The Pharma Innovation Journal 2023; SP-12(12): 1145-1157

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; SP-12(12): 1145-1157

www.thepharmajournal.com
Received: 05-10-2023
Accepted: 11-11-2023

Author's details are given below the reference section

From historical data to future predictions: Analyzing and forecasting oilseed yield trends in India using time series methods

N Revathi, Dixita Gourshetty, V Haritha, Noorbina Razak, Arsha Sugathan, Sankar M, Ashitha Thomas, Aveen Ragh, Humbare Mrunalini Dinkar and Vaisakh Venu

DOI: https://doi.org/10.22271/tpi.2023.v12.i12So.24704

Abstract
The study investigates the use of Autoregressive Integrated Moving Average (ARIMA) modeling techniques to predict oilseed yields using historical agricultural data. The dataset includes records of oilseed yields from multiple growing seasons in India. The study uses data preprocessing, cleaning, exploratory analysis, and rigorous stationarity checks. The ARIMA model's parameters are identified through Autocorrelation Function and Partial Autocorrelation Function plots. Performance is evaluated using Akaike Information Criterion corrected, Bayesian Information Criterion, Root Mean Squared Error, and residual analysis. The findings show the ARIMA model's effectiveness in capturing temporal patterns and seasonality in oilseed yield data, proving its ability to provide accurate forecasts.

Keywords: Oil seeds, time series analysis, ARIMA, ACF, PACF, AICc, RMSE

Introduction
India ranks fourth globally in terms of oilseed production. In India, oilseeds are cultivated on a significant amount of land and yield a high volume of output. They are the second most important determinant of agricultural economy, next to cereals within the segment of field crops. They contribute about 13% of the gross cropped area and 3% of the gross national product. Oilseeds provide edible oils, which are essential for human nutrition and health which also contributes energy, essential fatty acids, fat-soluble vitamins, and antioxidants. The country produces groundnut, soybean, sunflower, castor, linseed, sesamum, niger seed, rapeseed-mustard, and safflower oilseeds. Oilseeds are commercially important as they provide oil and oilcakes which are used to make lubricants, varnish, medicine, perfume, candles, soap, manure, and cattle feed. Major oilseed producing states in India include Gujarat, Rajasthan, Madhya Pradesh, Maharashtra, and Andhra Pradesh.
India holds first position in vegetable oils consumption with a per capita consumption of 18 kg oil per annum. However, the domestic production of oilseeds is not sufficient to meet the demand, and India imports vegetable oil to fill the gap this affects the domestic oilseed industry. Hence, increasing oilseeds production is important for the country to achieve food security, reduce foreign dependence and improve farmers economic and social status.

Castor seed
India is the largest producer of castor seeds (*Ricinus communis)*, accounting 36% of total output and 28% of the world's acreage. Gujarat is the largest castor seed producing state in India, followed by Rajasthan and Andhra Pradesh. The most widely used and significant seed oil globally in terms of economy is castor oil.
Approximately 40–55% of castor seed is oil, and the kernel has the greatest oil content of any farmed oil crop at 64–71%. The high concentration of ricin oleic acid (84.2–94%), an 18-carbon monounsaturated fatty acid, makes castor oil a special kind of vegetable oil [1]. Castor seeds contain 45-47% non-edible oil, which is used for industrial purposes in the manufacture of soaps, transparent paper, printing inks, varnishes, linoleum, and plasticizers and as a lubricant in all moving parts particularly high-speed engines and airplanes.

Corresponding Author:
Vaisakh Venu
Assistant Professor, Department of Basic Engineering and Applied Sciences, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Groundnut

Groundnut, also known as peanut (*Arachis hypogaea)*, is a leguminous crop that is widely cultivated for its edible seeds. The plant is native to South America but is now grown in various parts of the world, particularly in tropical and subtropical regions. Groundnut is one of the major oilseeds cultivated in India, and its cultivation is widespread across various states. Gujarat, Andhra Pradesh, Telangana, Karnataka, Tamil Nadu, and Maharashtra are among the leading producers.

One of the primary economic values lies in its role as a vital food source. The seeds, commonly known as peanuts, are consumed worldwide in various forms, including roasted, boiled, and as a key ingredient in peanut butter. Groundnut oil, extraction of the seeds, is widely utilized for cooking, making it a staple in many households. Beyond its role as a food source, groundnut is a significant contributor to the agricultural economy. There are numerous health benefits for humans by consuming Groundnuts [2, 3]. The cultivation of groundnuts provides a source of income for numerous farmers, particularly in regions with favourable growing conditions. As a cash crop, it plays a crucial role in supporting livelihoods and fostering economic development, especially in developing countries where agriculture is a primary economic activity.

Linseed: Linseed (*Linum usitatissimum* L.) is an one of the traditional crops that comes under the category of oilseeds. It has numerous applications in fact each part of the plant has a different usage. It is an annual plant belonging to the genus Linum and the family Linaceae. In fact, the name Linum originated from the Celtic word lin or "thread", and the name usitatissimum is Latin for "most useful". The oil from the linseed has much potential in industrial applications. Industries like food processing, eco materials and cosmetics are using this oil for making various products.

India is the second largest (21.21%) linseed- growing country in the world in terms of area of cultivation after Canada. Production-wise, India ranks 4th (8.20%) in the world after Canada (40.51%), China (18.68%) and the USA (10.89%) [4]. In India, Madhya Pradesh is the state that comes leading in area and production.

Niger seed: *Guizotia abyssinica* (L.f.) Cass. (*Asteraceae*), commonly known as Niger seed, is an important edible seed oil commonly grown in the Ethiopian highlands as well as in India, where it is an integral part of Indian and Ethiopian economy [5, 6]. India produced approximately 33 thousand metric tonnes of Niger seeds in 2022. These small black or dark grey coloured oilseeds are rich in oil (30-40%), protein (20-25%), soluble sugar (12-18%) vitamins and minerals like magnesium, calcium, and potassium. Niger oil offers various health benefits like promoting heart health, aiding sleep, and relieving gastrointestinal problems and. Niger seed are sown mostly in the summer season and harvested around 4 to 5 months after planting. Under ideal growing condition it exhibits a yield potential around 800-1000 kg/ha. Andhra Pradesh, Karnataka, Bihar, Madhya Pradesh, West Bengal and Orissa are the major Niger growing states of India.

Rapeseed-Mustard: Rapeseed and mustard are the third most important edible oilseeds globally, followed by soybeans and oil palms. They have oil content ranging from 37 to 49% and are used in pickles, curries, vegetables, hair oils, medicines, and fat production. The oil cake is also used as animal feed and fertiliser. Young plants' leaves are used as green vegetables, while their stems and leaves provide fodder for cattle. Mustard oil is used in the tanning industry to soften leather.

In 2018-19, the world's rapeseed mustard production reached 36.59 million hectares, 72.37 million metric tonnes, and 1980 kg/ha. India, accounting for 19.8% and 9.8% of total acreage and production, saw a significant increase in productivity over the past eight years. India's mustard production accounted for 75-80% of the 6.23 million hectares grown in the country during the 2018-19 growing season. By the end of 2023, India is estimated to produce over 12 million tons of rapeseed and mustard, an increase from the previous fiscal year. In fiscal 2022, India produced over 38 million tons of oilseeds.

Safflower: Safflower (*Carthamus tinctorius)* is one of the major oilseed crops in India which has the highest average contribution to total production of oilseeds in India (38%), followed by rapeseed-mustard (27%) and groundnut (27%). The main purpose of growing safflower crop in many countries around the world is from producing high quality oil reliable for human consumption. The percentage of the oil in the safflower seeds is ranging between 14.8% - 25.9% [7]. Safflower oilseed is a rich source of fat and vitamin E. One tablespoon (14 grams) of safflower oil provides 124 calories and 14 grams of fat, with no protein or carbs. Safflower oil also contains 32% of the daily value (DV) of vitamin E, which is an antioxidant that helps protect cells from damage. The glycemic load of safflower oil is zero, meaning it does not affect blood sugar levels. The major safflower growing states are Maharashtra and Karnataka contributing more than 90% of India’s production.

