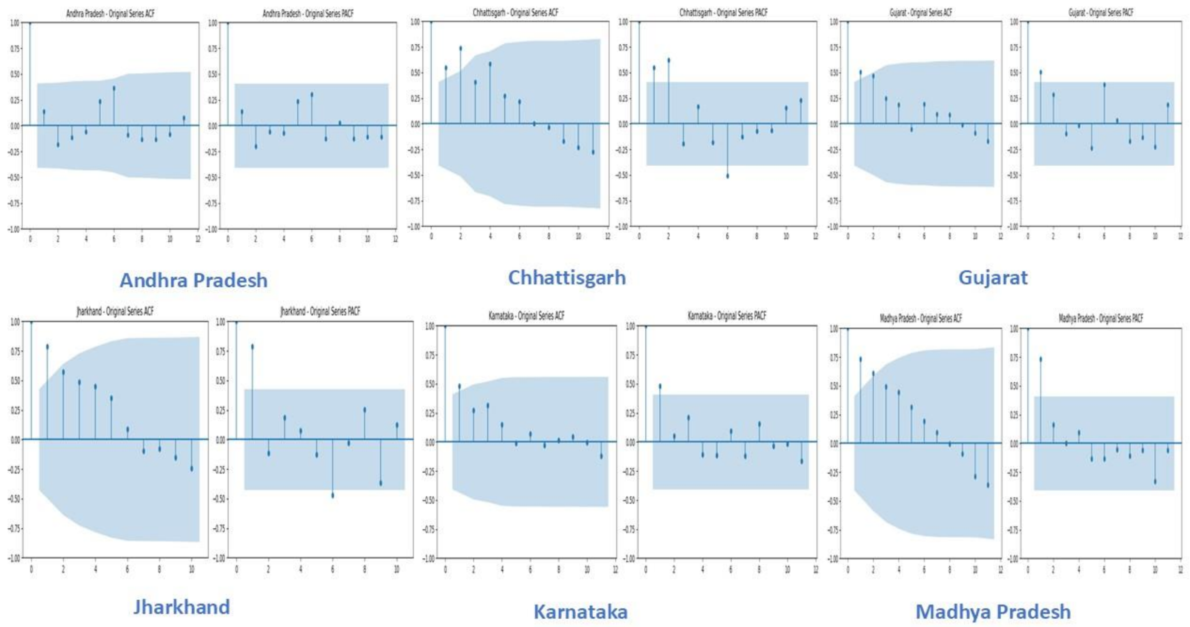

Fig 5.3.1.a ACF (left) and PACF (right) plots of the yield for the states AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, and MP

