International Journal of Statistics and Applied Mathematics 2021; 6(1): 51-58

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2021; 6(1): 51-58

www.mathsjournal.com
Received: 14-11-2020
Accepted: 26-12-2020

MO Wankhade
Department of Statistics, G.S. Science, Arts & Commerce College, Khamgaon, Maharashtra, India

UV Kale
Department of Statistics, Shri. Shivaji Science College, Amravati, Maharashtra, India

Corresponding Author:
MO Wankhade
Department of Statistics, G.S. Science, Arts & Commerce College, Khamgaon, Maharashtra, India

Application of ARIMA model for forecasts analysis of yield and area of peanut in India

MO Wankhade and UV Kale

DOI: https://doi.org/10.22271/maths.2021.v6.i1a.635

Abstract
The main objective of this study is to generate and analyze forecasts of area under Peanut and yield of peanut (in quintal/hectare) in India. The data related to area under the crop and yield per hectare is extracted from Government of India, Directorate of Economics and Statistics from the year 19701-71 to 2017-18. An Autoregressive Integrated Moving Average (ARIMA) model is used for analysis of the data in MINITAB19 statistical software. Randomness of error terms in the model are tested by using Ljung-Box Chi-Square Statistic and Autocorrelation function and Partial Autocorrelation functions. Forecasts related to yield and area of Peanut along with 95% confidence intervals for the period of next 15 years i.e. from 2018-19 to 2032-33 are generated by fitting an ARIMA model. We observed that even though there is downward trend for area under Peanut, yield quintal per hectare showed upward trend.

Keywords: ARIMA, ljung-box chi-square statistic, ACF, PACF, forecast, peanut, India

Introduction
India is most prominent importer of edible and vegetable oil, in which Palm oil and hydrogenated palm oil has the major share. India import about 160 lakh tons of edible oil, in which Palm oil contribute about 85-95 lakh tones. Indian continental is rich in flora and fauna, fertile and productive land in belt of Ganga, Bramhaputra, Godawari, Kaveri, Krushna, Narmada etc valley. In India mainly nine edible oilseed crops are grown. India need nearly Rs. 80000 crore forex for import of edible oil per year. Soybean, mustard and Rapeseed, Peanut and Sunflower contribute more in term of area under the crop and yield. After 1980's soybean became a major vegetable oilseed followed by mustard and peanut. India was the prominent producer of Peanut till 1990's now it is second largest producer of peanut after China. Peanut is cultivated in more than 100 countries and China is the most prominent producer of peanut while India is next to China in terms of production as well as average yield. India contributes about 21% of world production of peanut but average yield per hectare is significantly low as compared to United States and China. United States has recorded highest yield 4.1 tons per hectare, China 3.6 tons per hectare and India has recorded only 1.8 tons per hectare.

Review of literature
Agashe DR *et al.* [1] revealed on linear regression model for the analysis of trends in area, production and productivity of peanut in various districts of Chhattisgarh, India and noticed an increasing trends with respect to area under crop and production of peanut due to modern facilities and technologies of irrigation and high yield varieties of the crop. The analysis of trend in area under crops, production and productivity of non-food grains in India was made by Mukesh Kumar & Shallu Sehgal [2] adopting the compound growth rate model for peanut, sunflower, soya bean, rapeseed & mustard, coffee, cotton jute for the period of 1980-81 to 2012-13. They remarked that there is increasing trends in area and production of peanut in India. V. Balakrishnama Naidu, A. Siva Sankar, C. Leelavathi [3] presented rigorous study of trends in area and production of Peanut, Sesame, Sunflower oilseeds in the state of Andhra Pradesh, India and also in Chittoor district of the same state for the period 1996-97 to 2011-12. They assumed the compound growth rate model for the analysis of trend and observed that the area, production and yield of Peanut and sesame are decreasing during the study period i.e. from 1996-97 to 2011-12 in the state of Andhra Pradesh.

The work on the trend with respect to area, production and productivity of peanut was moved by G Shruthi *et al.* [5] on extracting the time series data from directorate of economics and statistics, India of the period 2000-01 to 2013-14. Their results are also based on compound growth rate model and showed that the area and production of peanut in India has a decreasing trend while the productivity was increasing trend over the years. Also Santosh Kumar *et al.* [6] focused on the problem of instability analysis with respect to area under crop, production and productivity of Soybean in India by applying compound growth rate model on the time series data for the period 1996-97 to 2015-16. They noted that growth rate of area, production and productivity of Soybean was increasing during the study period. They have examined instability with respect to area, production and productivity of Soybean on the basis of coefficient of variation and instability index based on coefficient of variation and R^2. A comparative study of trends of area, production and productivity of peanut among district level (Anantpur), State level (Andhra Pradesh) and national level (India) was conceived by Dr. B. Madhusudhana [7] by conducting a survey. He perceived that growth rate of area has declining trend in Anantpur district, Andhra Pradesh and at national level. Chandra Mohan Misra [8] examined the trends in area under cultivation, production and productivity of Peanut in India by applying the technique of orthogonal polynomial on the time series data and made some recommendations for resolving the challenges and issues related to peanut cultivation in India. A impact of cultivation technology, machinery, chemical control, sowing period, varieties of peanut, crop rotation, seed treatment, fertilizer management, irrigation management and inter cultural operations was explained by Ramapadmaja, T. Umamahaswara Rao [9].

