The Pharma Innovation Journal 2021; SP-10(11): 529-533

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; SP-10(11): 529-533

www.thepharmajournal.com
Received: 07-09-2021
Accepted: 09-10-2021

Neelam Chouksey
RMD College of Agriculture & Research Station, Ambikapur Chhattisgarh, India

Corresponding Author
Neelam Chouksey
RMD College of Agriculture & Research Station, Ambikapur Chhattisgarh, India

Forecasting of yield and production of groundnut using ARIMAX

Neelam Chouksey

Abstract
Historical data has been considered for forecasting of groundnut yield and production. For the purpose, autoregressive integrated moving average with explanatory variables has been applied along with all estimation procedures. There are ARIMAX technique employed for forecasting of groundnut yield and production on time based data of Surguja district of Chhattisgarh. Data of Weather variables viz. maximum temperature, minimum temperature rainfall and Precipitation (1966 to 2020) is taken from Meteorological observatory of Rajmohini Devi College of, Agriculture and Research Station Ambikapur, Chhattisgarh as input variables in ARIMAX model. Comparative study of the fitted models is carried out from the viewpoint of mean absolute percentage error (MAPE), root mean squared error (RMSE). After comparison resulted that ARIMAX (1,1,1) and ARIMAX (1,1,0) model provided a lower RMSE from other models for yield and production respectively.

Keywords: forecasting, ARIMAX, groundnut, yield, production

Introduction
In all oilseed crops, groundnut is one of the major crop for most of the tropical and subtropical regions in the world. Groundnut distribution and its productivity may be affected by temperature and other climatic factors. In India it has first position in area and second position in production.
According to FAOSTAT (2018), China produces the largest quantity of groundnut 171.50 lakh tones as well as it is the biggest consumer in the world. After that other major groundnut producing countries are India (91.79 lakh tonnes), United States (32.81 lakh tonnes), Nigeria (24.20 lakh tonnes) and Sudan (16.41 lakh tonnes). The current showed area of India is 4.75 lakh hectares as compared to higher than last year 4.59 lakh hectares (directorate of economics and statistics, 2020-21). The major groundnut coverage states are Telangana with 1.16 lakh ha acreage followed by Karnataka (1.07 lakh ha), Tamil Nadu (0.99 lakh ha), Odisha (0.70 lakh ha) and Andhra Pradesh (0.66 lakh ha).Theses states occupies almost 70% of total area of India. However groundnut is sensitive crop and it cannot stand in frost, long and severe drought and water stagnation condition but it is grown in most of variety of soil types. This crop is best for loamy soils, sandy-loam and also in black soils with good drainage facilities .Heavy and stiff clays are unsuitable for groundnut cultivation.
However the contribution of groundnut is higher in India's GDP but need of protein in diet is also increased therefore demand of groundnut increased for increasing population in India. Groundnut production should be increased to meet the required demand. In other words if, demand is increased for consumption of human then decreased the availability of groundnut for industries, which creates problem for providing required capacity for agriculture-based industry (Pal, and Mazumdar, 2015) [12]. So that, accurate and reliable forecasts of crop yield before the harvest may be solve the problem of tropical interest. Such type of forecast is needed for the policy makers, Government, agro industries, agriculturists and traders. Many linear and nonlinear forecasting methods are available in our literature. In all of them, univariate time-series method Box Jenkins' Autoregressive integrated moving average (ARIMA) (Box *et al*., 2007) [3] method is very important and widely used method. Mishra *et al*. (2014) [11] adopted ARIMA model for the forecasting of fertilizer application Celik *et al.* (2017) [4] applied ARIMA technique for forecasting of ground production for next 15 years. They resulted that, ARIMA (0, 1, 1) was best fitted model for groundnut forecasting. Sanjeev, Verma, Urmil, (2015) [13] were used ARIMA model for pre harvest sugarcane yield estimation. They used weather variables in ARIMA model as input variables i.e. the ARIMAX model and resulted that ARIMAX model was performed superior over univariate ARIMA model.

