International Journal of Statistics and Applied Mathematics 2023; 8(4): 102-106

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(4): 102-106

https://www.mathsjournal.com
Received: 04-05-2023
Accepted: 03-06-2023

Upendra Kurmi
Department of Mathematics and Statistics, CoA Jabalpur, JNKVV, Jabalpur, Madhya Pradesh, India

Umesh Singh
Department of Mathematics and Statistics, CoA Jabalpur, JNKVV, Jabalpur, Madhya Pradesh, India

RB Singh
College of Agriculture, Ganj Basodam JNKVV, Madhya Pradesh, India

Mujahida Sayyed
College of Agriculture, Ganj Basodam JNKVV, Madhya Pradesh, India

Corresponding Author:
Umesh Singh
Department of Mathematics and Statistics, CoA Jabalpur, JNKVV, Jabalpur, Madhya Pradesh, India

Forecasting of area, production, and productivity of groundnut using nonlinear growth model in India

Upendra Kurmi, Umesh Singh, RB Singh and Mujahida Sayyed

Abstract
India's largest agricultural export is groundnut, which is a significant source of oilseeds. By giving the expanding population access to inexpensive food goods and job opportunities, it might accelerate the country's economic growth. In this study, nonlinear growth models such the Monomolecular, Logistic, Gompertz, Richards, and Weibull are examined. Non-linear equations may be solved iteratively using a number of techniques, and parameters can be approximated using the least-squares approach. The parameters of these models were evaluated using the RStudio programme with the Levenberg-Marquardt approach being one of the most important. To select the best-fitted model, a number of goodness of fit metrics were utilised, such as R^2, MAE, MSE, and RMSE. Nonlinear models such as the monomolecular, logistic, gompertz, Richards, and weibull models were used to examine groundnut data of India. The goodness of fit criteria was used to choose the non-linear model that best suited the data series. It has been determined that the most appropriate models for the area, production, and productivity of groundnut are the Gompertz, Weibull, and Logistic models.

Keywords: Non-linear models, Levenberg-Marquardt technique, R^2, MAE, MSE, and RMSE

Introduction
Groundnut is cultivated on 6.09 million hectares, yielding 10.21 million metric tonnes with a productivity of 1676 kg per hectare (Agricultural Statistics at a Glance, 2021). It is referred to as the "king" of oilseeds and produces around 25% of all the oilseeds produced in the nation. Although groundnut can be grown in one or more of the seasons (Kharif, Rabi, and Summer), the Kharif crop makes up around 80% of the land and production (June to October).Oil, protein, and starch make up between 47 and 53 percent and 26 and 11.5 percent of its seeds, respectively. By giving the expanding population low-cost food options and creating jobs for them, it has the ability to accelerate the country's economic growth.
Forecasting is crucial to deciding whether an agribusiness succeeds or fails since farmers face a lot of issues, such as weather, supply and demand, and resource shortages. Producer yields are significantly impacted by the variable output, productivity, and value that are brought on by these factors. Additionally, the study that used various growth models to address these issues applied improper statistical techniques. Using several nonlinear regression growth models, the current study was conducted to examine the growth trend of groundnut in India from 1980–1981 to 2020–2021. In this study, the Levenberg-Marquardt method was used to anticipate groundnut output, productivity, and area in India utilising a number of significant nonlinear deterministic growth models, including the Monomolecular, Logistic, Gompertz, Richards, and Weibull models.

Methodology
The study was based onnonlinear statistical models that can be expressed as

$$Y(t_i) = f(t_i, \theta) + \xi_i, \quad i=1,2,3........n \qquad (1)$$

We assume that errors term ξ_i are iid with zero mean and constant variance σ^2

Nonlinear statistical model parameter estimation

Although this convergence is very sluggish towards the end of the iterative phase, the Steepest Descent approach may converge on actual parameter values even when the initial values are far from the real parameter values. Levenberg-Marquardt algorithm is the most extensively used technique for calculating nonlinear least squares estimates (Levenberg, 1944, Marquardt 1963) [5]. The gradient (or steepest descent) method and the linearization (Taylor series) method are effectively combined in this method. This study took into account the relative change in the residuals sum of squares on subsequent iterations, which eliminates the series disadvantage of both techniques.We employed the following well-known nonlinear growth models in this study:

Gompertz Model-In honour of Benjamin Gompertz, a sort of mathematical time series model is known as the Gompertz curve or Gompertz function (1779-1865). Based on a sigmoid function, growth is slowest at the beginning and end of a particular time period. The curve approaches the function's right-side or future value asymptote considerably more gradually than it does the left-side or lower valued asymptote.

$$y_t = \alpha \exp^{\{-\exp^{(\beta-\gamma^* t)}\}} + \xi_0$$

1. Monomolecular Model

$$Y_t = \alpha^*(1 - \beta^* exp(-k^* t))$$

2. Logistic Model

$$Y_t = \frac{\alpha}{1+\beta \exp^{(\gamma^* t)}} + \xi_0$$

3. Richards model

$$y(t) = \frac{1}{\{\alpha + \beta \exp^{(\gamma^* t)^{\delta}}\}} + \xi_0$$

4. Weibull model

$$y(t) = \alpha - \beta \exp^{(-\gamma^* t^{\delta})} + \xi_0$$

Goodness of fit criteria for nonlinear Statistical models

Here, we take into account the following goodness of fit statistic to compare distinct rival models.

Coefficient of Determination (R^2)

The percentage of the dependent variable's variance that is pre dicted from the independent variable is known as the coefficie nt of determination or R squared method. It shows how much variety there is in the given data collection.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

The formula for calculating adjusted R squared is as follows, we calculate the Adjusted R^2 for the model comparison because selected models have different parameters.

$$\text{Adjusted } R^2 = \frac{(1-R^2)(N-1)}{N-p-1}$$

Where R^2 is sample R Squared, p: Number of predictors and N is Sample size

Root Mean Square Errors (RMSE) -One of the methods most frequently used to assess the accuracy of forecasts is root mean square error, also known as root mean square deviation. It indicates the Euclidean distance between measured true values and forecasts.

$$\text{RMSE} = \sqrt{\left\{\sum_{i=1}^{n}\frac{(Y_i - \hat{Y})^2}{n}\right\}}$$

Mean Absolute Error (MAE)

$$MAE = \sum_{i=1}^{n}\frac{|Y_i - \hat{Y}|}{n}$$

Mean Square Error (MSE)

$$MSE = \sum_{i=1}^{n}\frac{(Y_i - \hat{Y})^2}{(n-P)}$$

Examination of residuals

The variations between what was actually observed and what the fitted model predicts are called residuals. The models' two key assumptions are as follows: (i). Residuals are random. (ii). Residuals are normally distributed. We now discuss the tests for assumptions (i) and (ii) above.

a) Test for the randomness of residuals (Run test)
b) Residuals are normally distributed (Shapiro-Wilk test)

Result and Discussion

Initially, attempts were undertaken to discover the non-linear growth model that best described the area, production, and productivity of groundnut in India.

For data analysis, R-studio and SPSS.26 statistical tools were used. Firstly, we separated the entire groundnut data into two parts: training data and testing data. Training data was collected from 1981 to 2016, and testing data were collected from 2017 to 2020. The best model was chosen from the training data and validation was done using the testing data. The best model selected from the testing data was used for prediction of area, production, and yield of groundnut from 2021 to 2025.

To assure "convergence," various initial values were tried, and the results were reported. The iterative procedure has converged for a set of initial parameter values, and the parameter estimates after convergence are provided. To examine model performance, coefficient of determination (R^2), Mean Absolute Error (MAE), Mean square error (MSE), and Root Mean Square Error (RMSE) were used as the goodness of fit criterion. The results of several non-linear growth models for area, production, and yield of testing data are represented, and the best model is chosen from the groundnut testing data. The selected best model is used for forecasting of area, production, and productivity of groundnut for the period 2021 to 2025.

Fitting of various Non-linear Growth Models for training data

On the basis of training data from 1981 to 2016, the most suited model for groundnut area, production, and yield in India was determined. To assure convergence of the parameter estimations for the groundnut data, different sets of initial parameter values were attempted. Tables 1(a) show the parameter estimates after convergence of groundnut time series data.