Methodology

The time series data for the analysis is extracted from Directorate of Economics and Statistics, Ministry of Agriculture and farmer welfare, Government of India, from 19701-71 to 2017-18. The same data is analyzed for forecast and trend with respect area under peanut and yield of peanut quintal/hectare by assuming the Autoregressive Integrated Moving Average (ARIMA) in MINITAB19 software. The forecasts for next 15 years from 2017-18, with respect to area and yield of peanut are generated along with 95% confidence limits. The tests for independence of error terms are made using Ljung-Box Chi-Square Statistic and Autocorrelation function (ACF) and Partial Autocorrelation function (PACF).

An ARIMA (p, d, q) model is mathematically described as (1);

$$Y_t = \theta_0 + \varepsilon_t + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \quad \dots (1)$$

Where Y_t and ε_t are the observed value and random error at time the period t respectively.
Also ε_t assumed to be Normal distribution N (0, σ^2).
φ_i (i=1, 2, p) and θ_j (j=0,1, 2, q) are termed as the parameters of the model.

Where p is the order of autoregressive, d is the order of integrated and q the order of the moving average. The parameters p, d and q are assumed to be integers greater than or equal to zero.
A mathematical formulation of ARIMA (p, d, q) model using lag polynomials is as shown in (2).

$$\varphi(L)(1-L)^d y_t = \theta(L)\epsilon_t \quad \dots (2)$$

This is rewritten as (3).

$$\left[1 - \sum_{i=1}^{p} \varphi_i L^i\right](1-L)^d y_t = \left[1 + \sum_{j=1}^{q} \theta_j L^j\right]\epsilon_t \quad \dots (3)$$

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) [4] functions are expressed (4) to (6);
Autocorrelation Coefficient at k^{th} lag is.

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad \dots (4)$$

$$\gamma_k = \mathrm{Cov}\,(y_t, y_{t+k}) \quad \dots (5)$$

Covariance at lag zero is;

$$\gamma_0 = \mathrm{Cov}(y_t, y_t) = \mathrm{Var}(y_t) \quad \dots (6)$$

Ljung-Box-Pierce statistic Q is.

$$Q = n(n+2)\sum_{i=1}^{k} \frac{{\rho_i}^2}{n-i} \quad \dots (7)$$

Where ρ_i is the autocorrelation at i^{th} lag and k is the number of lags to be tested.
Statistic Q follows chi-square distribution with k-p-q degrees of freedom.

Table 1: Area under peanut (million hectare) and yield (quintal/hectare) in India

Year	Area	Yield	Year	Area	Yield	Year	Area	Yield
1970-71	7.33	8.34	1986-87	6.98	8.41	2002-03	5.94	6.94
1971-72	7.51	8.23	1987-88	6.84	8.55	2003-04	5.99	13.57
1972-73	6.99	5.85	1988-89	8.53	11.32	2004-05	6.64	10.2
1973-74	7.02	8.45	1989-90	8.71	9.3	2005-06	6.74	11.87
1974-75	7.06	7.24	1990-91	8.31	9.04	2006-07	5.62	8.66
1975-76	7.22	9.35	1991-92	8.67	8.18	2007-08	6.29	14.59
1976-77	7.04	7.47	1992-93	8.17	10.49	2008-09	6.16	11.63

1977-78	7.03	8.66	1993-94	8.32	9.41	2009-10	5.48	9.91
1978-79	7.43	8.35	1994-95	7.85	10.27	2010-11	5.86	14.11
1979-80	7.17	8.05	1995-96	7.52	10.07	2011-12	5.26	13.23
1980-81	6.8	7.36	1996-97	7.6	11.38	2012-13	4.72	9.95
1981-82	7.43	9.72	1997-98	7.09	10.4	2013-14	5.51	17.64
1982-83	7.22	7.32	1998-99	7.4	12.14	2014-15	4.77	15.52
1983-84	7.54	9.4	1999-00	6.87	7.64	2015-16	4.6	14.65
1984-85	7.17	8.98	2000-01	6.56	9.77	2016-17	5.34	13.98
1985-86	7.12	7.19	2001-02	6.24	11.27	2017-18	4.91	18.68

Source: Directorate of Economics and Statistics, Government of India
Area in Million Hectare: Yield in Quintal per Hectare

Result and discussion

Setting the parameters values as p=1, d=1 and q=1 in ARIMA (p, d, q) model for area under peanut the results obtained using MINITAB19 software are presented in Table No 2.