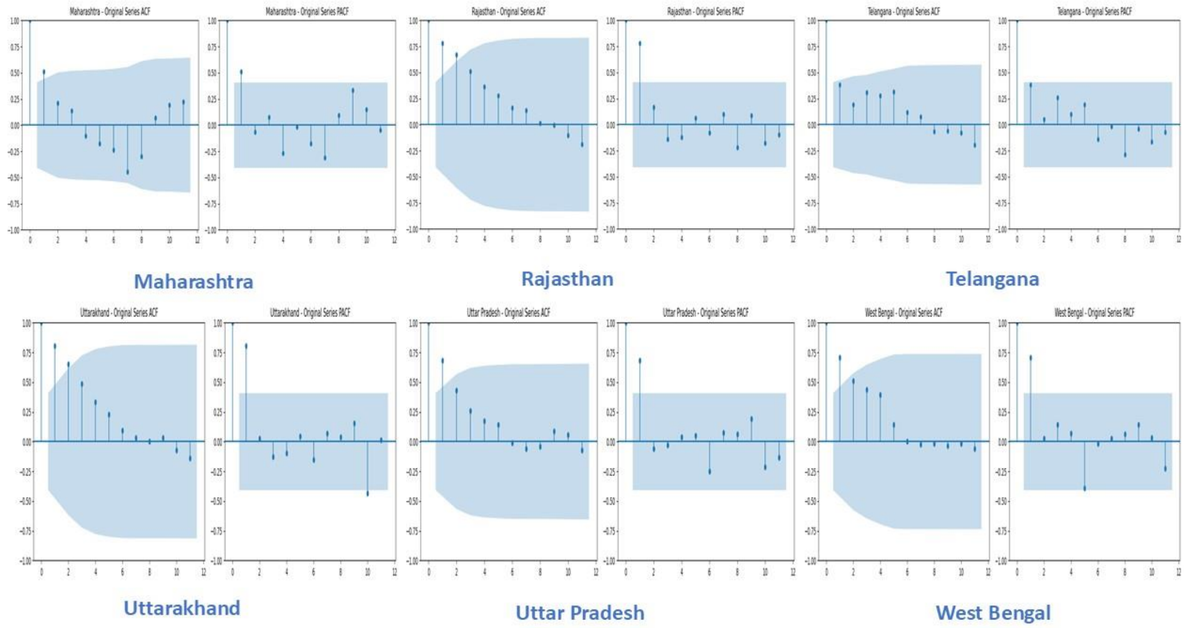

Fig 5.3.1.b ACF (left) and PACF (right) of yield for the states Maharashtra, Rajasthan, Telangana, Uttarakhand, UP, and WB

5.4 Differencing for non-stationary series: From the values and ACF/PACF plots it is clear that apart from Maharashtra, yield of wheat is non-stationary in each state available. So, to make them stationary, first order differencing is applied. After differencing, the value of wheat yield in each state is again plotted, which are presented in **Fig 5.4.1.a** and **Fig 5.4.1.b**.

