AATCC Review

20 April 2025: Received
10 June 2025: Revised
17 June 2025: Accepted
19 July 2025: Available Online

https://aatcc.peerjournals.net/

Original Research Article | Open Access

Yield forecasting of groundnut in Bihar through Auto-Regressive Integrated Moving Average (ARIMA) models

Ezhilmathi S[1], Mahesh Kumar*[1], Aarti Kumari[2], Ajeet Kumar[1], Sudhir Paswan[1] and Kiran[1]

[1]College of Basic Sciences and Humanities (CBS&H), Dr. Rajendra Prasad Central Agricultural University (RPCAU), Pusa, Bihar, 848125, India

[2]College of Agriculture, Dr. Rajendra Prasad Central Agricultural University (RPCAU), Pusa, Bihar, 848125, India

ABSTRACT

The present study entitled "Yield forecasting of groundnut in Bihar through Auto-Regressive Integrated Moving Average (ARIMA) models" is based on the ARIMA models for forecasting groundnut yield in Bihar. The secondary data on groundnut yield were collected from the year 1980 to 2018 from the Directorate of Groundnut Research, Directorate of Oilseeds Development and India Agril. Stat. The data from 1980 to 2016 were used for analysis of forecasting groundnut yield and the data for 2017 to 2018 were kept for model validation. Instead of conventional or econometric methods, the ARIMA models were used to forecast the productivity of groundnut in Bihar. The time series data of 37 years from 1980 to 2016 were used for the study. Models ARIMA (0,1,1), ARIMA (0,1,2), ARIMA (0,0,1), ARIMA (1,0,0), ARIMA (1,0,1), ARIMA (1,1,1), ARIMA (2,0,0) and ARIMA (2,0,1) were built. The parameters of all these models were computed and tested for their significance. Various statistics were also computed for selecting the adequate and parsimonious model i.e., t-test and chi-square test. This is supported by low values of MAPE, MAE, RMSE and BIC for forecasting of groundnut yield in Bihar. Using the selected ARIMA models ARIMA (1,0,1) the yield values were forecasted for five five-year period ahead i.e. from 2017 to 2021 in Bihar. The forecasted values of Bihar are 1026.72 kg/ ha, 1028.74 kg/ ha, 913.00 kg/ ha, 913.00 kg/ ha and 913.00 kg/ ha, respectively for 2017, 2018, 2019, 2020 and 2021. The forecasted values of Bihar exhibit an increasing trend, for 2017 and 2018, in the yield of groundnut. These yield values were presented along with their lower and upper limits with 95% confidence interval. Using the mathematically sound ARIMA models, the groundnut yield values were forecasted with 0.90 percent of one step ahead forecast errors for Bihar. The two steps ahead forecast errors are 1.69 per cent for Bihar. All the 8 models were subjected to critical examination. Among them ARIMA (1,0,1) model was chosen as it is stationary, invertible, parsimonious, stable and has minimum error. Thus, the forecast model for groundnut productivity in Bihar is,

$$Z_t - Z_{t-1} = 6.879 + 0.855\,(z_{t-1} - z_{t-2}) - 0.354\,(a_{t-1} - a_{t-2}) + a_t$$

Keywords: *forecasting of groundnut yield, ARIMA models, forecasting of crop yield..*

INTRODUCTION

Groundnut (*Arachis hypogaea* Linn.), also known as peanut, moongfali, monkey-nut or earthnut, is an annual herbaceous legume valued globally for its high oil and protein content. It belongs to the Fabaceae family and is a rich source of energy [7, 9, 11], providing approximately 5.6 to 5.8 calories per grain (raw and roasted respectively), along with essential amino acids, minerals, and vitamins. Groundnut pods, containing two to three seeds, have a hard shell and a shelling percentage ranging from 60 to 75%. Globally, groundnut is cultivated in over 80 countries across tropical, subtropical, and warm temperate regions, mainly between 40°N and 40°S latitudes. As per FAO estimates, global acreage and production stand at approximately 285 lakh hectares and 459.51 lakh tonnes, respectively, with an average yield of 1611 kg/ha. China, India, and Nigeria are the leading producers, together contributing over half of global output. Argentina leads in exports, followed by the U.S., India, China, and Brazil. In 2019–20, groundnut oil production accounted for 6.05 million metric tonnes—2.97% of global vegetable oil output. Historically introduced to India via the Philippines and South America, groundnut gained significance during the 16th century, especially in Tamil Nadu's South Arcot region. By the late 19th century, it spread across Maharashtra, Gujarat, Andhra Pradesh, and Karnataka. Today, India ranks first globally in groundnut acreage (approximately 70 lakh ha annually) and second in production (80–85 lakh MT), though its average yield lags behind major producers like China and the U.S. Gujarat leads India's production (40.1%), followed by Rajasthan, Andhra Pradesh, Karnataka, and Maharashtra. About 80% of cultivation occurs during the kharif season under rainfed conditions. In 2018–19, groundnut occupied 47.31 lakh ha in India, producing 67.27 lakh tonnes at an average yield of 1422 kg/ha. Export-wise, India shipped 6.64 lakh MT of peanuts worth ₹5,096 crore to Southeast Asian countries in 2019–20 [2,15]. The ICAR–Directorate of Groundnut Research (ICAR-DGR) at Junagadh, Gujarat, plays a central role in varietal development and technology dissemination, having released over 215 improved varieties and several agronomic innovations.

*Corresponding Author: **Mahesh Kumar***

DOI: https://doi.org/10.21276/AATCCReview.2025.13.03.431

In Bihar, groundnut remains a minor crop with a yield of just 1 kg/ha in kharif 2018–19, due to limited area (860 ha), rainfed dependence, erratic rainfall, old cultivars, and biotic stresses. Weeds, diseases, and aflatoxin contamination from Aspergillus flavus further constrain productivity. Recognizing groundnut's economic potential and the need for yield enhancement, the present study explores forecasting groundnut yield in Bihar using Auto-Regressive Integrated Moving Average (ARIMA) models to aid future planning and intervention.