ARIMA model (Area): ARIMA (1, 1, 1) Area under Peanut (in Million Hectare)

$$Y_t = -0.0513 - 0.004Y_{t-1} - 0.357 \epsilon_{t-1} \quad \dots (8)$$

Table 2: Estimates parameters at Each Iteration in ARIMA (1, 1, 1)

Iteration	SSE	Parameters		
0	11.8544	0.100	0.100	0.044
1	10.1825	-0.050	0.249	-0.077
2	10.0907	0.059	0.399	-0.054
3	10.0797	-0.024	0.336	-0.053
4	10.0790	0.008	0.368	-0.051
5	10.0788	-0.009	0.351	-0.052
6	10.0787	-0.001	0.359	-0.051
7	10.0787	-0.005	0.355	-0.051
8	10.0787	-0.003	0.357	-0.051
9	10.0787	-0.004	0.356	-0.051
10	10.0787	-0.004	0.357	-0.051
11	10.0787	-0.004	0.356	-0.051
12	10.0787	-0.004	0.357	-0.051
13	10.0787	-0.004	0.357	-0.051
14	10.0787	-0.004	0.357	-0.051

Table 3: Final Estimates of Parameters for area under peanut

Type	Coefficient	SE Coef	T-Value	p -Value
AR 1	-0.004	0.425	-0.01	0.993
MA 1	0.357	0.396	0.90	0.373
Constant	-0.0513	0.0450	-1.14	0.261

Table 4: Ljung-Box Chi-Square Statistic at different lags

Ljung-Box Chi-Square Statistic			
Lag	**12**	**24**	**36**
Chi-Square	7.35	16.20	19.94
DF	9	21	33
P-Value	0.600	0.758	0.964

It was is observed that the model converges in 14 iterations as there is no significant change in estimates of parameters after 14th iteration (Table No. 2). As the p-value for order of Auto regression AR1 is found to be 0.993 which indicates that order of auto regression is not significant at 5% level. Similarly an order of moving average MA1 is insignificant at 5% level of significance (Table No 3).

For testing randomness of error terms in the model we set up null hypothesis as

H_{01}: An error terms are random in nature in the ARIMA model

As the p-value of Ljung-Box Chi-Square Statistic for all lags is greater than 0.05 is an indication that error terms are random in nature (Table No. 4). Also all no correlation for ACF and PACF are significant (Figure 3 and Figure 4).

Forecast of Area under Peanut (Million Hectare) with 95% confidence Limits is shown in the following Table No. 5, Figure 1 and Figure 2. The forecasts estimates of area under peanut for the period of next 15 years i.e. from 2019 to 2033 exhibits that decreasing trend in the area of peanut in India. It will range from 4.90697 to 4.1915 million hectare with lower value 1.75386 and higher value of 6.62914 respectively.

Table 5: Forecast of area under peanut (Million Hectare) with 95% confidence limits

Year	Forecast	Lower	Upper	Year	Forecast	Lower	Upper	Year	Forecast	Lower	Upper
2019	4.906	3.969	5.845	2024	4.651	3.012	6.290	2029	4.396	2.275	6.516
2020	4.855	3.742	5.969	2025	4.600	2.854	6.346	2030	4.345	2.140	6.548
2021	4.804	3.538	6.070	2026	4.549	2.702	6.395	2031	4.294	2.009	6.578
2022	4.753	3.352	6.154	2027	4.498	2. 555	6.440	2032	4.242	1.880	6.604