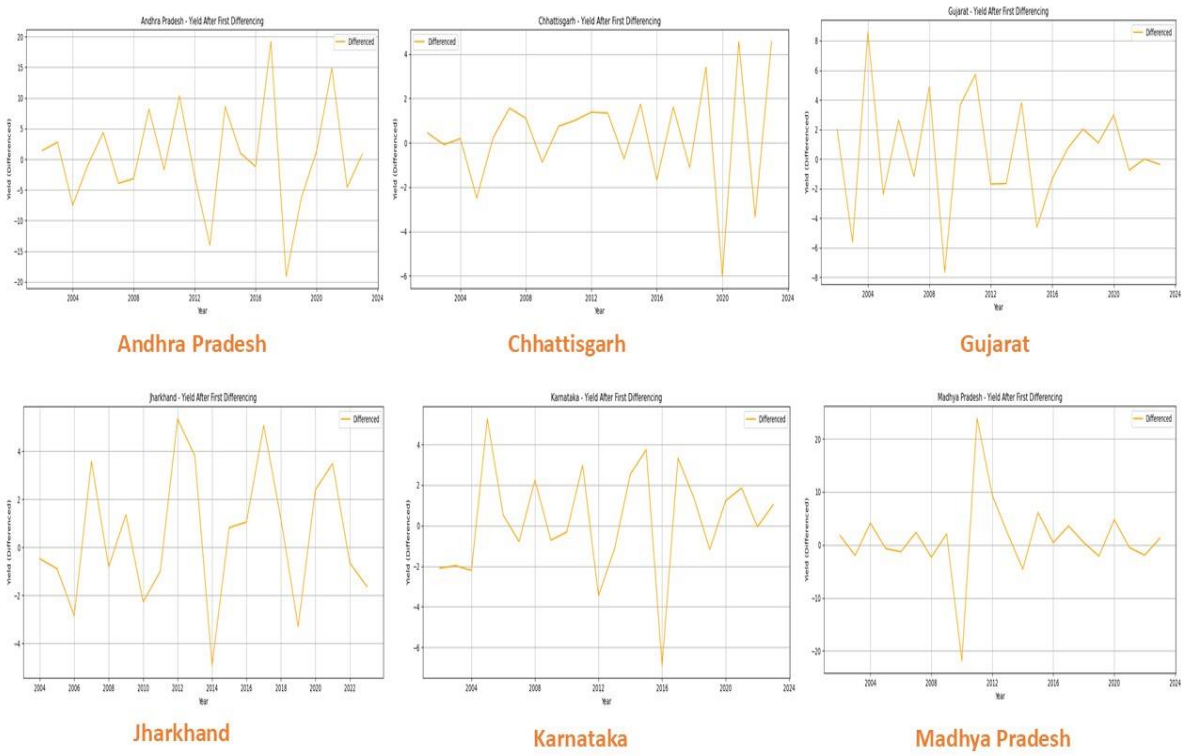

Fig 5.4.1.a yield v/s time plot after applying first order differencing for the states AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, and MP

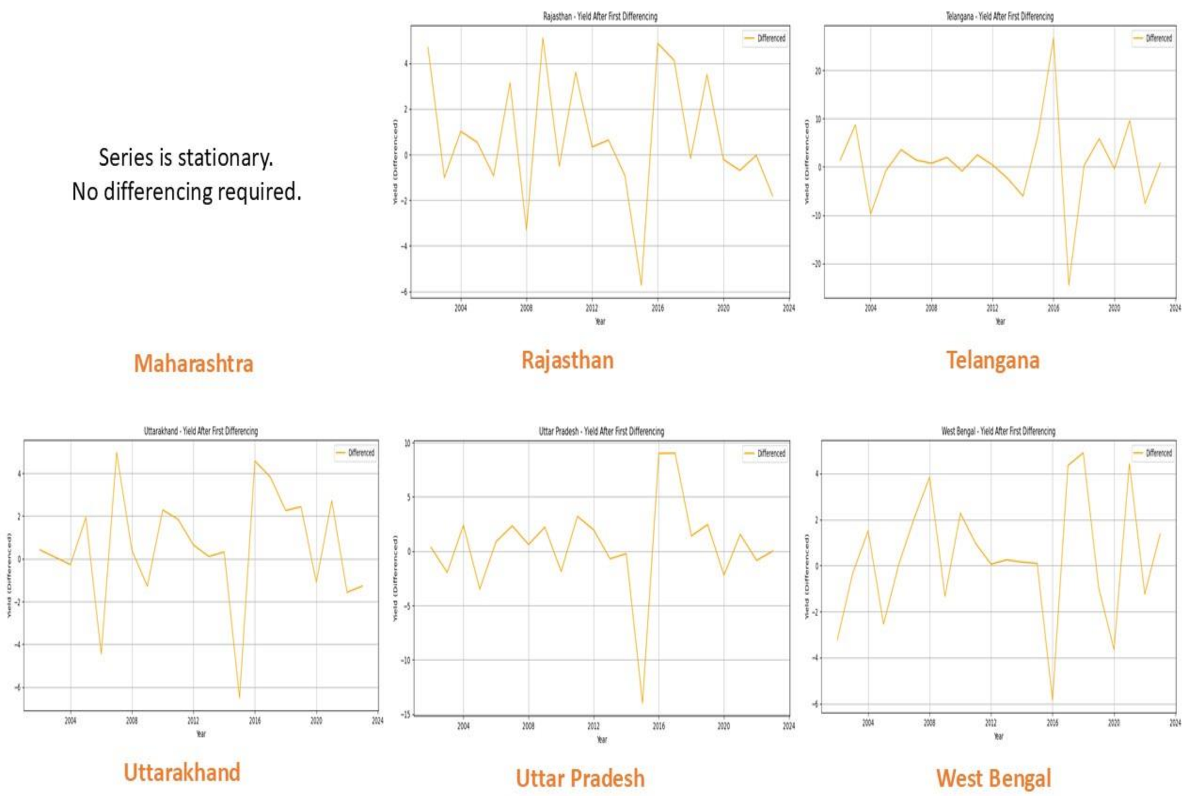

Fig 5.4.1.b yield v/s time plot after applying first order differencing for the states Maharashtra, Rajasthan, Telangana, Uttarakhand, UP, and WB

After plotting, the series is further decomposed into trend, seasonality and residuals in order to check their presence. The decomposed plots are presented in **Fig 5.4.2.a** and **Fig 5.4.2.b**