Table 1(a): Estimated parameter values for various non-linear growth models for the area, production, and yield of groundnut

	Initial value	Monomolecular	Gompertz	Logistic	Weibull	Richard
Area	α	0.098	7.898	7.937	-108.319	5.531
	β	-83.756	-4.495	-0.163	-16043.945	-5.410
	ϒ	0.011	-0.10	0.085	4.921	0.017
	δ				0.001	-0.017
Production	α	7.096	7.239	7.096	639.154	0.138
	β	4156.282	-0.907	-28732.184	631.858	0.065
	ϒ	9.557	0.316	-9.531	-0.003	-0.045
	δ				-0.998	7.409
Yield	α	-4.253	213454.452	844.251	28039.381	-0.125
	β	181.207	1.729	26.603	239.832	0.127
	ϒ	-0.0165	0.00316	0.090	-4.421	-0.001
	δ				-0.001	0.238

The estimated parameter values of Monomolecular, Gompertz, Logistic, Weibull, and Richard models for the area, production, and yield of groundnut for the period 1981 to 2016 have been given in table 1(a). These values were obtained when the iterative procedure is converged.

Goodness of fit statistic for the area, production, and yield of groundnut is represented in table 1(b) The highest R^2 was observed in case of Gompertz for area, Weibull for production, and logistic model for yield with the lowest value of MAE, MSE, and RMSE respectively in comparison to that of other growth models. Based on performance of these models, it was found that Gompertz for area, Weibull for production, and logistic model for yield fit very well with the data than other models. The performance of several non-linear growth models for area, production, and yield of testing data is presented in Table 2. The best model is chosen from total groundnut testing data. Selected best model is used to predict groundnut area, production, and yield from 2021 to 2025.

Table 1(b): Goodness of fit statistics for the area, production, and yield of groundnut for various non-linear growth models

		Monomolecular	Gompertz	Logistic	Weibull	Richard
Area	R^2	0.554	0.732	0.726	0.275	0.517
	MAE	0.65	0.489	0.493	0.78	0.67
	MSE	0.55	0.33	0.34	0.90	0.60
	RMSE	0.74	0.576	0.582	0.95	0.77
Production	R^2	0.28	0.35	0.29	0.54	0.43
	MAE	1.1	1.094	1.090	1.10	1.093
	MSE	1.92	1.87	1.92	1.89	1.87
	RMSE	1.39	1.367	1.39	1.37	1.366
Yield	R^2	0.496	0.493	0.535	0.345	0.514
	MAE	134.26	134.65	133.59	159.85	132.50
	MSE	31744.44	31968.30	29338.32	41310.21	30614.73
	RMSE	178.17	178.80	171.28	203.25	174.97

By comparing goodness of fit statistics for groundnut area, production, and yield from 1981 to 2016. According to Table 4.1(b) it was observed that Gompertz is found to be best fit model followed by Logistic model for the area, Weibull model is found to be best-fit followed by Richard model for production and Logistic is the best fit model followed by Richard for yield of groundnut with R^2 values 0.732, 0.54, and 0.535 respectively.

Fitting of various Non-linear Growth Models with testing data

Table 2 gives the performance of various non-linear growth models for area, production and yield of groundnut in India for the period 2017 to 2020. It is observed that Gompertz model is best fit for area, Weibull model is best fit for production, and logistic model is best fit for yield with minimum percent relative deviation of 0.30, 1.31, and 245.55 respectively for area, production and yield of groundnut in India.

Table 2: Performance or Validation of various non-linear growth models for area, production and yield of groundnut for the period 2017 to 2020

	Year	Observed value	Predicted Value				
			Monomolecular	Gompertz	Logistic	Weibull	Richards
Area	2017	5.34	5.56	5.03	4.15	6.61	5.60
	2018	4.89	5.50	4.79	3.82	6.59	5.55
	2019	4.73	5.44	4.55	3.45	6.58	5.51
	2020	4.89	5.38	4.29	3.05	6.56	5.46
Average of RD (%)			0.51	0.30	1.35	1.62	0.57
Production	2017	7.46	7.10	7.24	7.10	7.35	7.25
	2018	9.25	7.10	7.24	7.10	7.35	7.25
	2019	6.73	7.10	7.24	7.10	7.34	7.25
	2020	9.95	7.10	7.24	7.10	7.34	7.25
Average of RD (%)			1.43	1.36	1.43	1.31	1.36
Yield	2017	1398	1414.8	1416.7	1587.4	1218.9	1127.9
	2018	1893	1438.4	1439.2	1657.4	1223.4	1167.4
	2019	1422	1462.4	1462.0	1734.0	1227.9	1209.7
	2020	2063	1486.8	1485.1	1817.8	1232.2	1255.1
Average of RD (%)			272.00	272.5	245.5	468.3	503.9