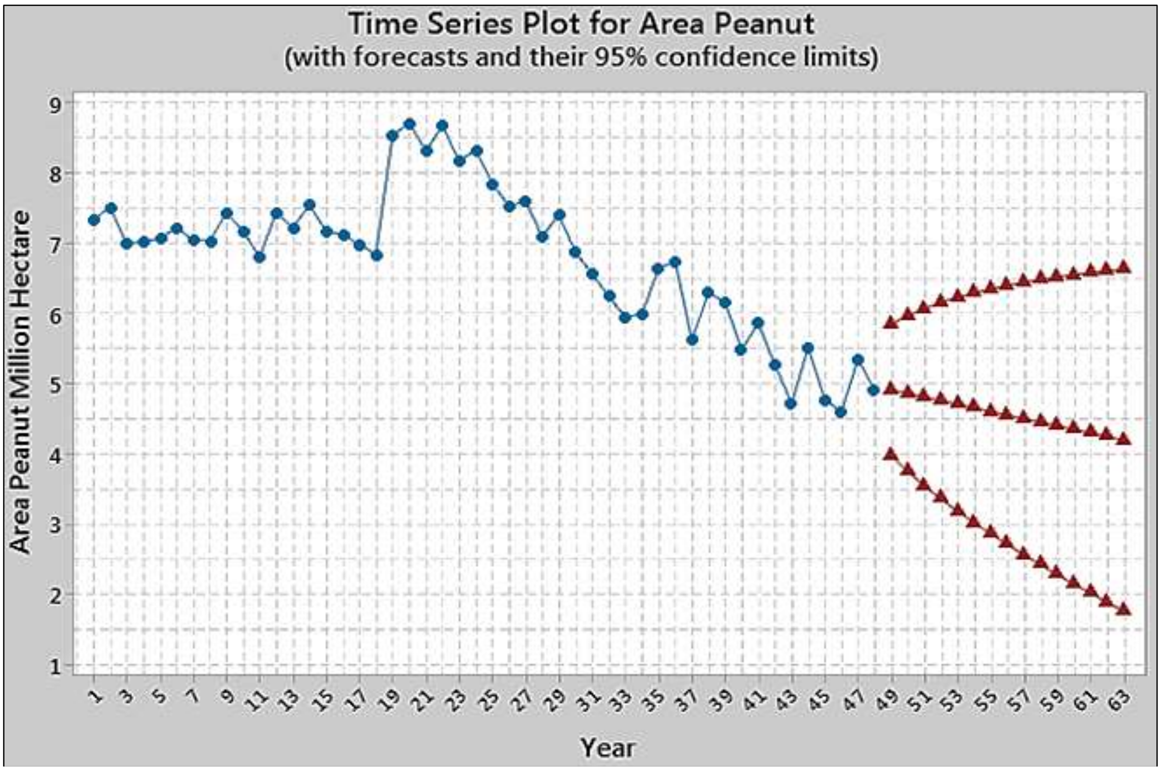

Fig 1: Time series plot of area under Peanut

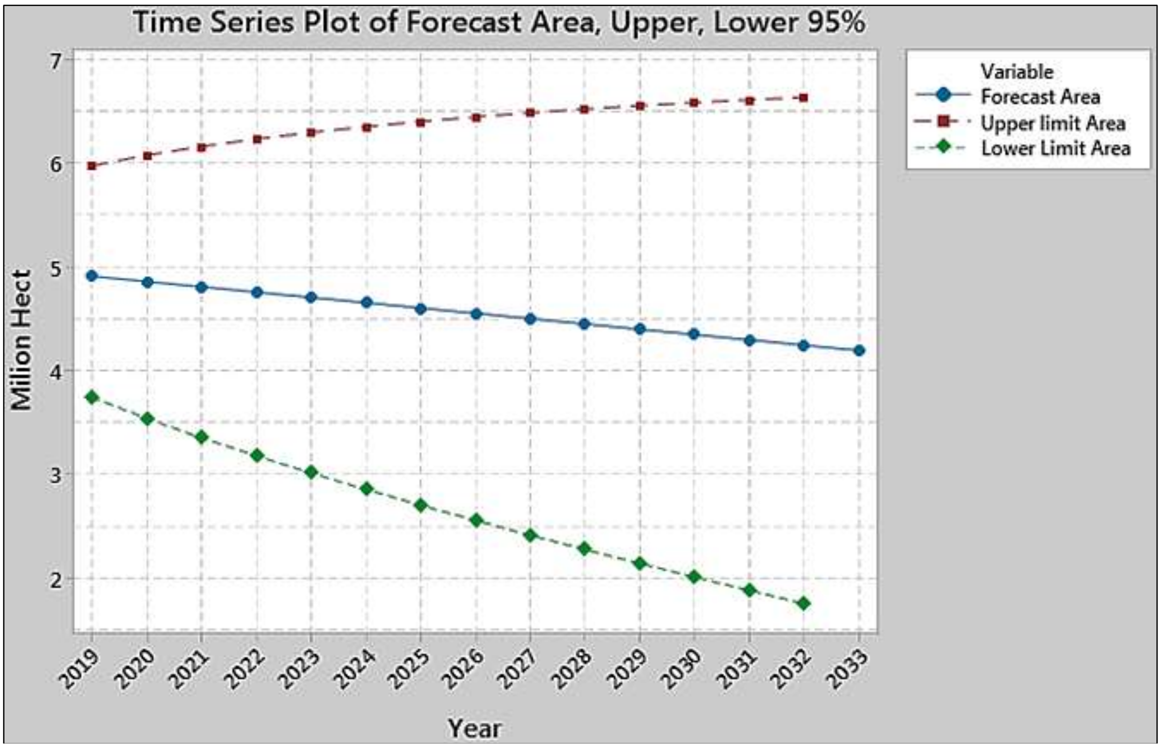

Fig 1: Time Series Plot of Forecast of Area under Peanut

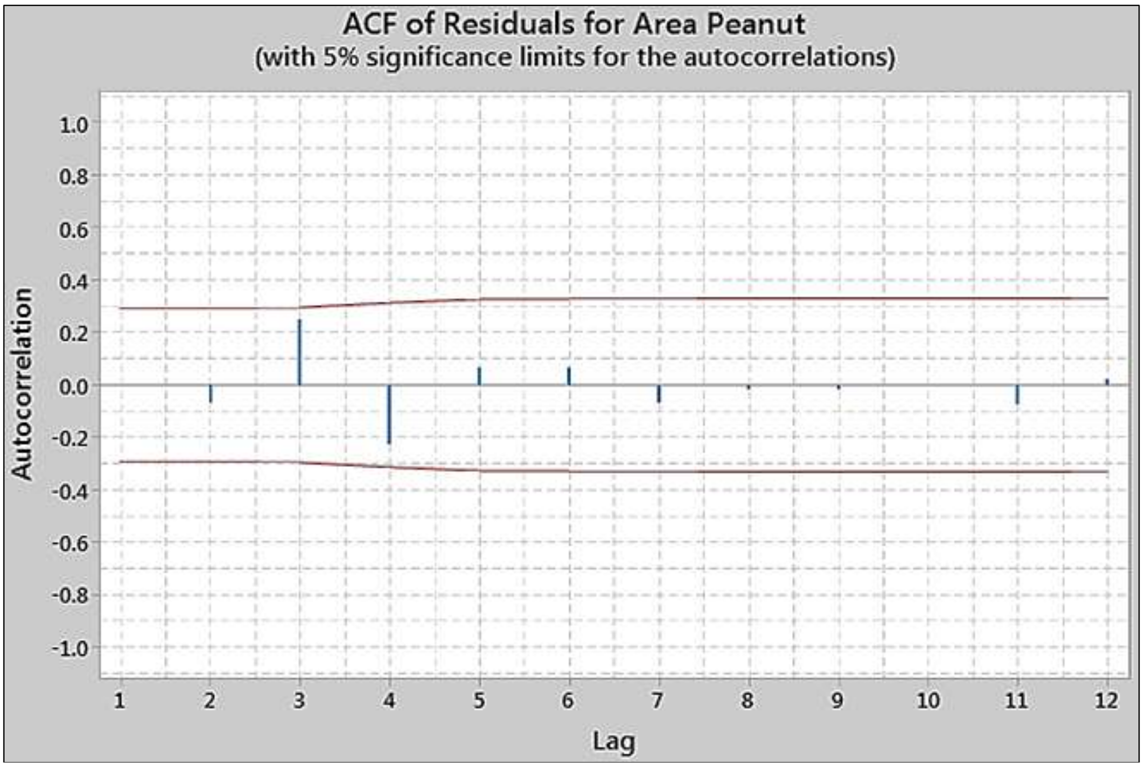

Fig 3: Graph of Autocorrelation function

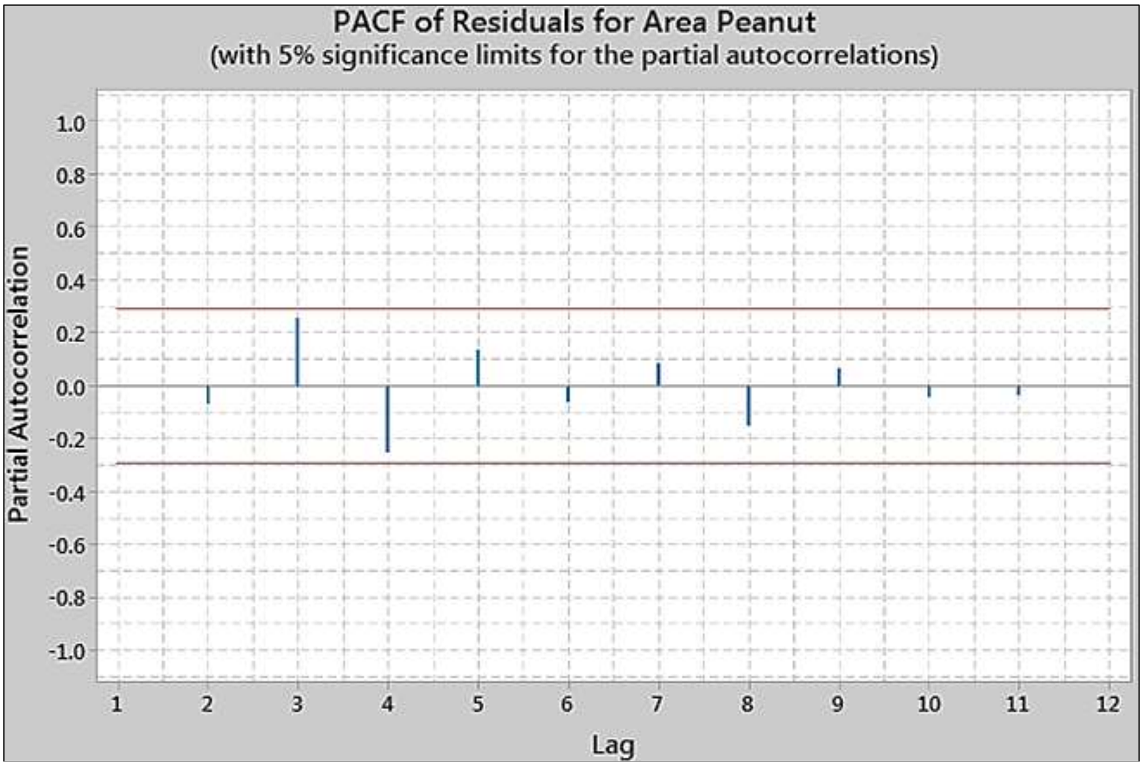

Fig 4: Graph of Partial Autocorrelation function

For generating the forecasts of yield of peanut with p=1, d=1 and q=1 in ARIMA (p, d, q) model it was observed that all the parameters were significant at 5%. Also p-value of Ljung-Box Chi-Square for 12 lags was less than 0.05. It was an indication that error terms are not independent. Thus we have changed the values of parameters p, d and q to fit a suitable ARIMA model for this data.