Fig 5.4.2.a Additive model decomposition after applying first order differencing for the states AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, and MP: (From top to bottom) Original, Trend, Seasonality and Random Components

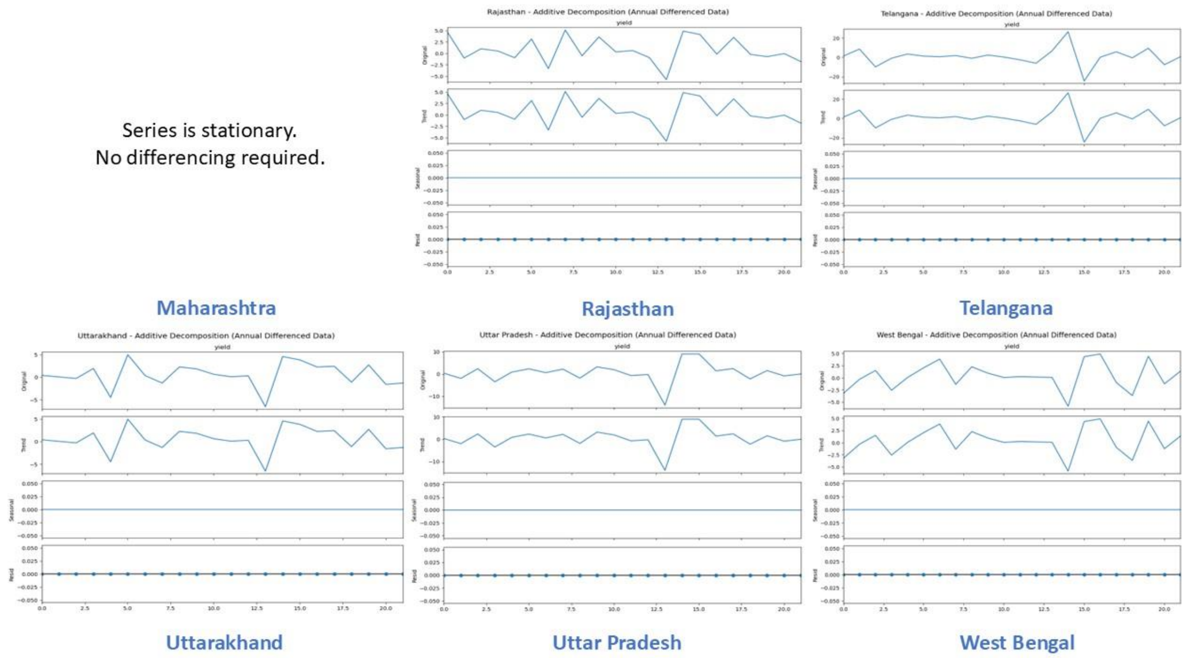

Fig 5.4.2.b Additive model decomposition after applying first differencing for the states Maharashtra, Rajasthan, Telangana, Uttarakhand, UP, and WB: (From top to bottom) Original, Trend, seasonality and random Components

Next, we applied ADF and KSS tests again on the differenced data. The values of test statistics and corresponding p – values are tabulated in **Table 5.4.1**.

Table 5.4.1. Measures of the tests for stationarity of yield data for all states after first differencing

State names	ADF test statistic	p-value	KPSS test statistic	p-value	Remark
AP	-4.2054	0.0006	0.1417	0.1000	Stationary
Chhattisgarh	-9.0771	0.0000	0.1981	0.1000	Stationary
Gujarat	-2.0835	0.2512	0.4107	0.0725	Non – Stationary
Jharkhand	-12.407	0.0000	0.0810	0.1000	Stationary
Karnataka	-6.1714	0.0000	0.1653	0.1000	Stationary
MP	-5.2095	0.0000	0.1238	0.1000	Stationary
Rajasthan	-1.9142	0.3254	0.1692	0.1000	Non – Stationary
Telangana	-4.9477	0.0000	0.1463	0.1000	Stationary
Uttarakhand	-0.3538	0.9175	0.0770	0.1000	Non – Stationary
UP	-2.9061	0.0446	0.0532	0.1000	Stationary
WB	-8.8549	0.0000	0.1228	0.1000	Stationary