Sesamum: Sesame (*Sesamum indicum* L.) is one of the oldest oilseed crops cultivated in India. Sesame is a drought resistant crop due to its extensive root system. It is known for its nutritional value and demand for foreign exchange has shown a substantial increase in the coming years. India ranks first in sesame production in the world with 19.47 lakh ha area and 8.66 lakh tones production. In India, 85% of the sesamum production comes from West Bengal, Madhya Pradesh, Rajasthan, Uttar Pradesh, Gujarat, Andhra Pradesh, and Telangana [8]. The country's weather condition provides a suitable ground condition for the cultivation of sesame, influencing the variations in its area, production, and productivity across different regions of the country. Sesamum acts as both as an oilseed and a rotational crop further marking its economic and ecological significance.

Soybean: Soybean (*Glycine max*) is number one oilseed crop in India, with an estimated production of 13.5 million tons in 2020-21. India produces around 10.8 million hectares of soybean every year, which contributes to 25% of the global edible oil, about two-thirds of the world’s protein concentrate for livestock feeding [9]**.** It is one of the most common cooking oils available globally, also rich in vitamin K, which is necessary for the protein synthesis that are crucial for maintaining bone mass. Soybean oil also strengthens the immune system and has antioxidant activity. It improves skin health and reduces signs of premature aging. Soybean oil promotes hair growth and healthy bones. It provides important omega-3 fatty acids. The crop is being cultivated in around 11.8 million hectares, with around 80% of the soybean produced in Maharashtra and Madhya Pradesh.

Sunflower: Sunflower (*Helianthus annuus*) is commonly grown as a crop for its edible oily seeds. Sunflower is one of the most important oilseed crops on a global level, grown on a total of over 22 million hectares worldwide [10]. Besides cooking oil production, it is also used in some industrial applications, as livestock forage, and as an ornamental in domestic gardens. In India, Karnataka is the leading state in the producer of sunflower. The major sunflower producing states of India are Karnataka having production of 3.04 lakh tons from an area of 7.94 lakh hectares followed by Andhra Pradesh, Maharashtra, Bihar, Orissa, and Tamil Nadu. The average productivity at all India level was 900 kg/ha. The production of sunflower in the world in 2022-23 was 52.42 million tons. For agricultural planning and decision making, forecasting oilseed production is found to be effective. Climate change, biotechnology, consumer preferences, and trade policies are some of the factors that affect oilseed production. Using advanced tools and methods such as crop modeling, remote sensing, bioinformatics and genomics, oilseed production can be predicted [11]. The purpose of this study is to model and compare the yield of castor seed, groundnut, linseed, Niger seed, rapeseed, safflower, sesamum, soybean and sunflower in India using the machine learning approach ARIMA (Auto Regressive Integrated Moving Average).

Materials and Methods: The analysis made use of a dataset that included historical information of oilseed yields gathered from 1966 to 2023. These data were taken from upag.gov.in and indicate Indian oilseed yields in tons per acre. R software is used for modeling and related analysis [12].
To deal with missing values and outliers, the first data preparation included stringent cleaning methods. To describe the dataset's temporal patterns, trends, and probable seasonality, the cleaned dataset underwent exploratory data analysis (EDA), which included descriptive statistics and visualizations. The Augmented Dickey-Fuller (ADF) test was used to assess stationarity in the time series, assessing the null hypothesis of the absence of a unit root. Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were also examined to identify probable autoregressive (AR) and moving average (MA) components in the oilseed yield time series.
For model building, parameters (p), (d), and (q) for the Autoregressive Integrated Moving Average (ARIMA) model were identified. These parameters were derived from careful examination and interpretation of the ACF and PACF plots, crucial in capturing the inherent autocorrelation and seasonality within the time series data. The ARIMA [13, 14] model for the study was formulated as:

$$Y(t) = c + \varphi_1 Y(t-1) + \varphi_2 Y(t-2) + \ldots + \varphi_p Y(t-_p) + \theta_1 \varepsilon(t-1) + \theta_2 \varepsilon(t-2) + \ldots + \theta q \varepsilon(t-q) + \varepsilon(t)$$

Here, "Y(t)" represents the oilseed yield at time "t," "c" is a constant term, φ_i (phi) denotes the autoregressive (AR) coefficients, θ_i (theta) represents the moving average (MA) coefficients, "p" is the order of the autoregressive part, "q" is the order of the moving average part, "d" is the degree of differencing, and ε(t) signifies the white noise at time "t."
For model selection, information metrics such as the Akaike Information Criterion corrected (AICc) and Bayesian Information Criterion (BIC) were used. The AICc and BIC help in the selection of the sparsest model that sufficiently captures the data while penalizing for model complexity. AICc and BIC values that are lower imply a better model fit. The Root Mean Squared Error (RMSE) was also generated to assess the accuracy of the ARIMA model's predictions in comparison to observed oilseed yields. RMSE calculates the average difference between anticipated and actual values to estimate the model's prediction error.
A comprehensive residual analysis was used to assess the ARIMA model to confirm its dependability and assumptions. Visual inspections, Q-Q plots, histograms, and the Autocorrelation Function (ACF) were used in the study to confirm that the distribution was normal and to look for any temporal trends or systematic mistakes. The residuals' lack of any obvious trends or correlations supported the model's prediction of oilseed yield accuracy going forward. The model's prediction reliability was increased because of this analysis's confirmation that the independence, normality, and constant variance assumptions were met.

Results and Discussions: Arima models were constructed for each oil seed separately and analysis were done for best fitted models.

Castor Seed: The study makes use of castor seed yield data spanning from 1967 to 2023. The augmented Dicky Fuller test (ADF test) is used to determine if the data is stationary. When the data was not stationary at first (p-value: 0.2514), the differencing approach was used to make it so. Using the ADF test, it was discovered that the first differenced data was stationary (p-value: 0.01). The ACF and PACF graphs of the differenced data are shown in Fig. 1.

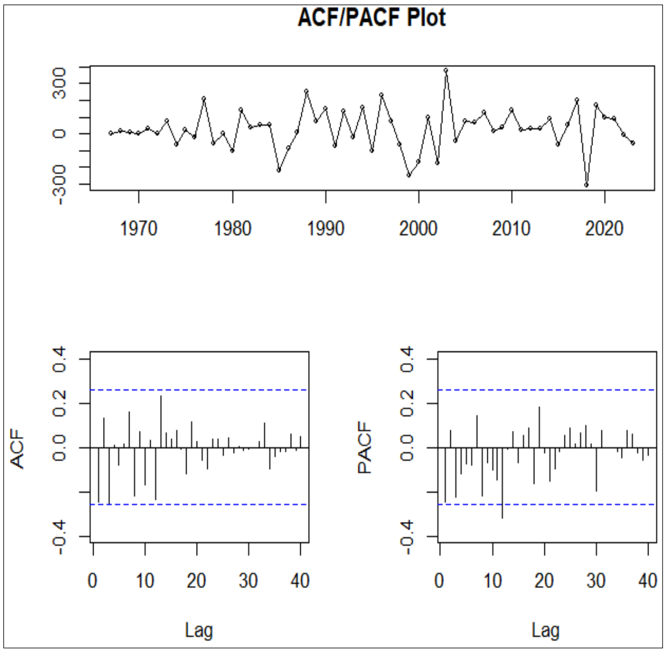

Fig 1: ACF and PACF plots of first differenced cast or seed yield data

Plots of the ACF and PACF provide potential values for p and q. AICc, BIC, and RMSE are used to build and compare various ARIMA models. The fitness matrices and some of the best fitted models are shown in Table 1.

Table 1: Best fitted ARIMA models for Castor seed Yield

Model	AICc	BIC	RMSE
ARIMA (1,0,1) with non-zero mean	713.45	720.85	116.9361
ARIMA (1,0,0) with non-zero mean	712.14	717.82	118.0074
ARIMA (1,0,0) with zero mean	714.61	718.47	123.0049
ARIMA (0,0,1) with non-zero mean	712.42	718.1	118.3005

Among the top models With a non-zero mean, ARIMA (1,0,0) is the best-fitted model with the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model is carried out and Fig. 2. gives the residual plots.