Review for forecasting of yield by ARIMA model

Scientists used the minimum Akaike's Information Criterion estimation procedure for deciding the order of an ARIMA model and checked it with the numerical examples [12]. Nochai R and Nochai T forecasted the pure oil price, farm price, and wholesale price of oil palm in Thailand. They predicted a time limit of 5 years since 2000. They selected the best models, considering the minimum MAPE values, as ARIMA (2,1,0), ARIMA (1,0,1) and ARIMA (3,0,0) for forecasting farm, wholesale and pure oil prices, respectively [11,20]. Many authors have analysed the annual soybean production data of India for 1970-2007 using ARIMA technique. They found ARIMA (2,1,0) as good model for predicting the production of the second largest oilseed crop of our country [19,21]. Researchers forecasted the area, production and yield of sugarcane in Tamil Nadu using a perfect fit of univariate Auto Regressive Integrated Moving Average (ARIMA) models. The crop data from 1950–2007 were utilized for this study. ARIMA (1, 1, 1) model is considered as best fit for sugarcane area and yield. ARIMA (2, 1, 2) is identified as suitable for predicting sugarcane production [18,22,26]. Padhan used ARIMA model to predict the yearly yield of 34 agricultural commodities, using the yearly data of yield from the year 1950 to 2010. He forecasted the yield for another 5 years since 2011 and verified the credibility of ARIMA model with Adj R^2, minimum AIC and lower MAPE values [13,25]. Debnath utilized the time series data for 1950 to 2010 to predict the values of area and production and yield, for the year 2020, of cultivated cotton in India. They used ARIMA (0,1,0), ARIMA (1,1,4) and ARIMA (0,1,1) as the best fitted models topredict the area, production and yield of cotton, respectively [4]. Liu proposed online learning algorithms for the estimation of ARIMA models which is applicable for higher computational efficiency and suitable to a wide range of applications. In addition to this they analysed the regret bounds of proposed algorithms and validated that their method was effective and robust [9]. Celik used ARIMA models to forecast the groundnut production in Turkey for the next fifteen years since 2016. They analysed the data from 1950 to 2015 and found the best-fit model ARIMA (0,1,1) among the other 6 models [3]. Jadhav demonstrated the utility of price forecasting of farm prices using univariate ARIMA techniques for major crops namely paddy, ragi and maize in Karnataka and validated the same using the Mean Squared Error, Mean Absolute Percentage Error and Theils U coefficient criteria [5]. Mahesh Kumar *et al.* took the data on the yield of sugarcane in Bihar. The crop data were taken from the year 1940 to 2010. The data was used in order to develop an ARIMA model and then validated using five years ofyield data from 2011 to 2015. ARIMA (0, 1, 1) model found as an appropriate model to predict the sugarcane yield in Bihar. They forecasted two years ahead yield of sugarcane in Bihar with less than 5% of prediction error [8]. Many scientists have study using ARIMA models to forecast the maize production in India for 4 years, 2018-2022. They analysed the Auto-Correlation Function and Partial Auto-Correlation Function of differenced series for the selection of more suitable model [16,23]. Siami-Namini compared the traditional algorithms for predicting time series data such as ARIMA model with innovatively created deep learning-based algorithms such as Long Short-Term Memory [17]. Biswas *et al.* used ARIMA model for short-term forecasting of the price of sunflower seeds for the market Kadiri of Anantpur district in the state of Andhra Pradesh. Mean Average Percentage Error and Root Mean Square Percentage Error for the selected model ARIMA (1,1,2) were 2.30% and 3.44%, respectively [1]. Scientists used ARIMA techniques to forecast the mustard yield in the study areas, Bhiwani and Hisar districts of Haryana, five years ahead. The validity of models checked by using MAPE, Akaike Information Criterion and Bayesian Information Criterion [6,14,24].

MATERIALS AND METHODS

The current study is yield forecasting of groundnut in the states of Bihar. The methodology is described as; Details of area under study; Data source; Techniques and tools employed in the analysis.

Details of area under study

The present study was experimented in Bihar the states situated in North India . The state Bihar is situated between 24°-20'-10" ~ 27°-31'-15" North latitude and 83°-19'-50" ~ 88°-17'-40" East longitude. Bihar joins the Ganga delta and Assam and it is located in sub-tropical to tropical. The North side located Himalayan Mountains plays an important role regarding the precipitation in Bihar. The state has the total area of 93.6 lakh hectares. That is about 3% of India's total geographical area. It is located in the Eastern India and the climate is subtropical. During the peak summer, from March- May, the average temperature will be nearly 40 degree Celsius. In the winter, from Dec- Jan, the temperature would drop down to 8 degree Celsius. The total area of Bihar, which is located 173 feet above sea level, is about 94,163 sq. km. The average number of rainy days in the state is 52.5 days in a year and the precipitation is about 976 mm. The Bihar plane mostly contains thick alluvial mantle of drift origin. The soil is mainly young loam rejuvenated year to year by continuous deposits of silt, clay and sand. These are carried by various river streams. Phosphoric acid, humus and nitrogen are deficient in the soil whereas lime and potash are in optimum amount. In total, Bihar consists of 38 districts in which agriculture and allied activities are of major concern. Cropping pattern of Bihar is almost stable and constant. The state is mainly a cereal economy where nearly 85% of Bihar's gross cropped area is cultivated with cereals. By the information provided by the state, Bihar accomplished self-sufficiency in food grain production. Particular schemes for production of oilseeds and pulses in rice fallow areas were exhibited in Bihar. That scheme widely known as 'Targeting Rice Fallow Areas (TRFA)' in the east part of India.