Apart from Gujarat, Rajasthan and Uttarakhand, differencing has resulted in Stationary series for yield in all other states, which is further established by ACF and PACF plots in **Fig 5.4.3.a** and **Fig 5.4.3.b.** The series of yield in Gujarat, Rajasthan and Uttarakhand is once again differenced and checked for stationarity. The values of the tests are tabulated in **Table 5.4.2.**

Table 5.4.2. Measures of the tests for stationarity of yield data after second differencing

State names	ADF test statistic	p-value	KPSS test statistic	p-value	Remark
Gujarat	-3.5486	0.0068	0.1726	0.1000	Stationary
Rajasthan	-5.4815	0.0000	0.2578	0.1000	Stationary
Uttarakhand	-3.9249	0.0019	0.2888	0.1000	Stationary

Since all the states are stationary, we can now move to forecasting.

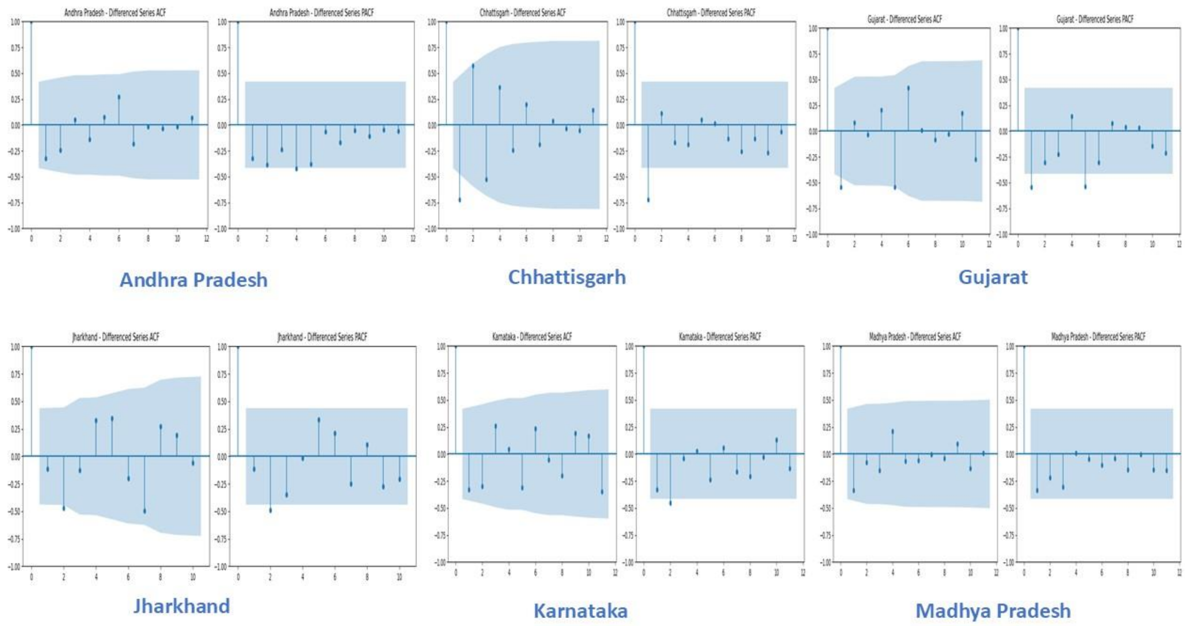

Fig 5.4.2.a ACF plot (left) and PACF plot (right) after applying first order differencing for the states AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, and MP