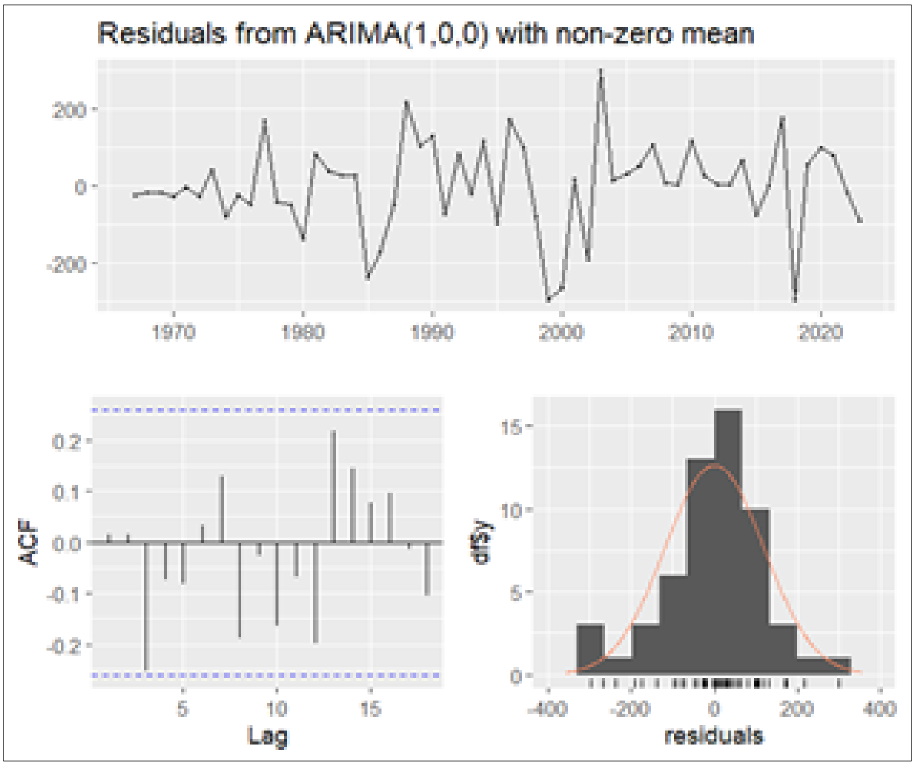

Fig 2: Residual plots for ARIMA (1,0,0) model with a non- zero mean.

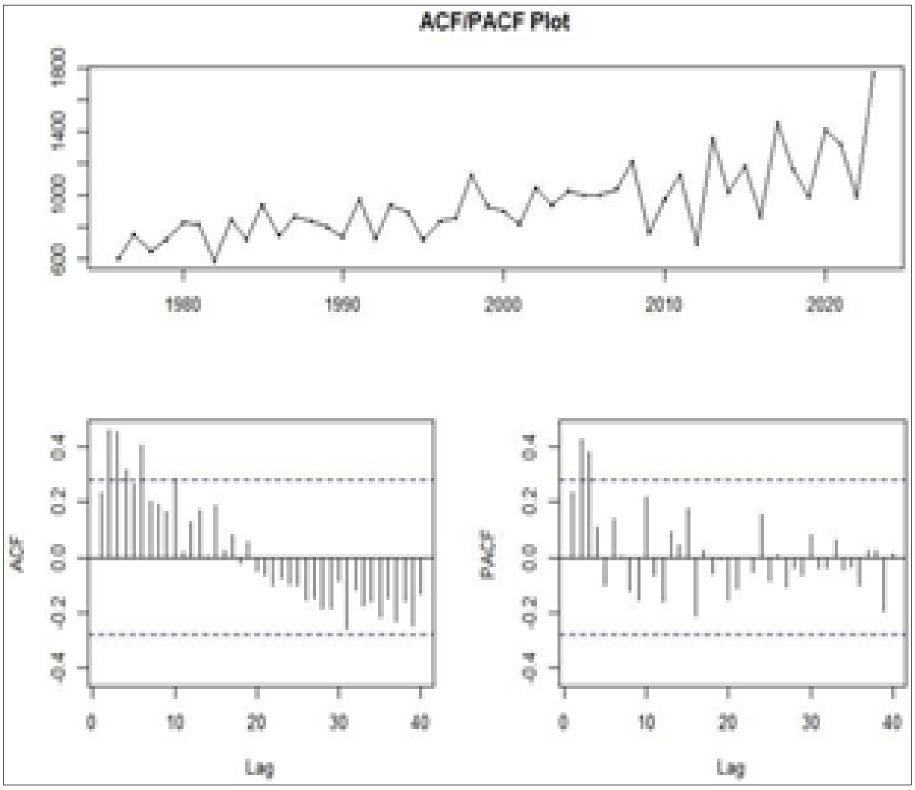

Fig 3: ACF and PACF plots of first differenced groundnut yield data

To determine if autocorrelation exists in the residuals for the best-fitted model, the Ljung-Box test is employed. The calculated p-value (0.3188) is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (1,0,0) model with a non-zero mean. The model is then used to predict the yield of castor seed for the next 5 years. Table 2 provides the forecasted values Plots of the ACF and PACF gives possible values for p and q. The fitness matrices and some of the best fitted models are shown in Table 3.

Table 2: Forecasted castor seed yield in India using ARIMA (1,0,0) model with a non-zero mean

2024	2025	2026	2027	2028
1936.97	1960.69	1990.60	2019.00	2047.76

Groundnut

Groundnut yield data from 1976 to 2023 is taken in this forecast study., Augmented Dicky Fuller test (ADF test) is performed to check the stationarity of data since obtained p-value is 0.4991, data was not stationary, in order to make it stationary the differencing approach was considered. It was found that the first differenced data was stationary by performing the ADF test, (p-value: 0.01). The ACF and PACF plots of the differenced data are depicted in Fig. (3).

Table 3: Best fitted ARIMA models for Groundnut Yield.

Model	AICc	BIC	RMSE
ARIMA (0,0,3) with non-zero mean	613.02	620.81	141.7594
ARIMA (1,0,2) with non-zero mean	613.06	620.84	141.8177
ARIMA (0,0,2) with non-zero mean	610.98	617.43	142.3844
ARIMA (2,0,2) with non-zero mean	615.15	624.15	140.9101

The best-fitted model among the top models is ARIMA (0,0,2) with a non-zero mean having the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model are also performed and Fig. 4. gives the residual plots.

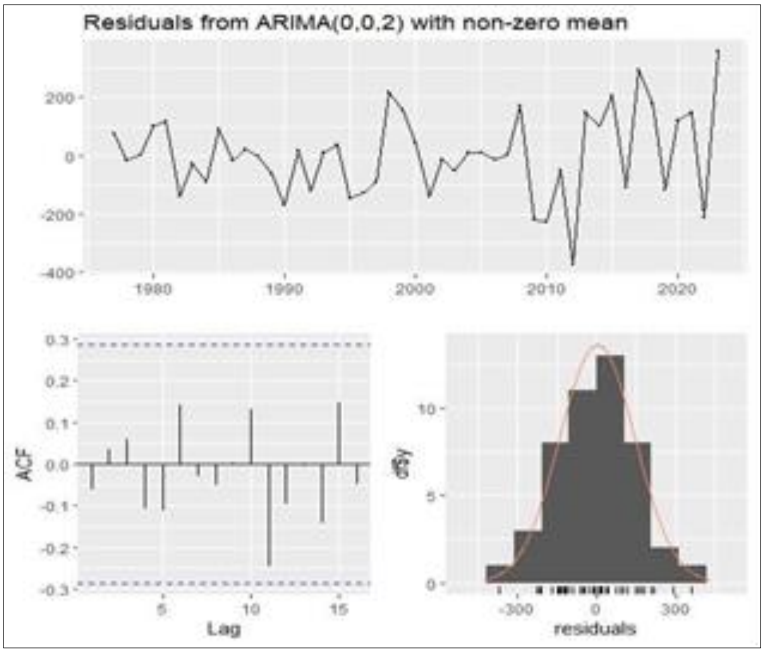

Fig 4: Residual plots for ARIMA (0,0,2) model with a non- zero mean.