Chouksey *et al*. (2018a) [6] employed ARIMAX models for forecasting of Rice yield for different nutrient combinations of Raipur district of Chhattisgarh. Chouksey *et al*. (2018b) [7] used combination of AR and MA order ARIMAX model for different nutrient combinations of nitrogen content and organic carbons included as an input variable. They found that ARIMAX model out performed as compared to ARIMA model. Hence many researchers stated that consideration of meteorological co-variables such as maximum temperature, minimum temperature rainfall and humidity etc. could be improve the prediction (Chadsuthi *et al*., 2012; Yogarajah *et al*. 2013; Arya *et al*. 2015) [5, 15, 1].

Data description and Statistical Methodology adopted

In the study, adopting modeling the time-series technique by using yield and production of Groundnut crop. Groundnut yield and production data form period 1966 to 2020 of Surguja district of Chhattisgarh is collected from the directorate of agriculture Chhattisgarh. Data of Weather variables viz. maximum temperature, minimum temperature rainfall and Precipitation (1966 to 2020) is taken from Meteorological observatory of Rajmohini Devi College of, Agriculture and Research Station Ambikapur, Chhattisgarh.

Autoregressive integrated moving average

ARIMA is linear time series model. In the model to achieve greater flexibility in fitting of actual time-series data, include both autoregressive and moving average processes. This leads to the mixed autoregressive-moving average model. Let us denote the values of a process at equally spaced time epochs

$t, t\text{-}1, t\text{-}2, \ldots$ by $y_t, y_{t-1}, y_{t-2}, \ldots$

"$y = j_1\, y_{t-1} + j_2\, y_{t-2} + j_3\, y_{t-3} + \ldots + j_p\, y_{t-p} + e_1 - q_1\, e_{t-1} - q_2\, e_{t-2} - q_3\, e_{t-3} + \ldots - qq\, e_{t-q}$ .....“(1)”

Where j is autoregressive coefficient; q is moving average coefficient and ε_t is white noise. This is written as ARMA (*p*, *q*) model. The models for nonstationary series are usually called by ARIMA models because are they use a combination of autoregressive (AR), integration (I) - referring to the differencing to produce the forecast, and moving average (MA) operations. An ARIMA model is usually stated as ARIMA (p,d,q) (Box & Jenkins; 1976) [3]. The“ARIMA methodology is carried out in four stages, viz. identification, estimation and diagnostic checking and forecasting. Parameters have estimated at the estimation stage of the tentatively selected ARIMA model at the identification stage and adequacy of tentatively selected model has been tested at the diagnostic checking stage. If the model has found to be inadequate, repeat the three stages until satisfactory ARIMA model is not select for the time-series under consideration.”

Autoregressive integrated moving average with exogenous variables - (ARIMAX)

ARIMA in one of the most important time series technique but if series is influenced by many other exogenous variable that case ARIMA not performed as our expectation. This condition necessary to create multivariate ARIMA model i.e. called ARIMAX

Let’s take two time series denoted Y_t and X_t, both the series have to be stationary. After that transfer function model is written as follows:

$Y_t =$ “$C + V(B)\, X_t + N_t$ … (2)

Where:

Y_t is the output series (yield);

X_t is the input series (nitrogen content and organic carbon);

C is constant term and

N_t is the stochastic disturbance, i.e. the noise series of the system that is independent of the input series; we have ARIMAX model formula with maximum lag denoted by *p* (free-form distributed lag model):

$y = C + v_0\, X_t + v_1\, X_{t-1} + v_2\, X_{t-2} + v_3\, X_{t-3} + \ldots + j_p\, y_{t-p} + \theta_q y_{t-q} + N_t$ ...(3)

Where j is autoregressive coefficient with p order and θ is moving average coefficient with q order. Construction of ARIMAX is similar iterative process as construction of univariate Box-Jenkins ARIMA model, i.e. identification, estimation and diagnostic checking. In the study, ARIMAX model is fitted to data by using R 4.0.5 software packages.”