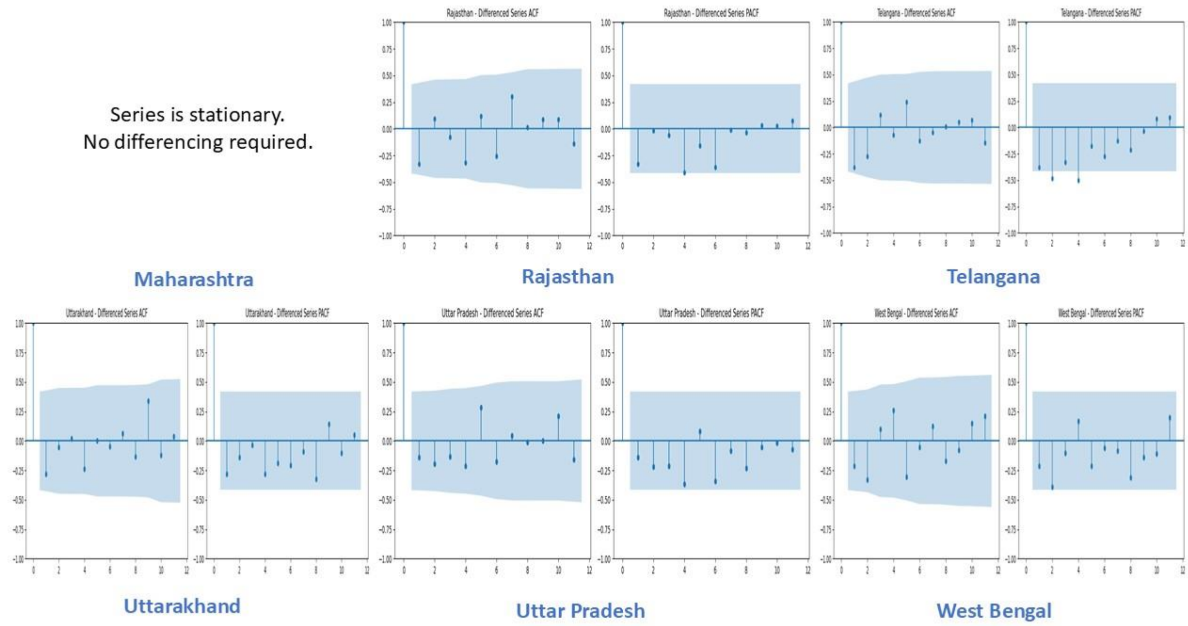

Fig 5.4.3.b ACF plot(left) and PACF plot(right) after first order differencing for the states Maharashtra, Rajasthan, Telangana, Uttarakhand, UP, and WB

5.5 ARIMAX forecasting and residual analysis: For each state, various ARIMAX (p, d, q) models are fitted for different values of p, d and q. The best ARIMAX model is then chosen based on the value of AIC. Further, the performance of the best model is evaluated using error metrics like MAE, MAPE and MSE. The results for each state are tabulated.

Table 5.5.1. Details of ARIMAX models and error metrics for each state

State	Best model	AIC	MAE	MSE	MAPE
AP	ARIMAX (0,1,1)	139.09	0.221	0.049	1.51%
Chhattisgarh	ARIMAX (0,1,1)	133.27	19.833	393.357	100%
Gujarat	ARIMAX (1,2,1)	93.68	8.190	67.071	21.50%
Jharkhand	ARIMAX (1,2,1)	83.21	1.234	1.523	4.79%
Karnataka	ARIMAX (1,1,1)	88.30	1.374	1.888	8.34%
MP	ARIMAX (0,2,1)	113.41	0.680	0.482	1.58%
Maharashtra	ARIMAX (0,1,1)	96.04	0.368	0.136	1.60%
Rajasthan	ARIMAX (0,2,1)	96.68	0.140	0.019	0.31%
Telangana	ARIMAX (0,1,1)	141.32	5.926	35.116	23.82%
Uttarakhand	ARIMAX (0,1,1)	86.12	1.551	2.407	4.43%
UP	ARIMAX (0,2,1)	120.98	3.710	14.212	8.41%
WB	ARIMAX (0,2,1)	98.57	2.247	5.049	6.06%

The forecasted values are then plotted against time year along with the observed values. The forecast plots are presented in **Fig 5.5.1.a** and **Fig 5.5.1.b**.