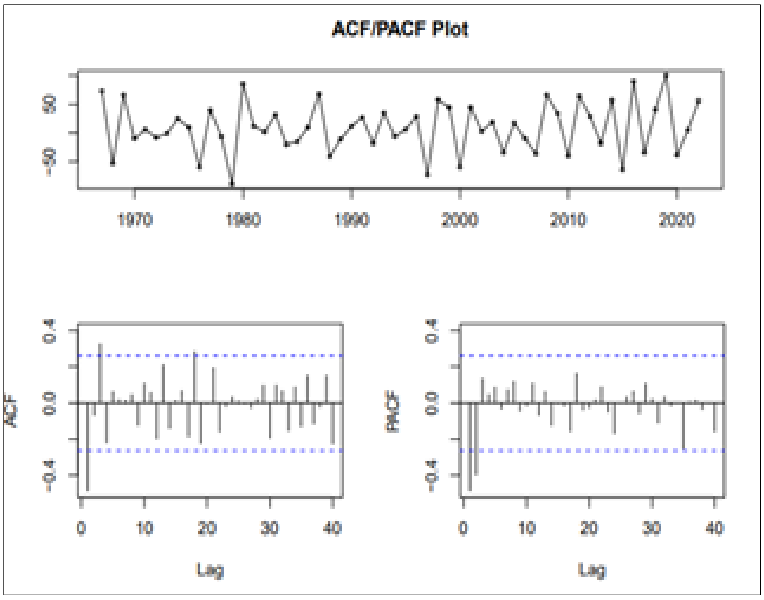

Fig 5: ACF and PACF plots of first differenced linseed yield data

The Ljung-Box test is used to check whether any autocorrelation exists in the residuals for the best-fitted model. The calculated p-value is 0.8755 which is larger than the standard significance level of 0.05, that indicates there is no significant autocorrelation present in the residuals of the ARIMA (0,0,2) model with a non-zero mean. The model is then utilized to predict the yield of Groundnut for the next 5 years. Table 4 shows the forecasted values

Plots of the ACF and PACF provide likely values for p and q. AICc, BIC, and RMSE are used to compare various ARIMA models. The fitness matrices and some of the best fitted models are shown in Table 5.

Table 4: Forecasted Groundnut yield in India using ARIMA (0,0,2) model with a non-zero mean

2024	2025	2026	2027	2028
1966.44	2003.77	2026.31	2048.86	2071.40

Linseed

The study was conducted using linseed yield data spanning from 1967 to 2023. To know if the data is stationary, ADF test is used. When the data was not stationary at first (p-value:0.99), the differencing approach was used to make it so. Using the ADF test, it was discovered that the first differenced data was stationary (p-value: 0.01). The ACF and PACF graphs of the differenced data are shown in Fig. 5.

Table 5: Best fitted ARIMA models for Linseed Yield.

Model	AICc	BIC	RMSE
ARIMA (2,1,0) with drift	564.14	571.45	33.94583
ARIMA (1,0,1) with drift	585.58	589.4	43.07132
ARIMA (2,1,1) with drift	565.42	574.35	33.5812
ARIMA (0,0,1) with non-zero mean	569.76	578.18	34.81209

To select the best-fitted model the model that is selected should have the lowest AICc, BIC, and RMSE values. So from the consideration, ARIMA (2,1,0) with drift is the best-fitted model. Residual analysis of the best fit model is carried out and Fig 6 shows the residual plots. The Ljung-Box test is preferred, to determine whether autocorrelation exists in the residuals for the best-fitted model

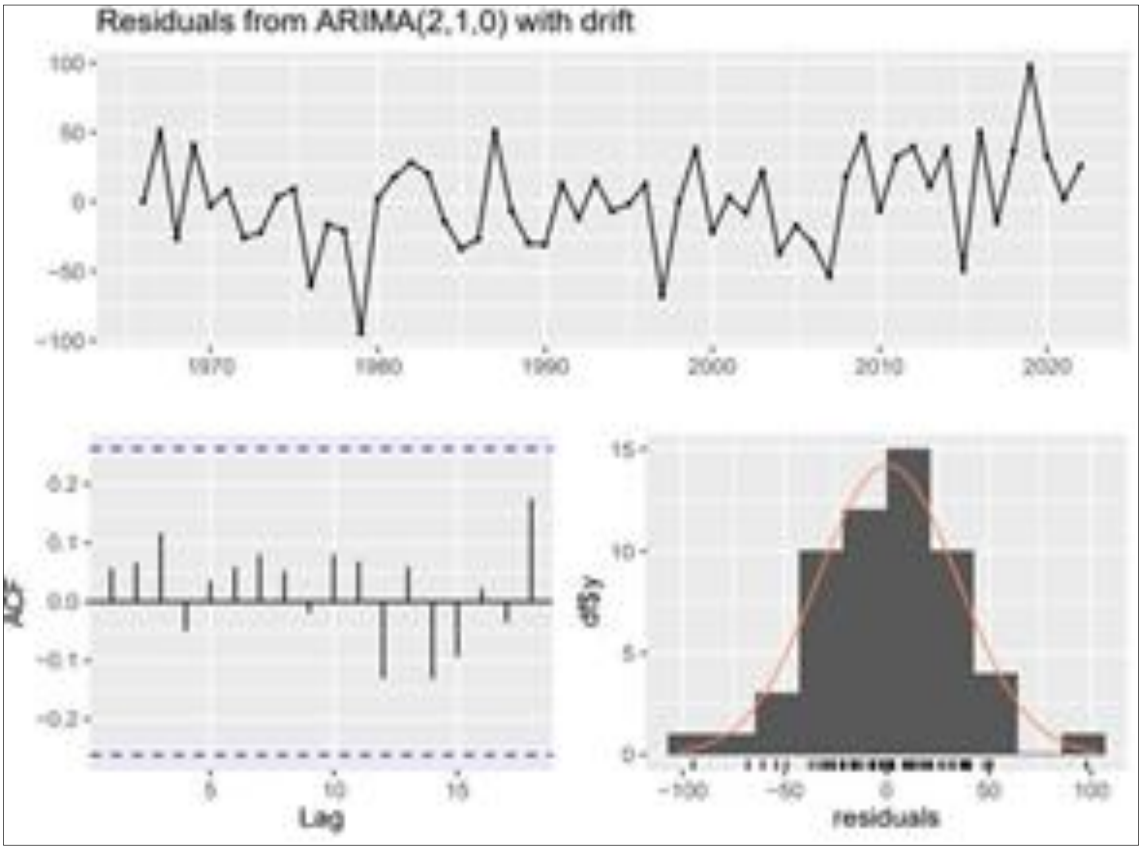

Fig 6: Residual plots for ARIMA (2,1,0) model with drift

The calculated p-value (0.9405) is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (2,1,0) model with drift. The model is then used to predict the yield of linseed for the next 5 years. Table 6 provides the forecasted values

Table 6: Forecasted castor seed yield in India using ARIMA (2,1,0) model with draft

2024	2025	2026	2027	2028
685.94	705.67	705.002	715.38	726.42

Niger Seed

The study makes use of niger seed yield data from 1976 to 2023. The ADF test resulted that the data was not stationary at first (p-value: 0.5263) then the differencing approach was used to make it so. It was discovered that the first differenced data was stationary (p-value: 0.01) using the ADF test. The ACF and PACF graphs of the differenced data are shown in Fig. 7.

