The Pharma Innovation Journal 2023; 12(2): 1900-1905

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(2): 1900-1905

www.thepharmajournal.com
Received: 22-11-2022
Accepted: 26-12-2022

Mahesh Kumar
Department of Basic Sciences and Languages (SMCA, P & L), Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar, India

Kalyan Singh Paikra
Department of Basic Sciences and Languages (SMCA, P & L), Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar, India

Corresponding Author:
Mahesh Kumar
Department of Basic Sciences and Languages (SMCA, P & L), Dr. Rajendra Prasad Central Agricultural University, Pusa, Samastipur, Bihar, India

Forecasting of chickpea yield using nonlinear model in Darbhanga district of Bihar

Mahesh Kumar and Kalyan Singh Paikra

Abstract
This research paper entitled "Forecasting of Chickpea yield using nonlinear model in Darbhanga district of Bihar" is based on the secondary data. Data was collected for the years 1966 to 2017 from the official sites of Department of Statistics and Economics of Bihar and ICRISAT, Hyderabad for achieving objective, data from 1966 to 2015 were analysed through R- Software and two years data 2016 and 2017 were kept for model validation of yield forecasting of Chickpea in Darbhanga district. For forecasting chickpea yield in Darbhanga district of Bihar, three different nonlinear models namely Logistic, Gompertz and Monomolecular were used.
All three non-linear models were fitted to data by using Statistical software R. For validation of assumptions of residuals i.e., randomness and normality of residuals, Run's test and Shapiro wilk's tests were employed respectively while for goodness of fit and validation of models, Chi-square test and eight steps ahead forecasting were done. For getting best fitted models for forecasting yield, models are compared by seven different statistics $R^{2,}$ RSS, MAPE, MAE, MSE, RMSE. So, after analysing the data, Logistic model is found better for Darbhanga district (Bihar) with FE% of 0.89% and 3.57% for year 2016 and 2017 respectively. Forecasting of chickpea yield is made for 2023 will be 1.18 t/h. Forecasting model of chickpea yield is best fitted model (i.e. Logistic) as below.
$\hat{Y}$ =1.3613/(1+(1.3613/0.3684-1) *exp(-0.0497*t))

Keywords: Nonlinear growth model, forecasting, run's test, Shapiro wilk's test

Introduction
Chickpea is known to have originated in western Asia (probably eastern Turkey). The cultivated chickpea is not found in the wild and *C. reticulatum* is its progenitor, while *C. echinospermum* is a close relative. From west Asia, it spread to Europe and in more recent times, it was introduced in tropical Africa, central and southern America and Australia.
Introduction of chickpea in India appears to be independent in the north and Peninsular India. The earliest record of chickpea in India is from Atranji Khera in Uttar Pradesh and this dates backs to 2000 BC. Its introduction to Peninsular India appears to be between 500 and 300 BC. In India chickpea is cultivated (2005-06) on 7.1 M ha, producing 5.7 M t with a productivity of 800 kg ha^{-1}. Madhya Pradesh with 2.6 M ha producing 2.4 M t with a productivity of 930 kg ha^{-1} ranks first in area, production and productivity. Rajasthan ranks second in area and production closely followed by Uttar Pradesh. Productivity is highest (1.6 t ha^{-1}) in AP followed by WB (1.1 t ha^{-1}). These three states along with Maharashtra account for 84 per cent of area and 86 per cent of total chickpea production in the country. The national average productivity is 800 kg ha^{-1} (2005-06), which is much less than the highest average productivity (1.8 t ha^{-1}) in Egypt.
Global production of chickpeas is around 7.5 M t from an area of 10 M ha with a productivity of 750 kg ha^{-1} India is the premier chickpea growing country accounting for 77 per cent of the total area and production in the world. Other important chickpea producing countries are Pakistan, Turkey, Mexico, Burma and Ethiopia. Major chickpea growing states in India are Madhya Pradesh, Rajasthan, Uttar Pradesh and Maharashtra.
Chickpea (*Cicer arietinum* L.) is a legume crop from Fabaceae family. Chickpea is one of most consumed pulses around the world. Global production of chickpea is around 15 million tonnes during year 2020. India is leading country in world production of chickpea with share of 71% of global total with production of around 11 million tonnes with productivity of 1036 kg/ha (FAOSTAT 2020). It is most important pulse crop in India. Chickpea contributes about 49.3% in total pulses production in India. Among other pulses, chickpea dominates in both area and production. In Bihar area under chickpea has decreased to 0.51 lac ha in 2020-21

from 0.52 lac ha in 2019-20 with production and productivity of 0.54 lac tonnes and 1052 kg/ha respectively (DSE, Bihar).

Review and Literature

Karadavut *et al.* (2010) [2] research was attempted to compare the performance of non-linear model based on leaf data of maize. Some non-linear models i.e. Richards model, Logistic model, Weibull model, MMF model and Gompertz model were used to fit with leaf data. Coefficient of determination (R^2), sum of squares error (SSE), root mean square error (RMSE), and mean relative error (MRE) were used to compare models. The result showed that to determine leaf growth of maize Richards, Logistic and Gompertz model were better than other models under consideration.