The parameters were set as p = 2, d =1 and q = 2, thus ARIMA model for yield forecasts is now expressed as;
ARIMA Model (Yield): ARIMA (2, 1, 2) Yield of Peanut (quintal/Hectare) obtained is as;

$$Y_t = 0.315 - 0.45Y_{t-1} - 0281Y_{t-2} + 0.772 \in_{t-1} - 0.21 \in_{t-2} \quad \dots (9)$$

Table 6: Estimates parameters at each iteration for yield of peanut

ARIMA model for yield of peanut (Quintal/ Hectare)						
Iteration	**SSE**	**Parameters**				
0	345.428	0.100	0.100	0.100	0.100	0.256
1	252.409	-0.050	0.025	0.250	0.175	0.195
2	243.357	0.065	-0.056	0.400	0.069	0.187
3	233.482	0.175	-0.133	0.550	-0.043	0.180
4	218.241	0.257	-0.207	0.700	-0.156	0.177
5	177.406	0.183	-0.271	0.850	-0.262	0.198
6	158.940	0.033	-0.253	0.828	-0.238	0.220
7	144.102	-0.117	-0.248	0.820	-0.232	0.246
8	135.309	-0.267	-0.250	0.810	-0.227	0.274
9	132.225	-0.409	-0.257	0.800	-0.224	0.303
10	132.185	-0.457	-0.281	0.763	-0.200	0.316
11	132.185	-0.435	-0.275	0.786	-0.222	0.312
12	132.181	-0.455	-0.283	0.766	-0.204	0.316
13	132.180	-0.446	-0.280	0.775	-0.213	0.315
14	132.180	-0.452	-0.282	0.770	-0.208	0.316
15	132.180	-0.448	-0.280	0.774	-0.212	0.315
16	132.180	-0.451	-0.282	0.771	-0.209	0.316
17	132.180	-0.449	-0.281	0.773	-0.211	0.315
18	132.180	-0.451	-0.281	0.771	-0.209	0.316
19	132.180	-0.449	-0.281	0.772	-0.211	0.315
20	132.180	-0.450	-0.281	0.772	-0.210	0.315
21	132.180	-0.450	-0.281	0.772	-0.210	0.315

Table 7: Final Estimates of Parameters for Yield of peanut

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.450	0.581	-0.77	0.443
AR 2	-0.281	0.299	-0.94	0.353
MA 1	0.772	0.575	1.34	0.187
MA 2	-0.210	0.455	-0.46	0.647
Constant	0.315	0.114	2.77	0.008

Table 8: Ljung-Box Chi-Square Statistic

Lag	12	24	36
Chi-Square	9.32	15.64	24.38
DF	7	19	31
P-Value	0.230	0.681	0.795

From the Table No. 6 there is no significant change in the estimates of the parameters after 21st iteration. Thus the model converges in 21 iterations. The p-values for order of Auto regression AR1, AR2 and moving average MA1, MA2 are greater than 0.05 which indicates that order of auto regression and moving average are not significant at 5% level of significance (Table-7).

H_{02}: An error terms are random in nature

The p-value of Ljung-Box Chi-Square Statistic for all lags is greater than 0.05 which indicates that error terms are independent (Table No.8). Also all no correlation for ACF and PACF are significant (Figure 7 and Figure 8).Forecast of yield Peanut (quintal/ Hectare) with 95% confidence Limits is shown in the following Table No. 9, and Figure 5 and Figure 6 shows trend of forecasts of productivity of peanut. The forecasts estimates of area under peanut for the period of next 15 years i.e. from 2019 to 2033 exhibits an increasing trend in the productivity (yield/Hectare) in India. It will range from 14.8914 to 18.6886 with lower 11.3524 and highest 23.5144 quintal/ hectare respectively.

Table 9: Forecast of yield of peanut (Quintal per Hectare) with 95% confidence Limits

Year	Forecast	Lower	Upper	Year	Forecast	Lower	Upper	Year	Forecast	Lower	Upper
2019	14.891	11.430	18.352	2024	17.114	13.079	21.150	2029	17.956	13.460	22.453
2020	16.112	12.567	19.657	2025	17.217	13.070	21.363	2030	18.142	13.562	22.723
2021	16.943	13.288	20.598	2026	17.399	13.166	21.632	2031	18.324	13.659	22.988
2022	16.542	12.652	20.432	2027	17.604	13.283	21.924	2032	18.505	13.760	23.251
2023	16.804	12.857	20.751	2028	17.776	13.365	22.187	2033	18.688	13.862	23.514