RD = Relative Deviation

Forecasted Values of groundnut area, production, and yield in India: Forecasted values of groundnut area, production, and yield in India for the period 2021 to 2025 by using the best-fitted model for testing data are presented in Table 3.

Table 3: Forecasted values of area, production, and yield of groundnut by using best-fitted model for the period 2021 to 2025

Year	Area	Production	Yield
2021	4.03	7.249	1909.56
2022	3.75	7.250	2009.89
2023	3.47	7.251	2119.66
2024	3.18	7.252	2239.77
2025	2.89	7.253	2371.20

Plot of Observed Data and Predicted Values for the Whole Period (1981-2020) of groundnut by the best-fitted model: An attempt has been made to plot observed and predicted groundnut area, production, and yield data by using the best-fitted model for testing data developed for the whole period 1981 to 2020are presented in fig. 4.4(a), 4.4(b) and 4.4(c). The observed and expected values for area, productivity, and yield are directly comparable according to the plot. It has been recognized from predicted values that India's groundnut production and yield are projected to increase, while the area is to decline slowly and steadily.

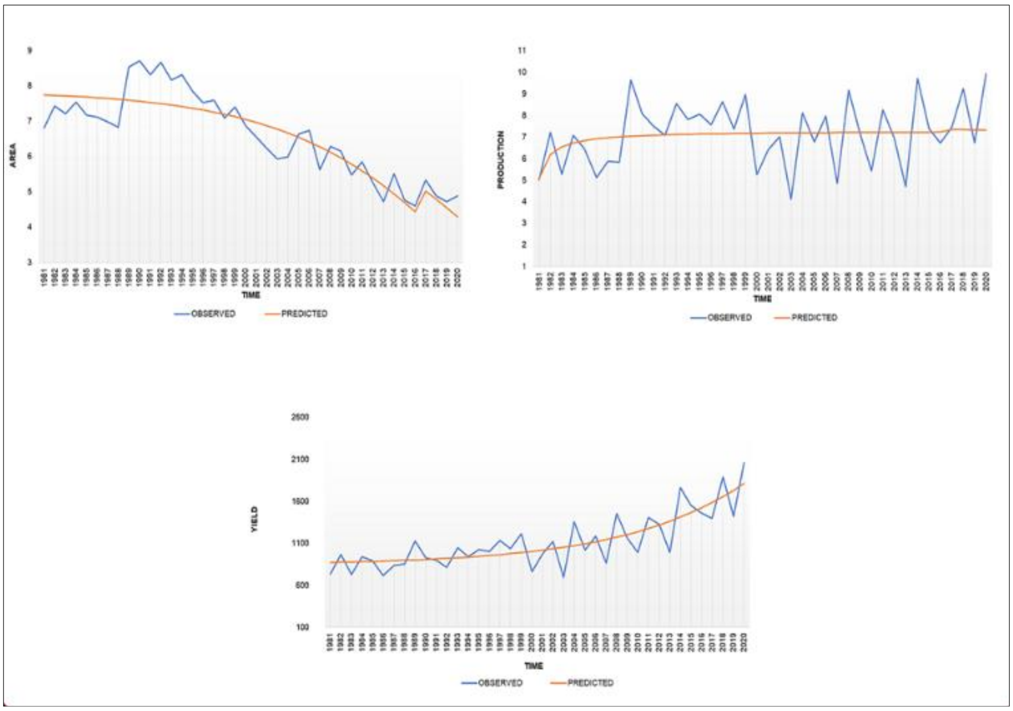

Fig 1: Plot of observed data and predicted value of groundnut area, production and yield obtained by best fitted models for the period 1981 to 2020

Examination of residuals

The model's two key assumptions are error randomness and normality. To validate this assumption, residuals analysis should be performed. If our fitted model is right, the residual analysis should support the model's assumptions or, at the very least, not deny them.