Results and Discussion

The ARIMAX models were fitted using the time series groundnut yield data for the period 1966 to 2020 of Surguja district of Chhattisgarh. As described in methodology section, ARIMAX model is applied to fitting in time series data with explanatory variables. Time series data of groundnut yield and production are taken for future forecasting. The annual data on weather variables for the same time period has been used as input variables in the fitting of ARIMAX model. Selection of input variable has been done on the basis of positive higher correlation between weather variables (maximum temperature, minimum temperature rainfall and Precipitation) with crop yield and production. Table 1 showed that maximum and minimum temperatures have been found positive highly correlated with yield and production. Therefore maximum and minimum temperatures were used as input variable for developing model.

Table 1: Correlation between yield, production, evaporation, maximum temperature, minimum temperature, rainfall and precipitation

Variables	Evaporation	Maximum temperature	Minimum temperature	Precipitation	Rainfall	Production	Yield
Evaporation	1						
Maximum temperature	-0.489*	1					
Minimum temperature	-0.431*	0.572*	1				
Precipitation	0.532*	-0.427*	-0.486*	1			
Rainfall	0.263*	-0.408*	-0.363*	0.444*	1		
Production	-0.196	0.453*	0.442*	-0.339*	-0.253	1	
Yield	-0.195	0.481*	0.516*	-0.289*	-0.252	0.874*	1

The time plot of the yield and production series are given in figure 1 and 2 respectively and plot of maximum and minimum temperature are given in figure 3 and 4. In the Figures 1 and 2 shows that yield and production data is non stationary nature and figures 3 and 4 illustrated that maximum and minimum temperature was near to stationarity. For claiming non stationarity, unit root test have been used which is described below.