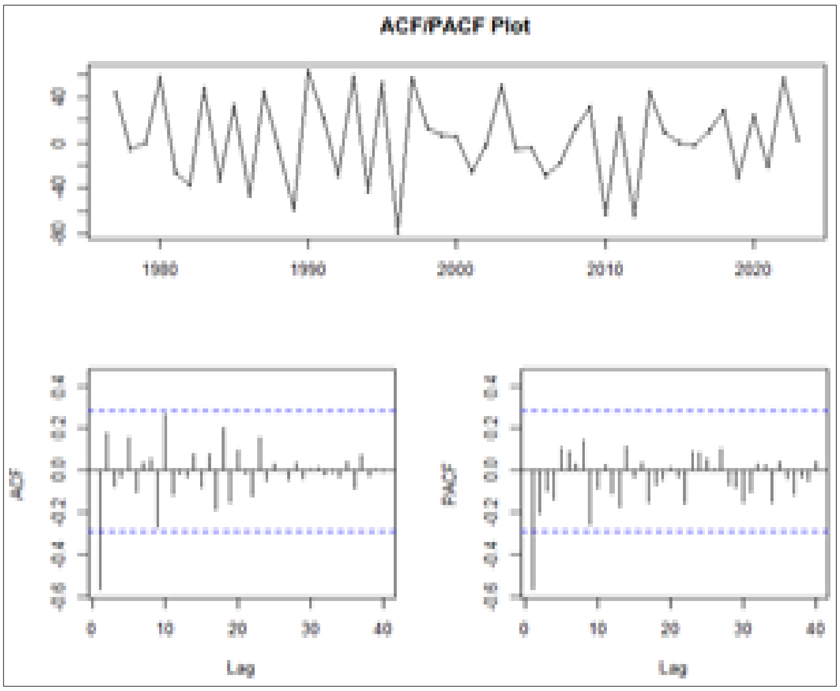

Fig 7: ACF and PACF plots of first differenced ni ger seed yield data

Plots of the ACF and PACF provide possible values for p and q. AICc, BIC, and RMSE are used to build and compare various ARIMA models. The fitness matrices and some of the best fitted models are shown in Table.7.

Table 7: Best fitted ARIMA models for Niger seed Yield

Model	AICc	BIC	RMSE
ARIMA (0,1,1)	458.58	462	29.9231
ARIMA (0,1,0)	474.74	476.5	36.5502
ARIMA (0,1,0) with drift	476.51	476.51	36.3907
ARIMA (1,1,0) with drift	460.65	465.64	29.8787

Among the top models ARIMA (0,1,1) is the best-fitted model with the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model is carried out and Fig.8 gives the residual plots.

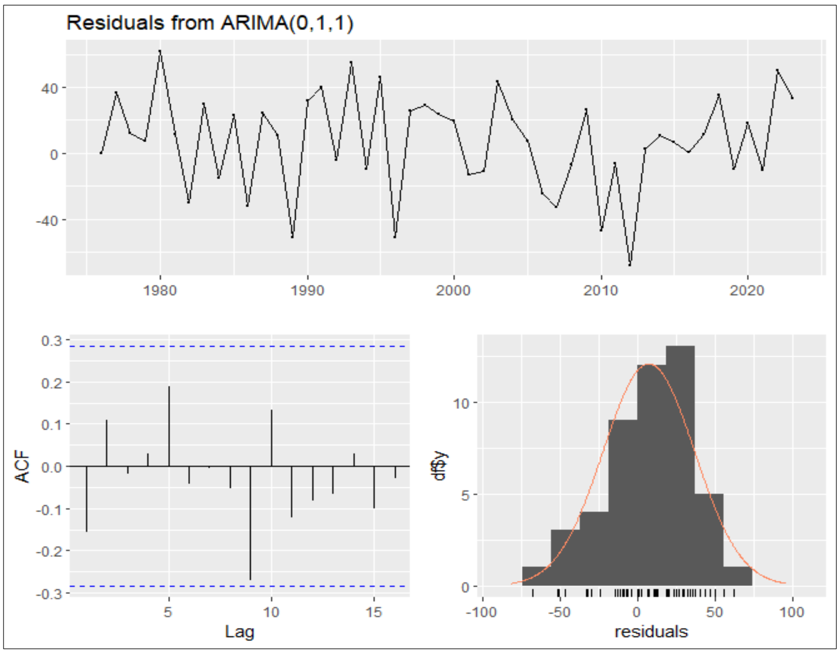

Fig 8: Residual plots for ARIMA (0,1,1) model with a non- zero mean.

From the Ljung-Box test, the calculated p-value (0.3695) is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (0,1,1) model. The model is then used to predict the yield of nigerseed for the next 5 years. Table 8

Table 8: Forecasted niger seed yield in India using ARIMA (0,1,1) model

2024	2025	2026	2027	2028
307.673	307.673	307.673	307.673	307.673

Rapeseed-Mustard

Rapeseed-mustard yield data from 1967 to 2022 is obtained for the study. The ADF test is used to determine if the data is stationary. Initially the data was not stationary (p-value: 0.3191), the differencing approach was utilized to make it stationary. Using the ADF test, it was discovered that the first differenced data was stationary (p-value: 0.01). The ACF and PACF graphs of the differenced data are shown in Fig. 9.

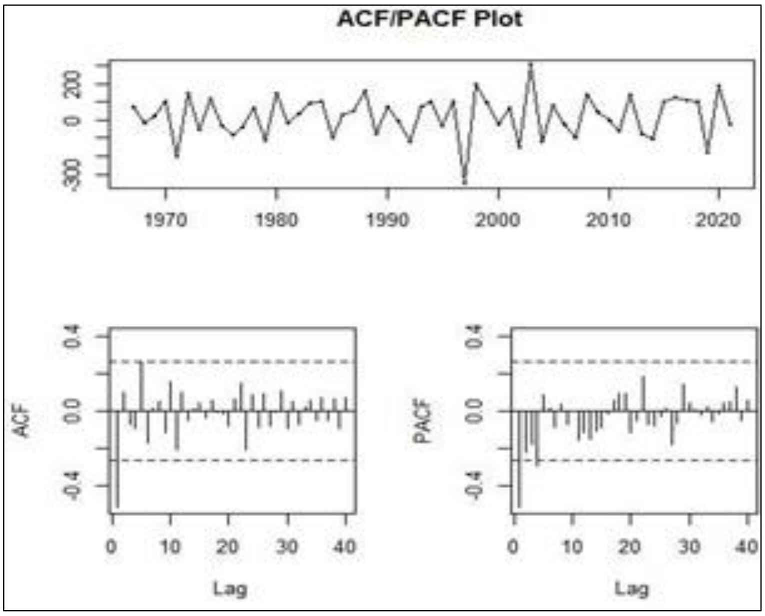

Fig 9: ACF and PACF plots of first differenced Rapeseed- Mustard yield data

Plots of the ACF and PACF provide potential values for p and q. For building and comparing various ARIMA models, AICc, BIC, and RMSE are used. The fitness matrices and some of the best fitted models are shown in Table 9.

Table 9: Best fitted ARIMA models for Rapeseed- Mustard Yield

Model		AICc	BIC	RMSE
ARIMA with drift	(1,1,1)	664.87	672.10	92.783
ARIMA with drift	(1,1,2)	667.23	676.04	92.733
ARIMA with drift	(0,1,1)	662.55	668.10	92.778
ARIMA with drift	(0,1,2)	664.87	672.10	92.782

Among the top models with a drift, ARIMA (0,1,1) is the best-fitted model with the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model is carried out and Fig. 10 gives the residual plots.