Table 4: Randomness test for the residuals and Shapiro-wilks of best-fitted model

	Run test			Normality Test		
Models	**No. of runs**	**Statistic (Z)**	**P-Value**	**N**	**Statistic (W)**	**P-value**
Area (Gompertz)	13	-2.553	0.0151	40	0.9308	0.01706
Production (Weibull)	24	1.269	0.2043	40	0.9258	0.01175
Yield (Logistic)	20	-0.1804	0.8569	40	0.919	0.007153

The randomness of residuals was checked using the run test, and the normality of the residuals was tested using the Shapiro-Wilk test, as detailed in the preceding chapter material and methods. Table 4(a) shows the value of z for the run test for each model, along with the number of runs and the test value. The values of z for area are significant, indicating rejection of the null hypothesis at the 1% and 5% level of significance, i.e., residuals are not random. Whereas the z value for the production and yield is non-significant, it indicates that the null hypothesis is accepted at the 5% level of significance, implying that the residuals are random. The Shapiro-Wilk W statistic is shown in table 4(b). The values are highly significant for all residuals at 5% level of significance, indicating that the model's normality assumption is not being fulfilled, but at 1% level of significance are non-significance for area and production, indicating that acceptance of the null hypothesis, i.e., residuals are normally distributed except yield.

Conclusion

The groundnut production, yield, and area of India from 1981 to 2020 were the subject of the study. The secondary data on groundnut were collected from the Directorate of Economics and Statistics, New Delhi. It is reasonable to study the application of a non-linear growth model to groundnut growth since growth in biological things is typically non-linear. Different non-linear growth models, including Monomolecular, Logistic, Gompertz, Richards, and Weibull are investigated in this research. As in case of non-linear model, parameters may be estimated using the least-squares approach and various iterative procedures are used for solving non-linear equations. Levenberg-Marquardt approach is among the most significant of them; the parameters of these models were estimated using the RStudio program and SPSS version 26.0. The best-fitted model was chosen using a variety of goodness of fit measures, includingR^2, MAE, MSE, and RMSE. For groundnut data of India, the non-linear models, including the Monomolecular, Logistic, Gompertz, Richards, and Weibull models, were used. The best non-linear model for the data series was chosen based on how well the goodness of fit criterion performed. In comparison to various models, the Gompertz model is determined to be the better suitable model for the area, Weibull model is best-fitted for production, and the Logistic model is best fit for yield of groundnut. Finally, we came to the conclusion that no model that had been evaluated was adequate for the groundnut.

Reference

1. Akshu Sharma L. Production of food grain in Haryana: A District wise analysis. International Journal of Academic Research and Development. 2017;2(3):182-187.
2. Archontoulis SV, Miguez FE. Non-linear Regression Models and Applications in Agricultural Research. Agronomy Journal. 2013;107(2):1-13.
3. Draper NR, Smith H. Applied Regression Analysis, 2nd Edition. John Wiley and Sons, New York; c1981.
4. Gompertz B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Method of Determining the Value of Life Contingencies. Phil. Trans. Roy. Soc. 1825;115:513-585.
5. Levenberg K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. Quarterly of Applied Mathematics. 1944;2:164–168.
6. Marquardt D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. SIAM J Appl. Math. 1963;11:431-441.
7. Mohanty M, Painuli DK, Mishra AK, Bandyopadhaya KK, Ghosh PK. Estimating impact of puddling, tillage, and residue management on wheat seedling emergence and growth in a rice-wheat system using non-linear regression models. Soil and Tillage Research. 2006;87(1):119-130.
8. Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples). Biometrical. 1965;52:591-611.
9. Singh U, Singh RB, Gautam SS. Stochastic models for describing growth of soybean production and soya oil production in India and Madhya Pradesh JNKVV Res. J. 2013;47(2):233-38.
10. SoyStats®. A reference guide to important Soybean facts & figures., A publication of the American Soybean Association; c2013.
11. Weibull WA. A statistical distribution of wide applicability. Journal of Applied Mechanics. 1951;18:293–297.