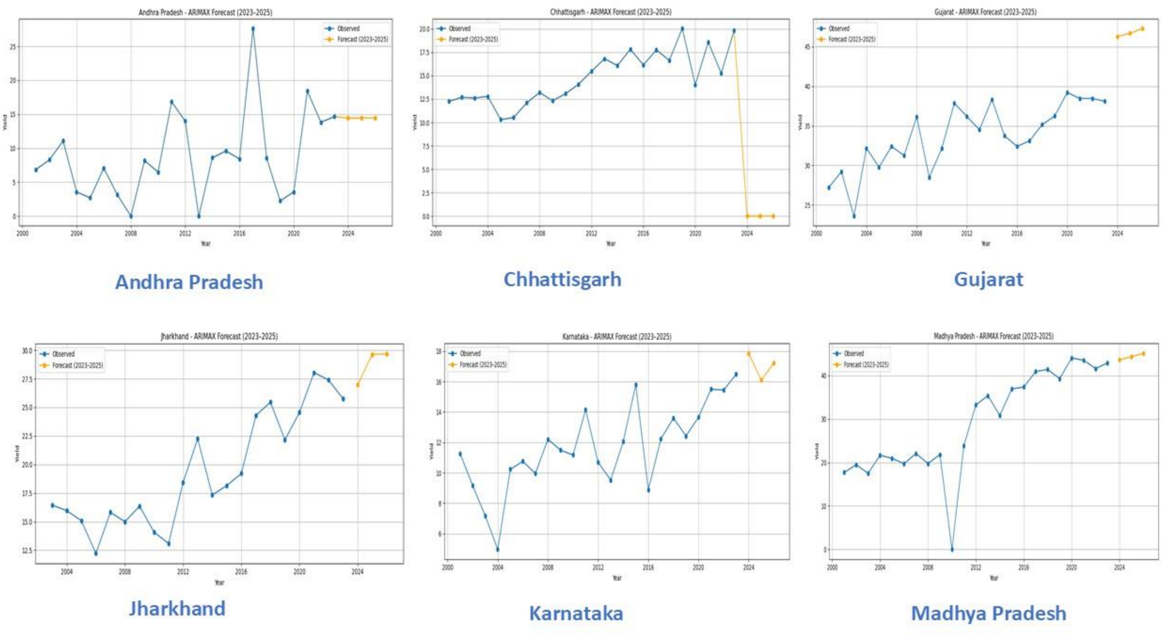

Fig 5.5.1.a Forecast of Wheat yield for the year 2023 – 2025 for the states AP, Chhattisgarh, Gujarat, Jharkhand, Karnataka, and MP