Data source

The present study is based on the secondary data on area, production and yield of groundnut. These datawere collected from the authenticated portals like Directorate of Groundnut Research, Directorate of Oilseeds Development and India Agri Stat.

Tools and techniques employed in the analysis

The groundnut crop data derived from the authenticated sources were used for the analysis as follow.

Method to analysed yield of Groundnut

ARIMA models for forecasting of yield

Box and Jenkins model (1976) is going to be applied for yield forecasting. Auto Regressive Integrated Moving Average (ARIMA) is the basic group of models for forecasting a time series. Various series coming in the forecasting equations are known as "Auto-Regressive" process. The showmanship of lags of the errors of forecast in the model is called "Moving Average" process. The ARIMA model is denoted by ARIMA (p,d,q), where 'p' stands for the order of the auto regressive process, 'd' is the order of the data stationary and 'q' is the order of the moving average process.

Auto Regressive process of order (p) is,

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

Moving Average process of order (q) is,

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

The basic formulation of ARIMA (p,d,q) could be described as,

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where, Y_t= yield (dependent variable) of groundnut at year t

Y_{t-1}, Y_{t-2}..., Y_{t-p} = response variable at time lags t-1, t-2..., t-p respectively; μ = constant mean of the process; ε_{t-1},ε_{t-2}..., ε_{t-q}= errors in the previous time periods; φ_s=coefficients to be estimated of AR process; θ_s =coefficients to be estimated of MA process; ε_t= forecast error, independently and normally distributed with zero mean &constant variance for t = 1, 2,..., n; d = fraction differenced during the interpretation of AR and MA.

The Box-Jenkins modeling procedure

Instead of any other traditional econometric methods, mathematically sound and reliable Box-Jenkins method is used for the forecasting. This method uses a series of stages in the ARIMA modeling procedure to create a model. The built models then tested for their accuracy by using the past data available. If the residues were little, contain useless information and distributed irregularly, the model is better fitted. If the model is not satisfactory, then the whole process should have been repeated once again to improve on the basic model by utilising the new available model. This series of procedure continues until a best fitted model has been found. The iterative phases (Fig-3.2) in developing anARIMA forecasting technique are found as:1. Model specification; 2. Model estimation; 3. Diagnostic checking; 4. Forecasting.

Model specification

The main goal in ARIMA modeling is to find out the best suitable values for p, d and q. This could be partly solved by analysing the Auto Correlation Function (ACF) and Partial Auto Correlation Functions (PACF) of the time series data (Pindyk & Rubinfeld, 1991). ACF indicates the order of autoregressive component 'q' of the model, while the PACF give an indication for the component 'p'. Initial stage is to check whether the data are stationary. The degree of the homogeneity, (d) i.e.no. of time series data to be differenced in order to result in a stationary series. It is decided based upon where the ACF fall out to zero. After deciding 'd', a stationary series, its ACF and PACF are analysed to determine the suitable values of p and q.

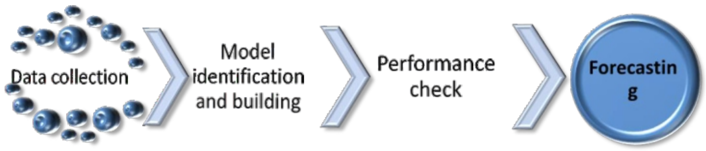

Fig-3.2: Steps in ARIMA forecasting

Model estimation

Next step is to estimate the model using computer package. The goal is to get the estimates of the tentatively built ARIMA model parameters in step 1, for the given values of p and q. The ARIMA coefficients (φ's and θ's) should be calculated with the use of a nonlinear least square procedure. The main important method of estimating the ARIMA models is called "Marquardt's compromise".

Diagnostic checking

Diagnostic checks are used to obtain the output. The initial diagnostic check is, residual analysis, to create a graph of time series plot of residuals. If the plot creates a rectangular scatter around a zero-horizontal level without any trend, applied model will be announced as normal. The second diagnostic check is testing of normality is. For first normality test, normal scores are marked against the residuals. If it made a straight line, applied model is declared proper fit. The second normality test is to plot a histogram of the residuals. Third check is identifying the fitness of good. For this purpose, residuals are marked against corresponding fitted values. Thus, the model will be announced as perfect fit when the plot shows no pattern.

Forecasting

After the assessment of the predicting capacity of fitted ARIMA model along with 95% confidence interval values. Five-years or less forecast is done because the forecasting errors increase rapidly if we go too far out in the future.

Trend fitting

Box-L-jung Q statistics is utilized to modify the non stationary data into stationarity data. Also used to evaluate the adequacy for the residuals. Additively various autocorrelations up to some numbers of lags are computed and their significance tested by Box-Ljung Q statistic. If none of the autocorrelations is significantly different from zero at any reasonable level, the particular ARIMA model should be proved as an appropriate model for forecasting. For checking the adequacy of AR, MA and ARIMA processes, various reliability statistics like R^2, stationary R^2, Root Mean Square Error (RMSE), Mean Absolute Percentage Error(MAPE), Mean Absolute Error (MAE)and Bayesian Information Criterion (BIC) [as suggested by Schwarz, 1978] were used. The model fitting statistics namely RMSE,MAE, MAPE,BIC and Q statistics are calculated as below;

$$Root\ Mean\ Square\ Error = \sqrt{\frac{\sum (Y_t - \widehat{Y}_t)^2}{n}}$$

$$Mean\ Absolute\ Percentage\ Error = \frac{1}{n}\sum_{t=1}^{n} \frac{|Y_t - \widehat{Y}_t|}{Y_t} \times 100$$

$$Mean\ Absolute\ Error = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \widehat{Y}_t|$$

where, $Y_t$is the original yield in different years and $\widehat{Y}_t$ is the predicted yield in the corresponding years and n is the no. of years utilised as predicting period.