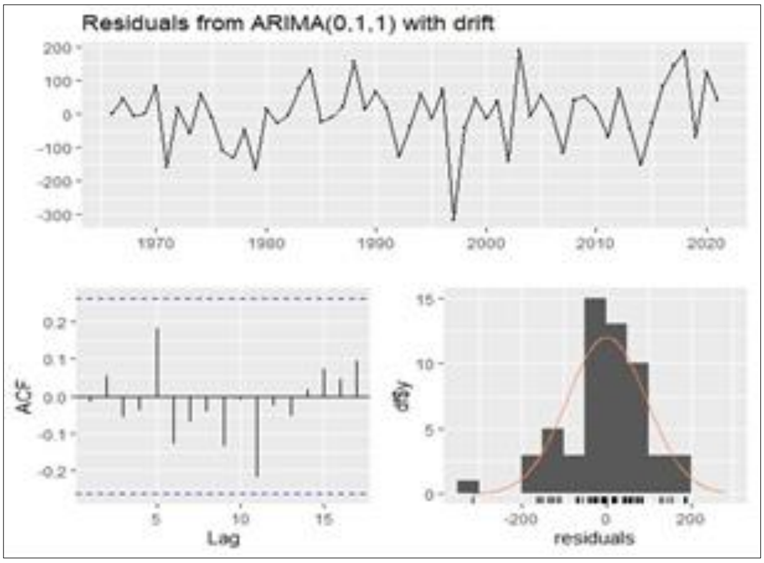

Fig 10: Residual plots for ARIMA (0,1,1) model with a non-zero mean.

The p-value (0.8083) obtained from the Ljung-Box test is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (0,1,1) model with drift. The model is then used to predict the yield of Rapeseed-Mustard for the next 5 years. Table 10 provides the forecasted values

Table 10: Forecasted Rapeseed-Mustard yield in India using ARIMA (0,1,1) model with drift

2022	2023	2024	2025	2026
1486.29	1505.24	1524.18	1543.12	1562.07

Safflower: The study makes use of safflower yield data spanning from 1966 to 2022. Using the ADF test, it was discovered that the data was stationary (p-value: 0.04). The ACF and PACF graphs of the differenced data are shown in Fig. 11.

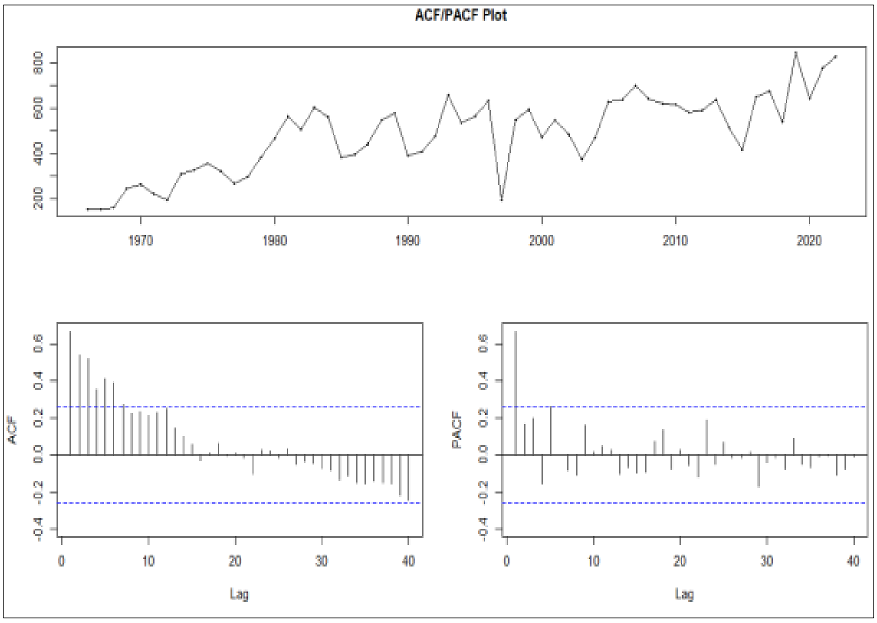

Fig 11: ACF and PACF plots of Safflower yield data

Plots of the ACF and PACF provide potential values for p and q. To build and compare various ARIMA models, AICc, BIC, and RMSE are used. The fitness matrices and some of the best fitted models are shown in Table 11.

Table 11: Best fitted ARIMA models for Safflower Yield.

Model	AICc	BIC	RMSE
ARIMA (0,1,0) with drift	701.09	703.11	123.2327
ARIMA (0,1,1) with drift	686.49	692.57	103.5871
ARIMA (1,1,0) with drift	696.08	702.15	113.5566
ARIMA (0,1,2) with drift	702.56	706.61	122.6522

Among the top models with a drift, ARIMA (0,1,1) is the best-fitted model with the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model is carried out and Fig. 12 gives the residual plots.

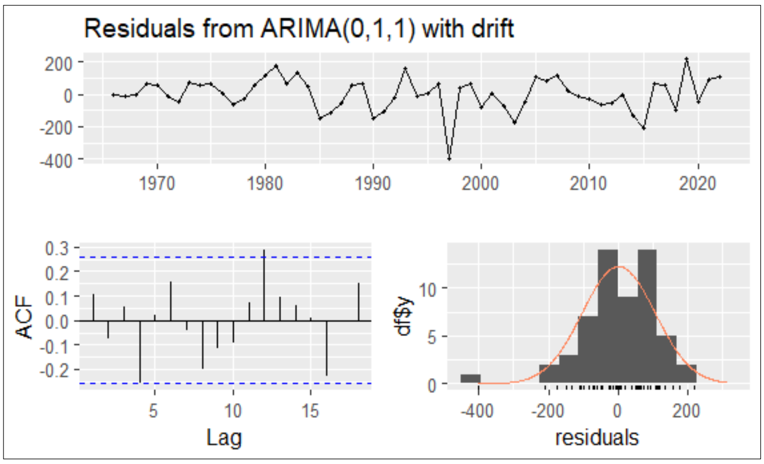

Fig 12: Residual plots for ARIMA (0,1,1) model with a drift.

The p-value (0.2431) resulted from the Ljung-Box test is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (0,1,1) model with drift. The model is then used to predict the yield of safflower seed for the next 5 years. Table 12 provides the

Sesamum

The study makes use of sesamum seed yield data from the period 1967 to 2023. At first the data was not stationary (p-value: 0.0535), then the differencing approach was used. Using the ADF test, it was found that the differenced data was stationary (p-value: 0.01). The ACF and PACF graphs of the differenced data are shown in Fig. 13.

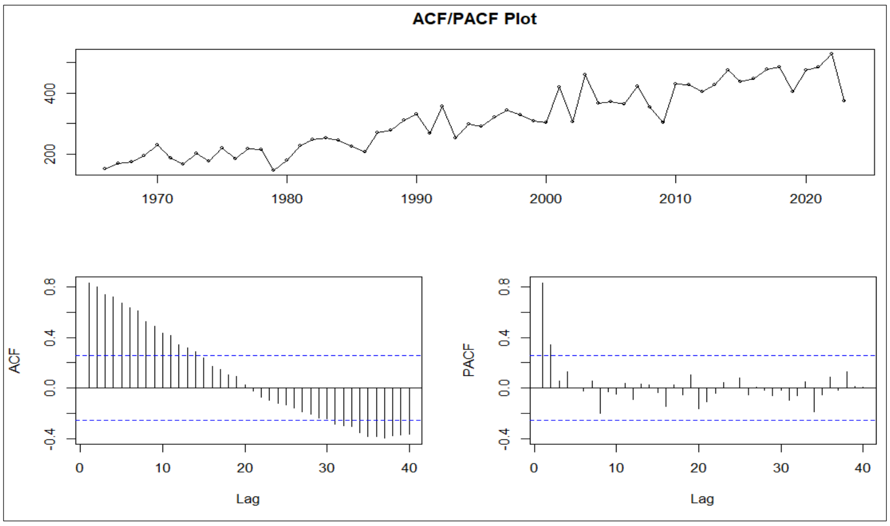

Fig 13: ACF and PACF plots of first differenced sesamum seed yield data

Plots of the ACF and PACF provide potential values for p and q. AICc, BIC, and RMSE are used to build and compare various ARIMA models. The fitness matrices and some of the best fitted models are shown in Table 13.

Table 13: Best fitted ARIMA models for Sesamu m seed Yield.