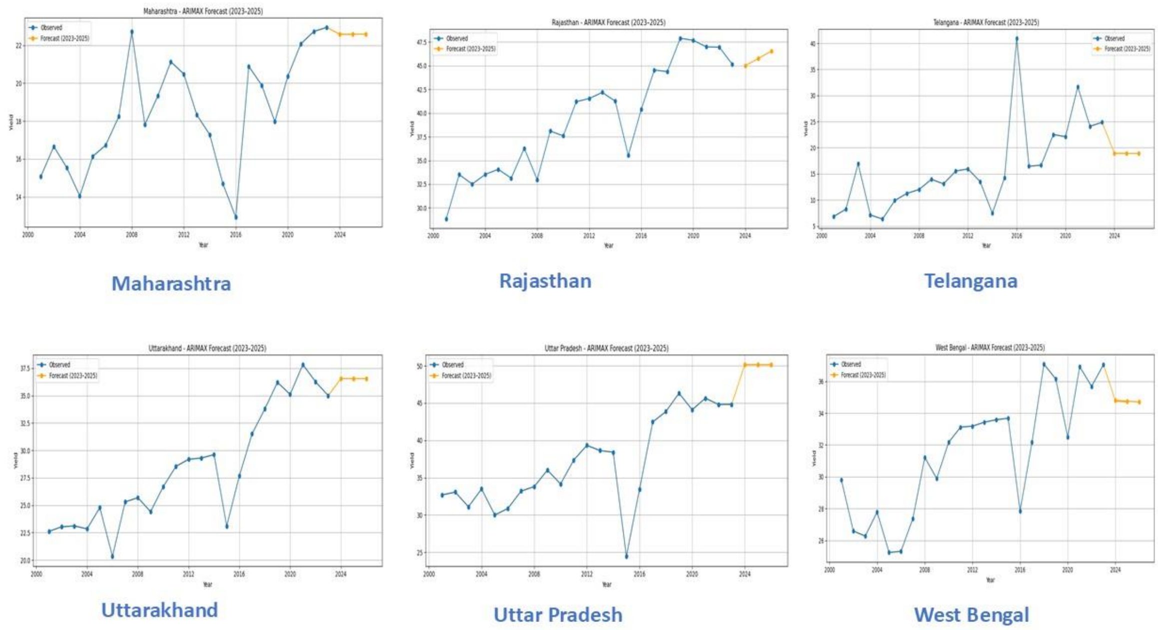

Fig 5.5.1.b Forecast of Wheat yield for the year 2023-2025 for the states Maharashtra, Rajasthan, Telangana, Uttarakhand, UP, and WB

6. Conclusion

Based on error metrics such as MAE, MAPE and MSE, the ARIMAX (0,2,1) model of the state Rajasthan appears to be the best fit followed by ARIMAX (0,1,1) of AP and ARIMAX (0,2,1) of MP. On the other hand, performance of ARIMAX (0,1,1) of Chhattisgarh was the worst followed by ARIMAX (0,1,1) of Telangana and ARIMAX (1,2,1) of Gujarat. However, there exists few areas where improvement can still be made. For instance,

1. The project is done based of annual data computed from daily data. We can convert the daily data into monthly data or even use the daily data only for forecasting.
2. The exogenous variables can be used individually or even as a group of same kind of variables (e.g. a group including temperature and rainfall, another one including FRL, live cap FRL etc.) to forecast yield.
3. If proper data is available wheat forecasting can be performed for other states of India as well or even for other crops.

7. References

1. FAS. (2025). *Production – Wheat:* Foreign Agricultural Service. U.S. Department of Agriculture. Retrieved October 16, 2025, from https://www.fas.usda.gov/data/production/commodity/0410000

2. FAOSTAT. (2025). *Crops and livestock products*. Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#data/QCL

3. Devi, M., Kumar, J., Malik, D. P., & Mishra, P. (2021). Forecasting of wheat production in Haryana using hybrid time series model. Journal of Agriculture and Food Research, 5, 100175.

4. Bali, N., & Singla, A. (2021). Deep learning-based wheat crop yield prediction model in Punjab region of North India. *Applied Artificial Intelligence*, *35*(15), 1304-1328.

5. Priya, S. K., & Kausalya, N. Forecasting Wheat Crop Area, Production & Productivity in India Using Arimax Model.

6. Banakara, K. B., Sharma, N., Sahoo, S., Dubey, S. K., & Chowdary, V. M. (2023). Evaluation of weather parameter-based pre-harvest yield forecast models for wheat crop: a case study in Saurashtra region of Gujarat. *Environmental Monitoring and Assessment*, *195*(1), 51.

7. Pandit, P., Sagar, A., Ghose, B., Dey, P., Paul, M., Alqadhi, S., ... & Abdo, H. G. (2023). Hybrid time series models with exogenous variable for improved yield forecasting of major Rabi crops in India. *Scientific Reports*, *13*(1), 22240.

8. Verma, U. (2022). ARIMA and ARIMAX models for sugarcane yield forecasting in Northern Agro-climatic zone of Haryana. Journal of Agrometeorology, 24(2), 200-202.

9. Das, Pankaj, and Samir Barman. "Perspective Chapter: An Overview of Time Series Decomposition and Its Applications." Applied and Theoretical Econometrics and Financial Crises (2025): 59.