Bayesian Information Criterion (p,q) = ln v*(p,q) + (p+q) [ln (n) / n]; where, p and q are the order of Auto Regressive and Moving Average processes respectively and n is the no. of observations in the time series and v* is the estimate of white noise variance σ^2.

$$Q = \frac{n(n+2)\sum_{i=1}^{k} rk^2}{(n-k)}$$

where, rk is the residuals autocorrelation at lag k and n is the number of residuals. The tentative ARIMA models differenced up to stationarity are discussed and the model which has the normalized BIC, stable R^2 and best model fitting statics (RMSE, MAPE and MAE) will be considered as the best fit.

RESULLTS AND DISCUSSION

The time series data for yield of groundnut in Bihar were collected from authenticated portals like Directorate of Groundnut Research, Directorate of Oilseeds Development and India Agri Stat. Data were collected and utilized from year 1980 to 2018. For forecasting of groundnut yield the data up to the year 2016 were utilised for developing the forecast model and remaining two years data were kept for validation of forecast model.

Yield in Bihar

After the study of figure- 4.1, it is noticed that the trend value of yield in 1980 is 893.77 kg/ha which is gradually increases and goes up to 939.81 kg/ha in 2018. Maximum actual yield in 2000 is 2000 kg/ha, whereas minimum actual yield in the consequent year 2001 is 62.5 kg/ha. The overall trend is stable and seems to be almost linear. Graph for residual v/s fitted values were given in figure -4.2.

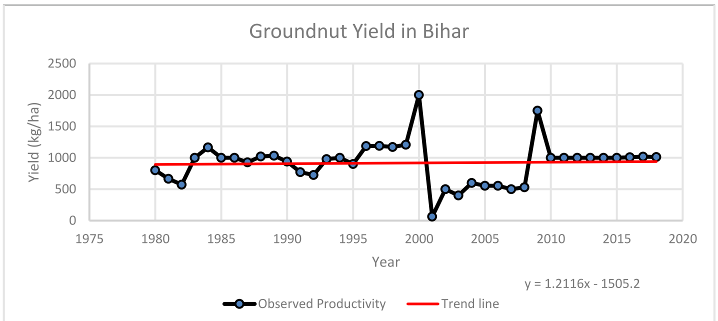

Fig. 4.1: Graph of trend value of groundnut yield in Bihar

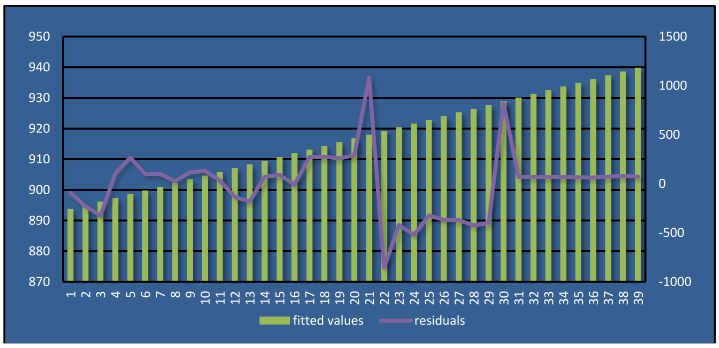

Fig. 4.2: Graph of residual v/s fitted value for groundnut yield in Bihar

Forecasting the yield of groundnut in Bihar through ARIMA models

In this study, we have used different models to find out the appropriate model for the forecast of groundnut yield in Bihar. For the model comparison, the yearly yield of groundnut was taken into consideration. The detailed analysis of groundnut yield forecasting in Bihar has been presented below.

Model Identification

The critical initiative stage in the finding process of an ARIMA is judging the stationary behavior of the under lying process. The original plot of groundnut yield data as shown in figure -4.3 reveals that the data is non stationary i.e. do not have constant mean and constant variance. So that the first order difference is plotted and shown in figure- 4.4. This plot revealed that the first order difference of data was found to be stationary i.e. having constant mean and constant variance.

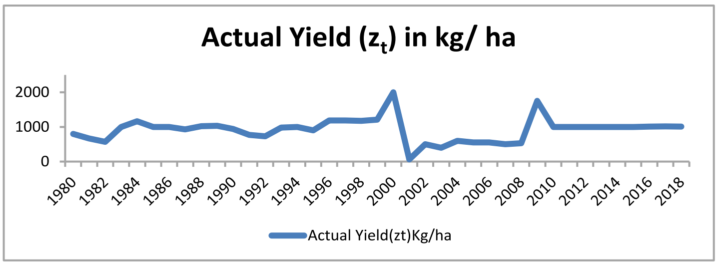

Fig. 4.3: Time series plot of groundnut yield in Bihar

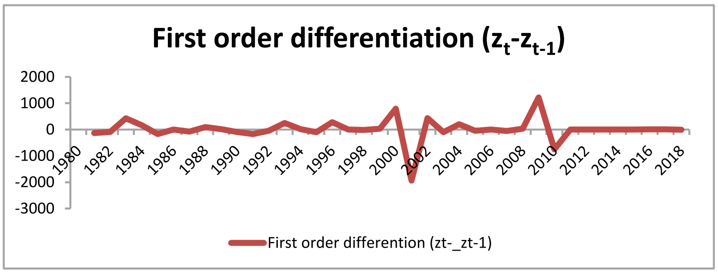

Fig. 4.4: Time series plot of first order difference series of groundnut yield in Bihar

The study of table- 4.1 and table - 4.2 show, Auto Correlation Function and Partial Auto Correlation Function of original (Z_t) and first order difference series(Z_t) for Bihar up to lag 20 (Twenty), respectively.