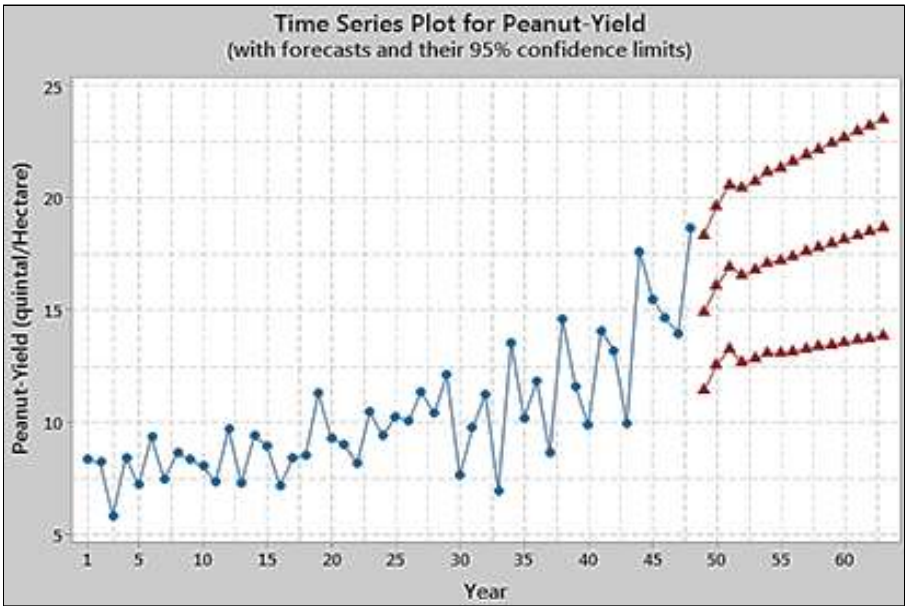

Fig 5: Time Series Plot of Yield

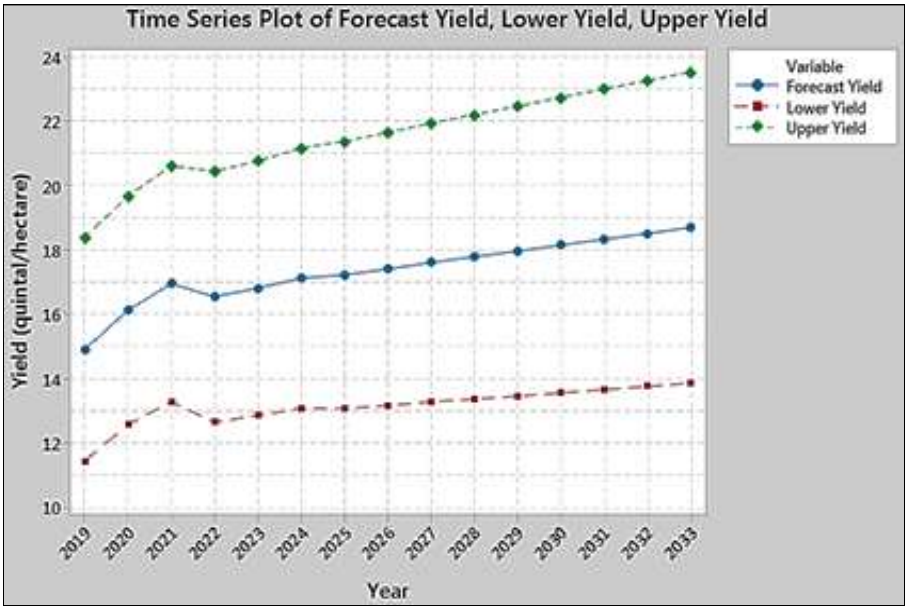

Fig 6: Time Series Plot of forecast Yield

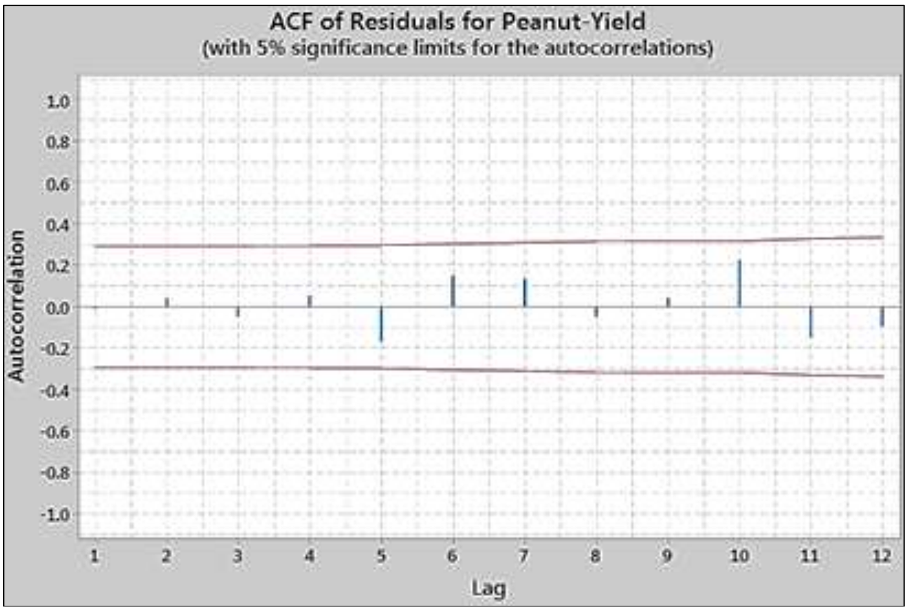

Fig 7: ACF for yield of peanut

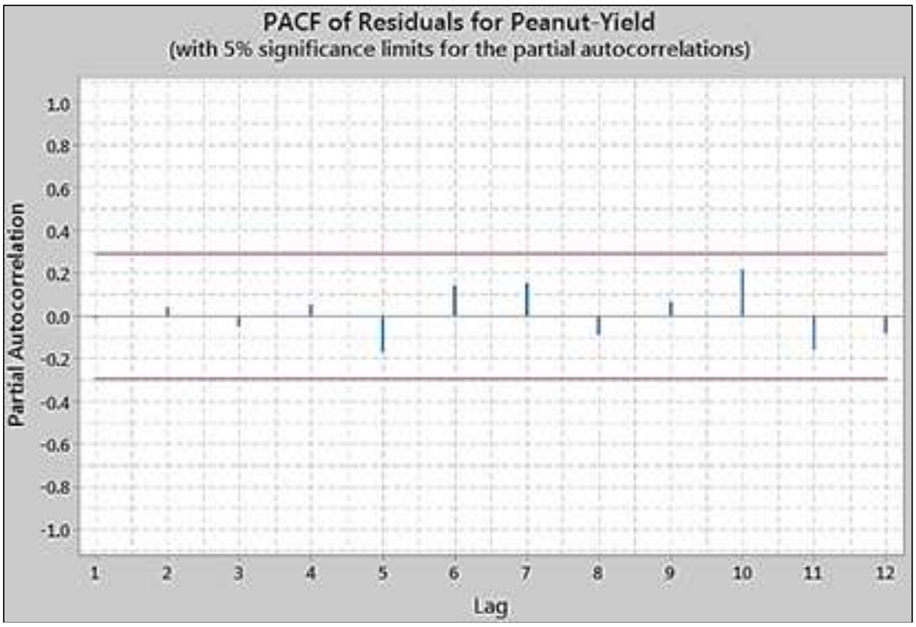

Fig 8: PACF for yield of peanut

Conclusion

We have adopted ARIMA model for analyzing the forecasts trend of area under peanut and productivity in quintal per hectare. ARIMA (1, 1, 1) model was proper model for estimating the forecasts of area under the crop peanut. All error terms were observed to independent in the said model. An ARIMA (1, 1, 1) model showed that there is downward trend with respect to area under peanut. While adopting ARIMA (1, 1, 1) model for analysis of forecasts with respect to yield (quintal/Hectare) of ground, showed insignificance of the model also error terms were observed to dependent. To find a suitable ARIMA model the values of p, d and q are changed and taken as p = 2, d = 1 and q = 2 resulting ARIMA (2, 1, 2) model. It was found that ARIMA (2, 1, 2) model showed insignificant results and also error term was observed to random in nature. Hence we may conclude that ARIMA (2, 1, 2) would be the proper model for analysis of forecasts of yield of peanut. Finally we conclude that ARIMA (1, 1, 1) model showed decreasing trend with respect to area under peanut from 2019 to 2033. While an ARIMA (2, 1, 2) model exhibits increasing trend with respect to productivity of peanut for the period 2019 to 2033. Area under peanut is decreasing while the yield of peanut in quintal per hectare is increasing. An increase in yield of peanut may be attributed to number of factors such as, quality of seed, irrigation facilities, fertilizers etc.

References