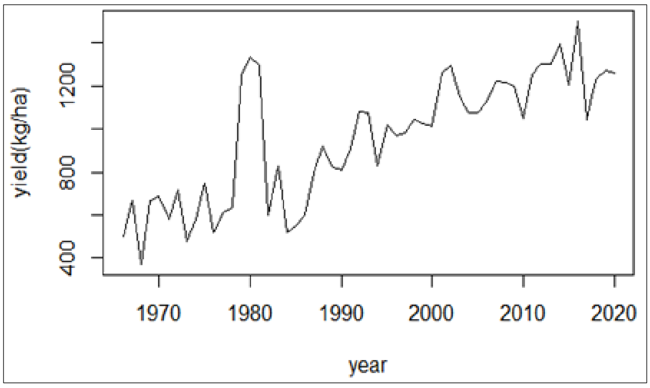

Fig 1: Time plot of yield

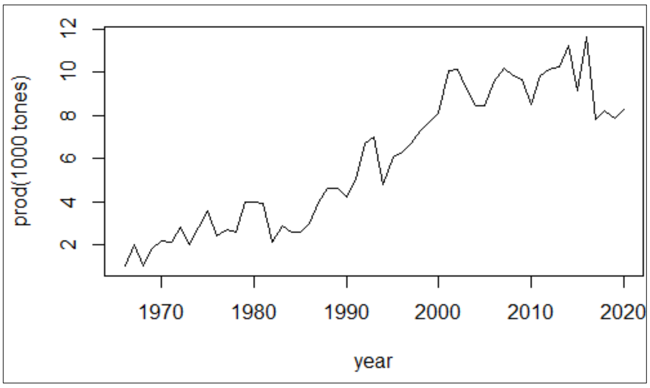

Fig 2: Time plot of Production

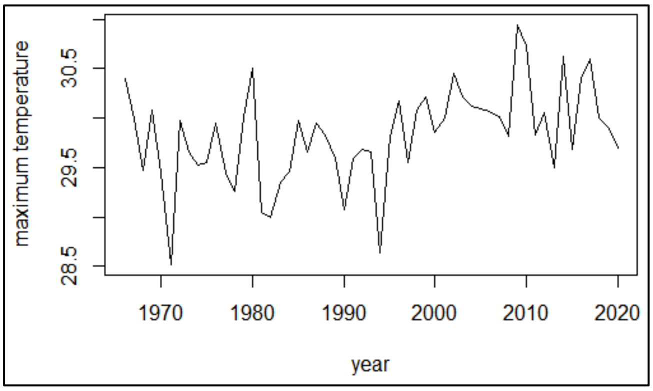

Fig 3: Time plot of maximum temperature

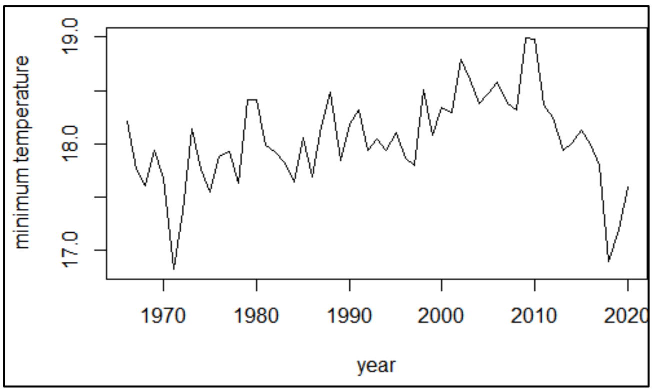

Fig 4: Time plot of maximum temperature

First of all variables i.e. yield, production, maximum and minimum temperature were tested for stationarity by using Augmented Dickey Fuller (ADF) test. Table 2 showed value of test statistics for ADF and it was found to be -4.959, -3.6178 -3.3760 and -5.8064 for yield, production, maximum temperature and minimum temperature series respectively. Hence after testing all the series are found to be nonstationary. Therefore, first differencing has been applied in all series and after first differencing all series becomes stationary.

Table 2: Augmented Dickey Fuller (ADF) test

Variable	Coefficient	Std. Error	t-Statistic	Prob.
yield	-0.6520	0.1314	-4.9594	0.0000
Production	-0.4194	0.1159	-3.6178	0.0007
Maximum temperature	-0.6974	0.2065	-3.3760	0.0015
Minimum temperature	-0.9377	0.1614	-5.8064	0.0000

There was used Marquardt algorithm (1963) to minimizing the sum of squared residuals. Schwarz's Bayesian Criterion, SBC (1978), minimum value of Akaike's Information Criterion, AIC (1969), and residual variance have been used for deciding the criteria to estimate autoregressive (AR) and Moving average(MA) in the model. Following are the Equation of best selecting model for yield and production.

Equation for yield

$\Delta y_1, = 14.8203 + 0.3324\, \Delta y_{1t} — 71.44\, x_1 — 9.31\, x_2 — 0.99\, \varepsilon_{t-i} + \varepsilon_t$

Equation for Production

$\Delta y_2, = 0.1284 - 0.3904\, \Delta y_{2t} — 0.5814\, x_1 — 0.0853\, x_2 + \varepsilon_t$

Where y_{1t} and y_{2t} denotes the yield and production of groundnut and x_1 and x_2 are maximum & minimum temperature in the year t. Δy_{1t} & Δy_{2t} are AR term and ε_{t-I} denotes MA term.