Model	AICc	BIC	RMSE
ARIMA (0,1,0)	622.88	624.85	55.59196
ARIMA (3,1,0) with drift	599.66	608.7	41.49848
ARIMA (0,1,0) with drift	624.75	628.61	55.4565
ARIMA (2,1,0) with drift	603.28	604.05	44.20401

Table 12: Forecasted safflower yield in India using ARIMA (0,1,1) model with a drift forecasted values

2023	2024	2025	2026	2027
753.95	764.12	774.30	784.47	794.65

Among the top models with drift, ARIMA (3,1,0) is the best-fitted model with the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model is carried out and Fig 14 gives the residual plots

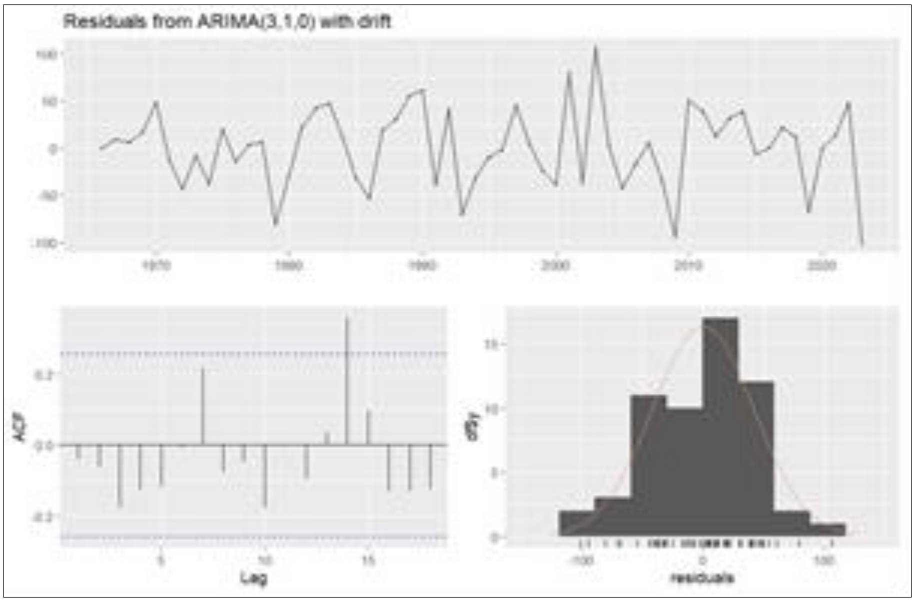

Fig 14: Residual plots for ARIMA (3,1,0) model with drift.

The calculated p-value (0.1757) from the Ljung-Box test is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (3,1,0) model with drift. The model is then used to predict the yield of sesamum seed for the next 5 years. Table 14 provides the forecasted values

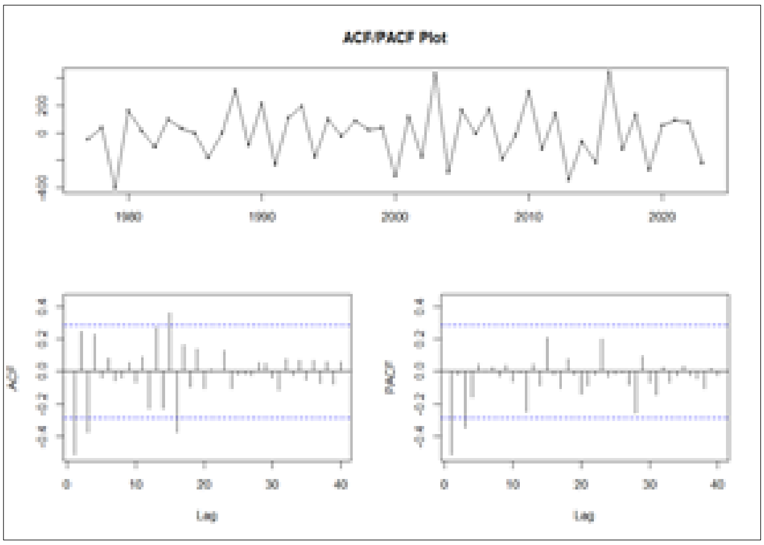

Fig 15: ACF and PACF plots of first differenced soybean seed yield data

Table 14: Forecasted sesamum seed yield in India using ARIMA (3,1,0) model with drift

2024	2025	2026	2027	2028
493.015	476.602	492.594	460.535	499.694

Soybean

The study makes use of soybean seed yield data spanning from 1976 to 2023. The data was found to be not stationary at first (p-value:0.49) using the ADF test, then differencing approach was used. It was discovered that the first differenced data was stationary (p-value: 0.01). The ACF and PACF graphs of the differenced data are shown in Fig. 15. Plots of the ACF and PACF provide potential values for p and q. AICc, BIC, and RMSE are used to build and compare various ARIMA models. The fitness matrices and some of the best fitted models are shown in Table 15. Among the top models, ARIMA (0,1,1) is the best-fitted model with the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model is carried out and fig 2 shows the residual plots

Table 15: Best fitted ARIMA models for soybean seed

Model	AICc	BIC	RMSE
ARIMA (0,1,1) with drift	616.5	621.49	156.3238
ARIMA (0,1,1)	614.43	617.86	156.8099
ARIMA (0,1,0) with drift	631.44	634.87	189.1537
ARIMA (0,1,0)	629.26	631.02	189.1588

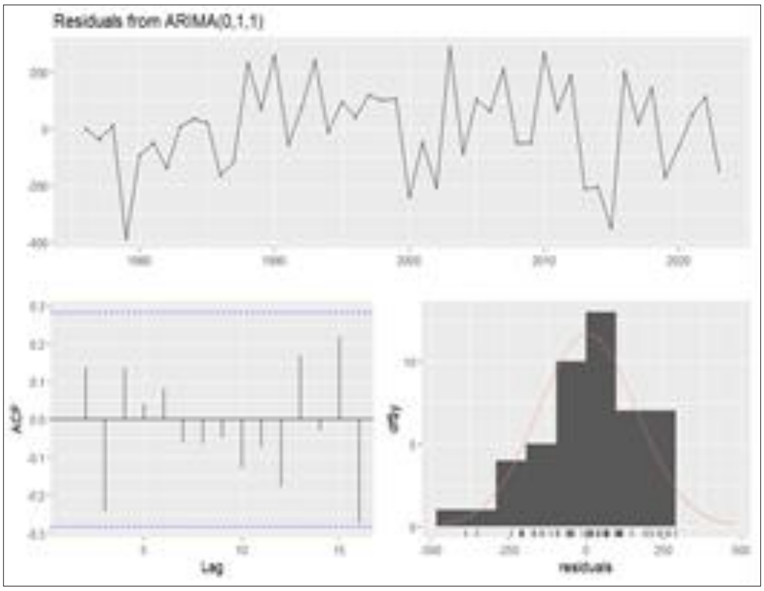

Fig 16: Residual plots for ARIMA (0,1,1) model Yield.

The Ljung-Box test is performed, to check if autocorrelation exists in the residuals for the best-fitted model. The calculated p-value (0.6109) is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (0,1,1) model. The model is then used to predict the yield of soybean seed for the next 5 years. Table 16 provides the forecasted values.

Table 16: Forecasted soybean seed yield in India using ARIMA (0,1,1) model

2024	2025	2026	2027	2028
1022.38	1022.38	1022.38	1022.38	1022.38

Sunflower

The study makes use of sunflower yield data spanning from 1976 to 2023. When the data was not stationary at first (p-value: 0.3053), the differencing approach was carried out to make it so. Using the ADF test, it was discovered that the first differenced data was stationary (p-value: 0.01). The ACF and PACF graphs of the differenced data are shown in Fig.17.