1. Agashe DR, Asras Sastri, Pandurang Bobade, Rajni Agashe. Trends of area, production and productivity of peanut in different districts of Chhattisgarh, Journal of Pharmacognosy and Phytochemistry 2018;7(4):1254-1259.
2. Mukesh Kumar, Shallu Sehgal. Trends in Area, Production and Productivity of Non-Food Grains in India, International Journal of Research 2015;2(2):993-1001.
3. Balakrishnama Naidu, Siva Sankar A, Leelavathi C. Trends in area, production and productivity of selected oil seed crops in Andhra Pradesh, International Journal of multidisciplinary research and development 2014;1(7):366-369.
4. Peter Zhang G. Time series forecasting using a hybrid ARIMA and neural network model Neuro computing 2003;50:159-175.
5. Shruthi G, Dayakar Rao B, Latika Devi Y, Jolly Masih. Analysis of Area, Production and Productivity of peanut crop in Telangana, Agric. Sci. Digest 2017; 37(2):151-153.
6. Santosh Kumar, Singh PK, Rathi D, Nahatkar SB, Choudhary VK, Parey SK. Growth and instability in area, production and productivity of soybean in India, International Journal of Science, Environment and Technology 2019;8(2):278-288.
7. Dr. B Madhusudhana. A Survey on area, production and productivity of peanut crop in India, IOSR Journal of Economics and Finance 2013;1(3):01-07.
8. Chandra Mohan Misra. Trends in area production and productivity of peanut in India: Issues & Challenges, Journal of research in agriculture and animal science. 2017;4(7):01-06.
9. Ramapadmaja, Umamahaswara Rao T. Impact of technology on peanut production of the Chittoor District, International Journal of recent technology and engineering 2019;8:1S4.
10. Gurudev A Tularam, Tareq Saeed. Oil price forecasting based on various univariate time series models, American Journal of operations research 2016;6:226-235.