Table 4.1: ACF of original (Z_t) and first order difference series (ΔZ_t) for Bihar

Lag	Z_t	t-value	ΔZ_t	Δt –value
1	0.13	0.793	-0.046	-0.261
2	0.176	1.054	0.147	0.885
3	0.029	0.169	0.006	0.035
4	0.023	0.134	0.055	0.32
5	-0.032	-0.186	-0.005	-0.029
6	-0.027	-0.157	0.076	0.438
7	-0.103	-0.595	0.275	1.577
8	-0.378	-2.172	-0.316	-1.854
9	-0.062	-0.318	-0.026	-0.134
10	-0.036	-0.184	0.018	0.092
11	-0.054	-0.276	0.011	0.056
12	-0.065	-0.332	-0.009	-0.047
13	-0.056	-0.284	-0.01	-0.051
14	-0.046	-0.234	0.006	0.029
15	-0.052	-0.263	-0.054	-0.273
16	0.002	0.01	0.106	0.535
17	-0.104	-0.525	0.019	0.093
18	-0.123	-0.618	-0.15	-0.752
19	0.027	0.134	-0.015	-0.074
20	0.042	0.208	0.042	0.208

Table 4.2: PACF of original (Zt) and first order difference series (∆ Zt) for Bihar

Lag	Z_t	t-value	ΔZ_t	∆t–Value
1	0.13	0.793	-0.032	-0.195
2	0.162	0.988	0.173	1.055
3	-0.011	-0.067	-0.003	-0.018
4	-0.008	-0.049	0.03	0.183
5	-0.038	-0.232	-0.016	-0.098
6	-0.022	-0.134	0.068	0.415
7	-0.09	-0.549	0.277	1.689
8	-0.367	-2.238	-0.413	-2.518
9	0.046	0.28	-0.06	-0.366
10	0.106	0.646	0.167	1.018
11	-0.061	-0.372	0.018	0.11
12	-0.079	-0.482	-0.024	-0.146
13	-0.055	-0.335	-0.04	-0.244
14	-0.015	-0.091	0.086	0.524
15	-0.101	-0.616	0.051	0.311
16	-0.152	-0.927	-0.06	-0.366
17	-0.092	-0.561	-0.006	-0.037
18	-0.086	-0.524	-0.124	-0.756
19	0.038	0.232	0.027	0.165
20	0.011	0.067	0.011	0.067

Parameter Estimation

The estimates of different parameters of groundnut yield model of Bihar are given in output tables- 4.3.For this model the parameter estimates along with standard deviation and t-ratio have been computed. The output table contains the AR factor and/or MA factor along with respective forecast model. The autocorrelation of the residuals has been computed for the diagnostic checking of the model. L-Jung Box test is used as a measure of Q-statistics for testing the significance of residual autocorrelations. The output table 4.3 shows the fitting of ARIMA (1, 0, 1) for Bihar. For the forecasting of groundnut yield in Bihar, the best-fit model is ARIMA (1,0,1) as it has low values of RMSE, MAPE, MAE and BIC. Fig-4.5 and Fig.-5.6. Residual ACF and PACF plot for ARIMA (1,0,1) for the yield of groundnut and Graph between time (year) and yield for ARIMA (1,0,1) in Bihar.

Selection of good model

As discussed the earlier, an appropriate model should be parsimonious, stationary and invertible. Parameters of coefficients estimated have been of good quality and stable. Out of eight ARIMA models in this study all models satisfies the stationarity and invertibility condition. Although for ARIMA(0,1,1),ARIMA(0,1,2) and ARIMA(0,0,1) ,p=0 it means it is pure MA model or a white noise –series. Literature says that for all pure MA models white-noise is stationary, so there are no needs to check stationarity condition (Box and Jenkins 1976).Similarly, we have q=0 a pure number in the case of ARIMA(1,0,0) and ARIMA(2,0,0).If q=0, we have a pure AR process or white-noise series. Literature again says that all pure AR process(or white noise) are invertible and no further checks are required(Box and Jenkins 1976). For the selection of parsimonious model, the principle guidelines were suggested by Box and Jenkins. ARIMA (0,1,1), ARIMA (0,0,1), ARIMA (1,0,0), ARIMA (1,0,1) and ARIMA (2,0,0) satisfy parsimonious condition, because it has lower parameters(i.e. two parameters) in comparison to ARIMA (0,1,2), ARIMA(1,1,1) and ARIMA (2,0,1) i.e. having three parameters, which are not satisfying parsimonious condition. The t-test and chi square test for all the models have been computed. Based on estimated coefficients significance ARIMA (0,1,1), ARIMA (0,0,1), ARIMA (1,0,0), ARIMA (1,0,1), and ARIMA (2,0,0) result as parsimonious models.

A comparison of various models suggests ARIMA (1,0,1) as the appropriate model. A cross validation of the selected best fit ARIMA (1,0,1) for forecasting the groundnut yield in Bihar in table -4.4, shows that the RMSE, MAPE, MAE and BIC values are quite low. Thus, the selected model is successfully validated.