Table 3: coefficients of ARIMAX (1,1,1) for yield

	Constant	AR1	MA1	Maximum temperature	Minimum temperature
Coefficient	14.820	0.3324	-0.999	-71.436	-9.312
Standard error	2.0187	0.1336	0.0482	45.202	44.651
z value	7.3417	2.4879	-20.7550	-1.5804	-0.2086
Pr(>|z|)	2.109e-13 ***	0.01285 *	2.2e-16 ***	0.11402	0.83480
AIC	717.03				
BIC	728.96				

Table 4: coefficients of ARIMAX (1,1,0) for production

	Constant	AR1	Maximum temperature	Minimum temperature
Coefficient	0.1284	-0.3904	-0.5814	-0.0853
Standard error	0.0921	0.1261	0.2353	0.2362
z value	1.3939	-3.0968	-2.4714	-0.3613
Pr(>|z|)	0.163334	0.001956 **	0.013458 *	0.717902
AIC	156.3			
BIC	166.24			

The models ARIMAX(1,1,1) and ARIMAX(1,1,0) considered (Tables3 and Table4) in the identification stage for yield and production. After that ARIMAX estimation was carried out using a "non-linear least squares" method and finally residual has been checked for white noise in the diagnostic checking stage of ARIMA. Chi-Squared statistic(χ^2) has been used by Ljung-Box method (1978) [3]. Result of the statistic ruled out any systematic pattern in the residuals (Table5).

Table 5: Ljung-Box test for yield and production

Yield			Production		
Chi-square	d.f	P value	Chi-square	d.f	P value
11.615	10	0.3117	5.637	10	0.8448

After the comparison of best performing model, ARIMAX (1,1,1) for yield and ARIMAX (1,1,0) for production found to be the best fitted model for groundnut (Table 6). The selected models have been used to obtaining the groundnut yield and production forecasts. Figure 5 & 6 exhibited line graph of actual, fitted and forecasted values for yield and production.

Table 6: Performance of model

	Model	ME	RMSE	MAPE
Yield	ARIMAX(1,1,1)	1.909431	158.6178	13.751
Production	ARIMAX(1,1,0)	0.002314637	0.9277023	15.451