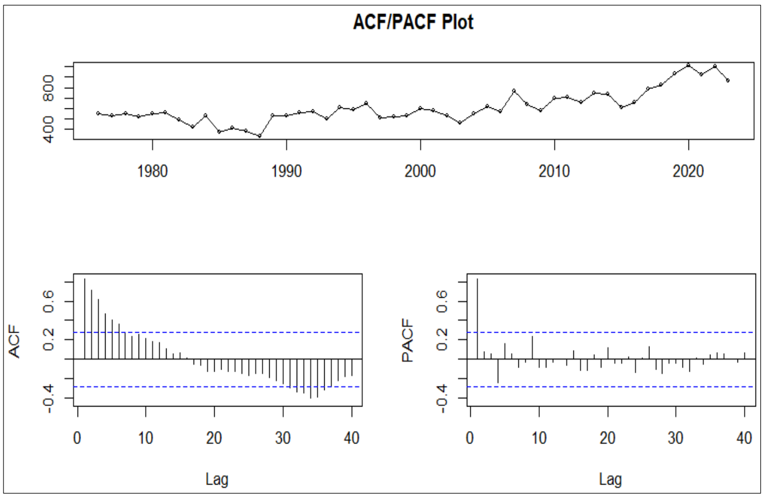

Fig 17: ACF and PACF plots of first differenced sunflower yield data

Plots of the ACF and PACF provide potential values for p and q. AICc, BIC, and RMSE are used to build and compare various ARIMA models. The fitness matrices and some of the best fitted models are shown in Table 17.

Table 17: Best fitted ARIMA models for Sunflower

Model	AICc	BIC	RMSE
ARIMA (0,1,0) with drift	584.67	588.05	131.4903
ARIMA (1,1,0) with drift	567.82	572.74	106.2912
ARIMA (0,1,1) with drift	547.46	552.38	82.08091
ARIMA (0,1,1)	545.51	548.52	82.384

Among the top models, ARIMA (0,1,1) is the best-fitted model with the lowest AICc, BIC, and RMSE values. Residual analysis of the best fit model is carried out and Fig. 18 gives the residual plots.

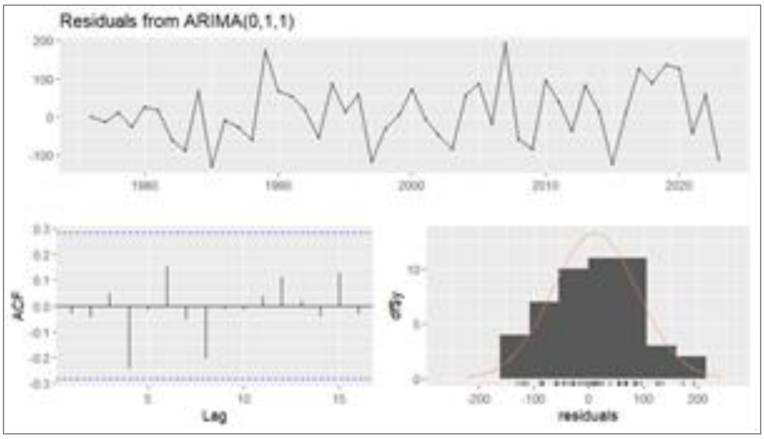

Fig 18: Residual plots for ARIMA (0,1,1) model

To determine if autocorrelation exists in the residuals for the best-fitted model, the Ljung-Box test is employed. The calculated p-value (0.5684) is larger than the standard significance level of 0.05, indicating that there is no significant autocorrelation present in the residuals of the ARIMA (0,1,1) model with a non-zero mean. The model is then used to predict the yield of sunflower seed for the next 5 years. Table 18 provides the forecasted values

Table 18: Forecasted castor seed yield in India using ARIMA (0,1,1) model with a non-zero mean

2024	2025	2026	2027	2028
903.699	903.699	903.699	903.699	903.699

Conclusion

The study used ARIMA modeling to predict oilseed yields based on historical data. The model accurately captured temporal patterns and seasonality, demonstrating satisfactory accuracy in predicting future yields. The use of metrics like AICc and BIC helped select the most appropriate model, considering goodness of fit and model complexity. The RMSE metric provided insights into the model's predictive performance, demonstrating its ability to closely approximate observed values. The robustness of the ARIMA model was confirmed through residual analysis, confirming its adherence to underlying assumptions and reliability for forecasting oilseed yields. This study demonstrates the viability of

ARIMA modeling in agricultural planning and decision-making processes.

References

1. Chakrabarty S, Kalam Mohammad Aminul Islam A, Yaakob Z. (Castor (Ricinus communis): An Underutilized Oil Crop in the South East Asia. Agroecosystems – Very Complex Environmental Systems; c2021.
2. Etherton PM, Hu FB, Ros E, Sabate J. The role of tree nuts and groundnuts in the prevention of coronary heart disease: multiple potential mechanisms. Journal of Nutrition. 2008;138(9):1746-1751.
3. Sabat J, Oda K, Ros E. Nut consumption and blood lipid levels: a pooled analysis of 25 intervention trials. Archives of Internal Medicine. 2010;170(9):821-827.
4. Patel RK, Tomar GS, Dwivedi SK. Effect of irrigation scheduling and nitrogen levels on growth, yield and water productivity of linseed (*Linum usitatissimum* L.) under Vertisols. Journal of Applied and Natural Science. 2017;*9*(2):698-705.
5. Weiss EA. Oilseed Crops. Blackwell Science, Oxford, UK; c2000.
6. Robbelen G, Downey KR, Ashri A. Oil Crops of The World. McGraw-Hill Inc, USA; c1989.
7. Elfadl EY, Reinbrecht C, Frick C, Witzke-Ehbrecht S, Rudolphi S, Claupein W, *et al*. Genotype by environment interaction in safflower (*Carthamus tinctorius* L.) grown under organic farming system in Central Europe. In: Esendal, E. (Ed.), Proceedings of the 6 th International Safflower Conference, Istanbul, Turkey; c2005. p. 39-43.
8. Kundatiri, Chetan, Maitra, Sagar, Shankar, Tanmoy, *et al*. Growth and productivity of sesame (*Sesamum indicum* L.) as influenced by spacing and nitrogen levels. Crop Research. 2022;57:190-194.
9. Agarwal DK, Billore SD, Sharma AN, Dupare BU, Srivastava SK. Soybean: Introduction, Improvement and Utilization in India- Problems and Prospects. Agric Res. 2013;2:293-300
10. Skoric D, Jocic S, Lecic N, Sakac Z. Genetic possibilities for altering sunflower oil quality to obtain novel oils. Can. J Physiol. Pharmacol. 2008;86(4):215-221
11. Venu, Vaisakh, Sreenath B, Ramdas RE. Various Mathematical Models in Agricultural Engineering. Current Journal of Applied Science and Technology. 2023;42(41):13-20. ISSN 2457-1024
12. Bowlekar AP, Sathian KK, Venu, Vaisakh. Rainfall Versus Water Balance Components Using R Software: A Case Study on a Micro-watershed in the Mid-lands of Kerala. Journal of Experimental Agriculture International. 2023;45(10):355-366. ISSN 2457-0591
13. Venu, Vaisakh R, Vipin P, Prajitha KN. A Comparative Analytical Study of Many Regression Model Approaches, Arima Model and a Hybrid Model for Forecasting Area, Production, and Productivity of Coconut in Kerala, India. Current Journal of Applied Science and Technology. 2023;42(40):37-49. ISSN 2457-1024
14. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; c2015.

Author's Details

N Revathi
M.Tech Scholars, Department of Soil and Water Conservation Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, 0Kerala, India

Dixita Gourshetty
M.Tech Scholars, Department of Soil and Water Conservation Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

V Haritha
M.Tech Scholars, Department of Processing and Food Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Noorbina Razak
M.Tech Scholars, Department of Processing and Food Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Arsha Sugathan
M.Tech Scholars, Department of Processing and Food Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Sankar M
M.Tech Scholars, Department of Processing and Food Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Ashitha Thomas
M.Tech Scholars, Department of Processing and Food Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Aveen Ragh
M.Tech Scholar, Farm Machinery and Power Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Humbare Mrunalini Dinkar
Ph.D Scholar, Department of Soil and Water Conservation Engineering, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India

Vaisakh Venu
Assistant Professor, Department of Basic Engineering and Applied Sciences, Kelappaji College of Agricultural Engineering and Technology, Tavanur, Kerala Agricultural University, Kerala, India