Table 4.3 Output of fitting ARIMA (1, 0, 1) for groundnut yield of Bihar

Model parameters

Parameters	Estimates	SE	t- Value
Constants	6.879	0.120	57.288
φ_1	0.855	0.133	6.418**
θ_1	0.354	0.165	2.141

Auto Regressive Factor: - (B) = 1 - 0.855B; Moving Average Factor: -(B) = 1 - 0.354B; Forecast Model: - $Z_t - Z_{t-1} = 6.879 + 0.855 (z_{t-1} - z_{t-2}) - 0.354 (a_{t-1} - a_{t-2}) + a_t$

Diagnostic Check:

Lags	Residual ACF	t- Value
1	0.106	0.646
2	-0.133	-0.801
3	-0.179	-1.059
4	-0.043	-0.247
5	0.022	0.126
6	-0.074	-0.423
7	-0.205	-1.171
8	-0.104	-0.571
9	-0.135	-0.738
10	0.059	0.317
11	0.221	1.188
12	0.043	0.223
13	-0.128	-0.660
14	0.003	0.015
15	0.044	0.22
16	0.062	0.316
17	-0.022	-0.112
18	0.030	0.152
19	-0.082	-0.416
20	0.074	0.374

Table: 4.4: Model Fit Parameter:

R-squared	RMSE	MAPE	MAE	BIC
0.909	120.338	9.868	81.519	10.459

Q-Statistics (L-Jung Box Test) = 10.879; D.F. = 16

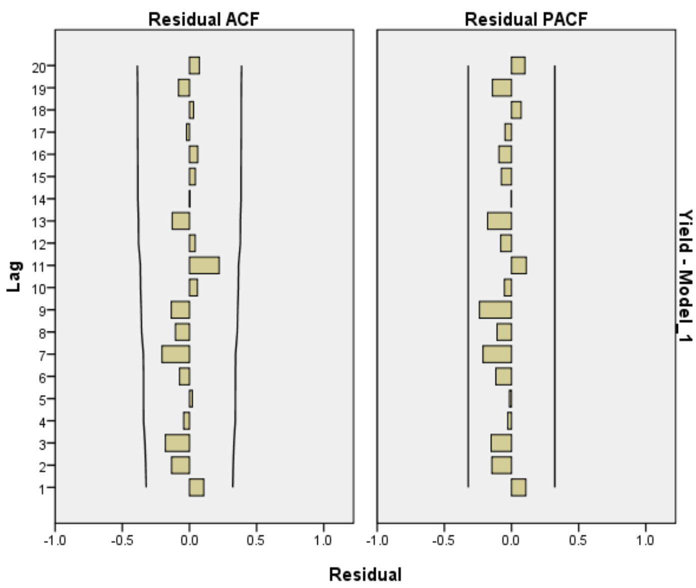

Fig. 4.5: Residual ACF and PACF plot for ARIMA (1,0,1) for yield of groundnut

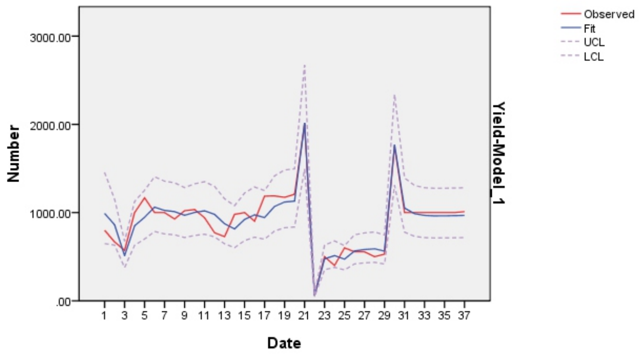

Fig. 4.6: Graph between time (year) and yield for ARIMA (1,0,1) in Bihar

Diagnostic check

To check whether the models are statistically adequate, the study of diagnostic checks have been performed. The diagnostic checks are in terms of independence of random shocks.

A statistically accurate model will have non-auto correlated random shocks with constant mean and constant variance. The residual ACF were calculated and given in the above-mentioned result tables for this study. The t-test was performed to test the significance of the null hypothesis Ho: $p_k(a) = 0$ for each residual autocorrelation coefficient. The respective standard error has been computed using Bartlette's approximation formulae. All the t-values of the residual ACF for the model under selection are non-significant, i.e., less than the critical values suggest the independence of the random shocks. This is also confirmed by a chi-square test given by Ljung and Box with the use of Q-statistics. The Q-statistics of L-jung and Box test for the appropriate model ARIMA (1,0,1) is 10.879, which is less than chi-square value at 16 degrees of freedom (table 4.4).Thus, the selected ARIMA model of the order (1,0,1) is found a perfect fit. The forecast error for the one step ahead and two steps ahead has been computed as 0.90% and 1.69% respectively (table 4.5).

Table 4.5: Selection of good model for groundnut yield of Bihar

Models	RMSE	MAPE	MAE	BIC	Stationary	Invertible	Parsimony	Stable/ R^2	Diagnostic Check		% Forecast Error	
									t-test	χ^2 test	1st step (2017)	2nd step (2018)
0,1,1	311.113	18.749	172.557	11.878	NA	✓	✓	0.301	NS	NS	0.48	2.68
0,1,2	184.562	12.074	109.840	11.232	NA	✓	X	0.785	** (For θ_1)	NS	0.22	3.44
0,0,1	142.768	12.547	100.429	10.606	NA	✓	✓	0.862	**	NS	6.30	5.99
1,0,0	144.723	13.215	106.085	10.633	✓	NA	✓	0.859	**	NS	4.82	5.02
1,0,1	120.338	9.868	81.519	10.459	✓	✓	✓	0.909	** (For φ_1)	NS	**0.90**	**1.69**
1,1,1	184.571	12.041	109.857	11.232	✓	✓	X	0.785	** (For θ_1)	NS	0.38	3.56
2,0,0	120.969	10.458	86.026	10.469	✓	NA	✓	0.908	** (For φ_1)	NS	5.13	5.64
2,0,1	122.586	9.860	81.464	10.594	✓	✓	X	0.909	NS	NS	4.25	3.10

*✓: Indicates That Condition is satisfied; X: Indicates That Condition is not satisfied; NA: Not applicable, if p=0 and q=0, i.e. for all pure AR and MA process(or white noise) are stationary and invertible respectively and no further checks are required; NS: Non-Significant; **: Highly significant*

Forecasting the yield of groundnut in Bihar through ARIMA models

Computation of forecasts and their confidence intervals

After confirming the validity of the model, it has been utilized to predict the future yields of the observed time series. Forecasts would be arrived in few years, which is called lead time. Thus, using ARIMA (1,0,1) the forecasts along the confidence intervals have been computed five years period ahead and presented in table -4.6. The table shows an increasing tendency of groundnut yield in Bihar for 2017 and 2018.