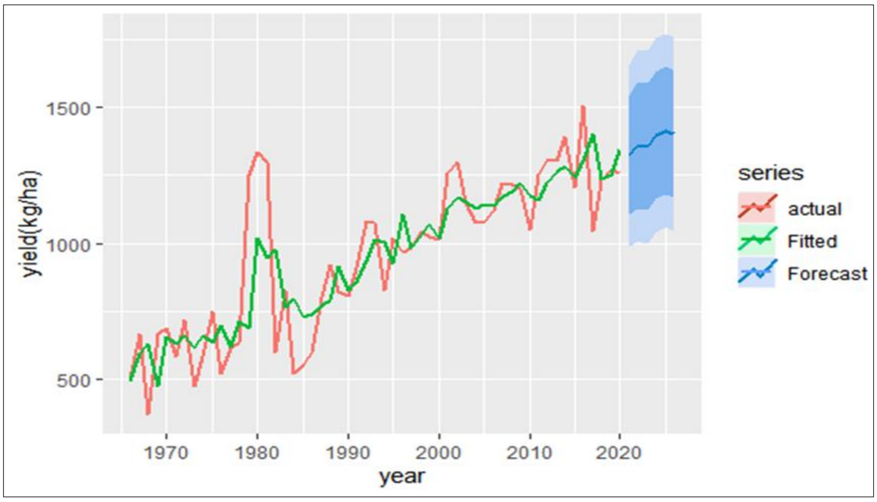

Fig 5: Plot of actual fitted and forecasted value for yield

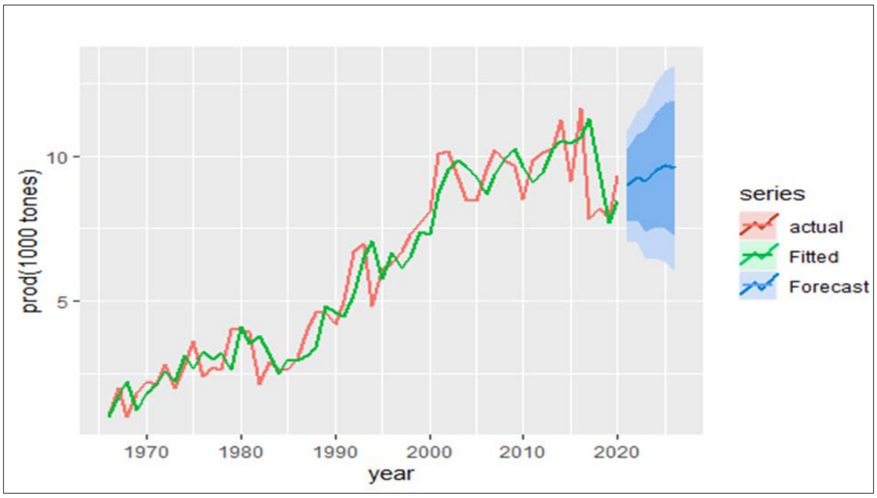

Fig 6: Plot of actual fitted and forecasted value for production

Conclusion

Groundnut yield and production data of Raipur district of Chhattisgarh have been analyzed and considering most important weather variables (maximum & minimum temperature) as input variables in ARIMAX model. Prediction for future value of yield and production has been done by the best selected ARIMAX model. Forecasted values of groundnut yield and production showed increasing trend. There also concluded that maximum and minimum temperatures are highly associated with groundnut yield and production in comparison to other weather variables.

References

1. Arya P, Paul RK, Kumar AK, Singh N, Sivaramne N, Chaudhary P. Predicting pest population using weather variables: an ARIMAX time series framework, Int. J Agricult. Stat. Sci 2015;11(2):381-386.
2. Ayyoob KC, Krishnadas M. Production forecast of groundnut (*Arachis hypogaea* L.) using crop yield-weather model, Agriculture Update 2013;8(3):436-439.
3. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control. Holden-Day, San Fransisco 1976.
4. Celik S, Karadas K, Eyduran E. Forecasting the production of groundnut in Turkey using arima model, The Journal of Animal & Plant Sciences 2017;27(3):920-928.
5. Chadsuthi S, Modchang Lenbury CY, Iamsirithaworn S, Triampo W. Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and arimax analyses. Asian Pacific Journal of Tropical Medicine 2012;5:539-546.
6. Chouksey N, Mishra GC, Singh SK, Singh M, Chouksey R., Forecasting of Nitrogen Content in the Soil by Hybrid Time Series Model, International Journal of Current Microbiology and Applied Science 2018a;7:7.
7. Chouksey N, Mishra GC, Singh SK, Singh M, Chouksey R. Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Rice Yield on Different Fertilizer Combinations, International Journal of Agriculture, Environment and Biotechnology 2018b;11:1029-1033.
8. Dickey D, Fuller W. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. American Statistical Association 1979;74:427-31.
9. Donald Marquardt W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics 1963;11(2):431-441
10. Akaike H. "Fitting autoregressive models for prediction", Ann. Inst. Statist. Math 1969;21:243-247.
11. Mishra P, Sahu PK, Uday JPS. ARIMA Modeling Technique in Analyzing and Forecasting, Fertilizer Statistics in India 2014;7(13):170-176.
12. Pal S, Mazumdar D. Forecasting groundnut production of India using nonlinear growth models. J Crop and Weed. 2015;11(Special Issue):67- 70.
13. Sanjeev Verma Urmil. Time-series modeling and forecasting of sugarcane yield in Haryana, M. Sc. Dissertations, Chaudhary Charan Singh Haryana Agricultural University 2015.
14. Schwarz G. Estimating the dimension of a model. Annual Statistics 1978;6:461-464.
15. Yogarajah B, Elankumaran C, Vigneswaran R. Application of arimax model for forecasting paddy production in trincomalee district in Sri Lanka. Proceedings of the Third International Symposium, SEUSL: Oluvil, Sri Lanka, 2013.