Table 4.6: Forecast along with their confidence interval of ARIMA (1,0,1) for groundnut yield of Bihar

Periods	Forecast	95% limits		Actual	% forecast error
		Lower	Upper		
2017	1026.72	285.26	2715.78	1017.54	0.90
2018	1028.74	285.82	2721.13	1011.63	1.69
2019	913.00	245.026	1580.97		
2020	913.00	245.026	1580.97		
2021	913.00	245.026	1580.97		

CONCLUSIONS

Groundnut is an important oilseed crop in India. Still the date, groundnut oil holds an important place in most of the Indian kitchens. So, it is important to study the trend of its area, production and yield. In this concern, ARIMA forecasting techniques were applied for the yield forecast. The linear trend analysis was done for the time series data of groundnut yield in Bihar . The trend line of groundnut yield in Bihar seems to be stable and linear. For the measure of accuracy, coefficient of determination (R^2) values are calculated for each one of the graphs.

The residuals examination was also done by fitting the graph of fitted versus residual values for each one of them. Instead of conventional or econometric methods, the ARIMA models were used to forecast the yield of groundnut in Bihar. The time series data of 37 years from 1980 to 2016 were used for the study. As the first step, the yield data were made stationary by the first order differentiation. Models ARIMA (0,1,1), ARIMA (0,1,2), ARIMA (0,0,1), ARIMA (1,0,0), ARIMA (1,0,1), ARIMA (1,1,1), ARIMA (2,0,0) and ARIMA (2,0,1) were built. All the 8 models were subjected to critical examination. Among them ARIMA (1,0,1) model was chosen as it is stationary, invertible, parsimonious, stable and has minimum error. Thus, the forecast model for groundnut yield in Bihar is,

$$\mathbf{Z_t - Z_{t-1} = 6.879 + 0.855\,(z_{t-1} - z_{t-2}) - 0.354\,(a_{t-1} - a_{t-2}) + a_t}$$

Using the selected ARIMA models ARIMA (1,0,1) and ARIMA (1,0,0) the yield values were forecasted for five years period ahead i.e. from 2017 to 2021 in Bihar and Tamil Nadu, respectively. The forecasted values of Bihar are 1026.72 kg/ ha, 1028.74 kg/ ha, 913.00 kg/ ha, 913.00 kg/ ha and 913.00 kg/ ha, respectively for 2017, 2018, 2019, 2020 and 2021. The forecasted values of Bihar exhibit an increasing trend, for 2017 and 2018, in the yield of groundnut. These yield values were presented along with their lower and upper limits with 95% confidence interval. Using the mathematically sound ARIMA models, the groundnut yield values were forecasted with 0.90 per cent of one step ahead forecast errors for Bihar. The two steps ahead forecast errors are 1.69 per cent for Bihar. The groundnut yield values were forecasted for a time period of five more years from 2017 to 2021 for Bihar.

The years 2017 and 2018 were kept for data validation. From the forecasted values of Bihar, it was found that the yield will remain more or less constant over the years.

Future scope of study

The current study has effectively demonstrated the utility of ARIMA models in forecasting groundnut yield using historical data. However, there remains a broad scope for expanding this research. Future studies can explore hybrid forecasting models by integrating ARIMA with advanced machine learning techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), or Long Short-Term Memory (LSTM) networks to improve predictive accuracy, especially for non-linear and seasonal agricultural data. Additionally, spatial forecasting using geostatistical or GIS-integrated time-series models can offer region-specific yield predictions in Bihar, allowing better targeting of policy and agronomic interventions. Climate variables such as rainfall, temperature, and relative humidity could be incorporated into multivariate time series models to account for climate-induced yield variability. There is also a need to validate and recalibrate the models with post-2021 data to ensure continued reliability. Moreover, similar modeling approaches can be extended to other major oilseed crops across different agro-climatic zones to create a comprehensive yield forecasting system. Lastly, operationalizing real-time forecasting platforms linked with remote sensing and weather forecasting data could help in developing a robust decision-support system for stakeholders involved in groundnut production, procurement, and planning.

Acknowledgment

The authors express their gratitude to the Dean, College of Basic Sciences and Humanities (CBS&H), Rajendra Prasad Central Agricultural University, Pusa, for approving the research work. We are also deeply appreciative of the technical staff of the Department, whose invaluable assistance greatly contributed to the success of this study. Their unwavering support and commitment were integral to the completion of this research.

Conflict of interest: The authors declare that they have no conflict of interest.

REFERENCES

1. Biswas, R., KumarMahto, A., &Alam, M. A. (2019). Short - term Forecasting of Agriculture Commodity Price by Using ARIMA: Based on Indian Market. In *International Conference on Advances in Computing and Data Sciences* (pp. 452-461). Springer, Singapore.

2. Kumar, Ajeet., Meena, S.K., Sinha, S.K., Singh, A.K., Minnatullah and Singh, S. K. (2024) Isolation and biochemical characterization of endophytic bacterium *Gluconacetobacter diazotrophocus* from native sugarcane cultivar of middle gangetic plains of India. *Indian Journal of ecology*, 51(1): 104-112. https://doi.org/10.55362/IJE/2024/4202.

3. Celik, S., Karadas, K., & Eyduran, E. (2017). Forecasting the production of Groundnut in Turkey Using ARIMA Model. *World*, *17373490*(150819), 25.

4. Debnath, M. K., Bera, K., & Mishra, P. (2013). Forecasting Area, Production and Yield of cotton in India using ARIMA model. *Research & Reviews: Journal of Space Science & Technology*, *1*(2), 16-20.

5. Jadhav, V., Chinnappa, R. B., &Gaddi, G. M. (2017). Application of ARIMA Model for forecasting agricultural prices.*Journal of Agricultural Science and Technology, 19*(4), 981-992.

6. Kumar, A., & Verma, U. (2020). Forecasting Mustard Yield in Haryana with ARIMA Model.*The Pharma Innovation*, *9*(4), 136-140.

7. Kumar A., Sinha S.K., Singh S.K., Rana L., Singh A.K., Kumari S., Kumar A., Singh H. and Paswan S. (2025) Influence of Intercropping and Planting Techniques on Sugarcane Yield and Nutrient Absorption in the North West Alluvial Plains of Bihar. *AATCC Review*, 13 (1): 144-153. https://hal.science/hal-04962582.

8. Kumar, M., Raman, R. K., and Kumar, S. (2017). Sugarcane Yield in Bihar: A Forecast through ARIMA Model. *Int. J. Pure App. Biosci*, *5*(6), 1042-1051.

9. Liu, C., Hoi, S. C., Zhao, P., &Sun, J. (2016). Online ARIMA Algorithms for Time Series Prediction. *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 30, No. 1).

10. Minnatullah, M., Kumar, A., Viswanathan, R., Meena, S.K., Paswan, S., & Kumar, P (2025). Enhancing chemical delivery efficiency through sett treatment device for management of red rot in sugarcane (Saccharum officinarum). *The Indian Journal of Agricultural Sciences*, *95*(4), 395–399. https://doi.org/10.56093/ijas.v95i4.157032.

11. Nochai, R., & Nochai, T. (2006). ARIMA Model for Forecasting Oil Palm Price. In *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and applications* (pp. 13-15).

12. Ozaki, T. (1977). On the order determination of ARIMA models. *Journalofthe Royal Statistical Society: Series C (Applied Statistics)*, *26*(3), 290-301.

13. Padhan, P. C. (2012). Application of ARIMA Model for Forecasting Agricultural Yield in India. *Journal of Agriculture and Social Sciences*, *8*(2), 50-56.

14. Pankratz, A. (2009). *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases* (Vol. 224). John Wiley & Sons.

15. Singh, S. K., Kantawa, M., Kumar, A., Kumari, S., Singh, K.K., Tiwari, S., Singh, P. and Singh, A.K. (2025). Groundwater Quality Assessment for Drinking and Agricultural Use in Motipur Block, Muzaffarpur District, Bihar, India. *International Journal of Environment and Climate Change*, 15 (5): 9-24. https://doi.org/10.9734/ijecc/2025/v15i54830.

16. Sharma, P. K., Dwivedi, S., Ali, L., & Arora, R. K. (2018). Forecasting MaizeProduction in India Using ARIMA Model. *Agro Economist-An International Journal*, *5*(1), 1-6.

17. Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1394-1401).

18. Suresh. K. K. and Krishna Priya, S. R. (2011). Forecasting sugarcane yield of Tamil Nadu using ARIMA models.*Sugar Tech*. 13(1):23-26.

19. Suvarna, M., Rachana, W., & Sonal, G. (2009). ARIMA Model for Forecasting Soybean Production in India. *Annals of Plant Physiology*, *23*(2), 246-248.

20. Kumar, Ajeet., Jha, C.K. and Kumari S. (2022) Study on feasibility of growing oil palm (*Elaeis guineensis*) under alluvial plain of Bihar. *Environment and ecology. 40 (2): 238-245, April –June 2022.* https://hal.science/hal-04676953.

21. Singh, S.P., Nema, S., Kumar, Ajeet., Minnatullah, and Kumar, N. (2024). Resource Use and Technical Efficiency of Soybean Production in Madhya Pradesh: A Data Envelopment Analysis Approach. *Journal of Advances in Biology & Biotechnology*, *27*(5), 551 - 557. https://doi.org/10.9734/jabb/2024/v27i5817.

22. Kumar A., Bairwa M., Meena S. K., Singh S. K., Sinha S.K., Rana L., Singh Y.V., Kumar A., Singh H. and Paswan S. (2025) Comparative Analysis of Macronutrient, Soil Sulphur Availability and Micronutrient Dynamics across Diverse Land Use Types of Indo-Gangetic plains of India. AATCC Review. 13 (01): 467-472. https://hal.science/hal-05081536.

23. Kumar, A and Kumar M. (2017) Effect of different levels of phosphorus on the yield and yield components of maize under agro-climatic zone- II of Bihar. *International Journal of Agricultural Sciences.* 13 (2): 266-270. http://researchjournal.co.in/online/IJAS/IJAS%2013(2)/13_266-270_A.pdf https://hal.science/hal-04676729.

24. Singh, S.K. Kumar, A and Singh A.K. (2019) Technical Bulletin On *Integrated Nutrient Management in Mustard Crop*, Pages 1-31. vide Technical Bulletin-INM in Mustard crop/SS-1/TCA-Dholi/2019, https://hal.science/hal-04960904.

25. Kumar, A., Singh, S. K., Meena, S. K., Sinha, S., Rana, L., Kumari, A., Singh, A.K., & Kumar, A. (2024). Human Health Risk Assessment of Nitrate-Contaminated Groundwater in North East Alluvial Plains of India. In Book: Geography, Earth Science and Environment: Print ISBN: 978-93-48388-08-7, eBook ISBN: 978-93-48388-35-3. *Geography, Earth Science and Environment: Research Highlights Vol. 2*, 1–26. https://doi.org/10.9734/bpi/geserh/v2/1424.

26. Kumar, A., Choudhary A., and Kumar M (2018) Impact of climate change and their mitigation for better sugarcane production-A Review; *International Journal of Agricultural Sciences.* 14 (2): 431-441. https://www.cabidigitallibrary.org/doi/pdf/10.5555/20183280336.