Pal *et al.* (2015) [4] study was carried out on forecasting of production of groundnut using time series data from period 1950-51 to 2011-12. Analysis was carried out using three different nonlinear models. Gauss- Newton algorithm was used to estimate the parameters of model and monomolecular and logistic model was found better then gompertz model.

Rajan *et al.* (2017) [7] study was conducted to compare six different non-linear growth model Logistic, Gompertz, Monomolecular, Richards, Quadratic and Reciprocal model fitness to data based on cotton in India. The data was collected from period of 1980-2013 for fitting models and R^2, RSS and RMSE were used check goodness of fit. To check randomness and normality of error term “Run Test” and “Shapiro-Wilks” test was used respectively.

Materials and Methods

The study was done with objective forecast yield of chickpea by fitting different non-linear growth models. The study was carried out using secondary data of yield of chickpea crop of Darbhanga district (Bihar). The time series data were gathered from the official site of ICRISAT, Hyderabad and Department of Statistics and Economics of Bihar. The data were taken from period of 1966 to 2017 i.e., 52 years, where the data from 1966 to 2015 were taken for analysis while data of year 2016 to 2017 were used for purpose of model validation.

In practice, it’s May not be always linear relationship among explanatory and response variables. So, nonlinear model can be employed to describe exact relationship. In this study, nonlinear growth models are employed in the way to get best fit model for forecasting yield of chickpea crop in Darbhanga district from Bihar.

Non-Linear Model

Every statistical inquiry where principles from some body of knowledge enter into the analysis most likely it will lead to a nonlinear model (Seber and Wild, 2003) [6]. These nonlinear models play crucial role to get complex relationship between variables. A model which exhibits nonlinearity for at least one parameter is called as nonlinear model. Occurrence of growth leads to the model needed for a particular time. The process involves making assumptions about the type of growth that is occurring, highlighting differential equations that represents those assumptions and then solving those equations to frame an appropriate model. The nonlinear models used in this study are

I. Logistic model.
II. Gompertz model.
III. Monomolecular model.

I. Logistic model

The logistic model is appropriate in situation when growth rate is directly proportional to the product of the present size and the further amount of growth. it is given by

$$\frac{dy}{dt} = r*y(1-y/k) \qquad \dots \text{(i)}$$

On integrating it,

$$Y_t = k/(1+(k/y0-1)*\exp(-r*t) \qquad \dots \text{(ii)}$$

II. Gompertz model

Differential equation of this model is expressed as

$$\frac{dy}{dt} = r*y*\ln(k/y) \qquad \dots \text{(iii)}$$

On integrating it, solution of equation will be

$$Y_{(t)} = k*\exp(\ln(y0/k)*\exp(-r*t) \qquad \dots \text{(iv)}$$

III. Monomolecular model

The model describes the progress of a growth at any time is proportional to the resources yet to be achieved. Differential equation of this model is given by

$$\frac{dy}{dt} = r\,(k-y) \qquad \dots \text{(v)}$$

While its solution is given by

$$Y(t)=k-(k-y0)*\exp(-r*t) \qquad \dots \text{(vi)}$$

Parameter Estimation

Let’s consider nonlinear model represented by

$$Y_t = f(x_t,\beta) + u_t;\ t = 1, 2 \dots, n \qquad \dots \text{(vii)}$$

Y_t = a independent variable
X_t= explanatory variables
β = unknown parameters

u_t are error terms which are unobservable and identical, independent and follows normal distribution with mean zero and constant variance.

The non-linear least-squares (NLLS) estimator is that estimator which minimize the value of sum of squared residuals and denoted by $\hat{\beta}$.

$$Sn(\beta)=\sum_{t=1}^{n}[Y_t - f(X_t,\beta)]^2$$

The differentiation is to be made of equation (3.10) with respect to β which will give normal equations and by solving them, we get values of β.

The normal equation can be expressed as

$$\sum_{t=1}^{n}\left[y_t - f\left(X_t,\hat{\beta}\right)\right]^2 \left(\frac{\partial f(X_t,\beta)}{\partial \beta_1}\right)_{\beta=\hat{\beta}} = 0$$

There doesn’t exist any explicit formula to get nonlinear least squares estimator $\hat{\beta}$, so to minimize equation. Some iterative method to be used. So, some nonlinear estimation procedure like Gauss-Newton algorithm, steepest descent algorithm and

Levenberg-Marquardt algorithm, can be employed to fit models. The detailed description of these algorithm is given in Draper and Smith (1998) [1]. Gauss-Newton algorithm is used in this study for fitting models and analysis is diner using R software.

For goodness of fit and validation of models, Chi-square test and OSAF (one step ahead forecast) test are used. For getting best fitted models for forecasting yield, models are compared by nine different statistics $R^{2,}$ RSS, MAPE, MAE, MSE, RMSE The residuals are examined by run test (Ratkowsky 1990) and Shapiro wilk's test for randomness and normality assumptions respectively.

Results and Discussion

For forecasting chickpea yield for Darbhanga district (Bihar), three different nonlinear models have been used. The models have been fitted on chickpea yield data from period of 1966 to 2015, while data of year 2016 and 2017 were kept for model validation purpose.

Fitting of different nonlinear model

Initial values for parameters of nonlinear model are required while doing estimation by fitting nonlinear model. The values are estimated by using Draper and Smith (1998) [1] method and used to nonlinear models. The estimated initial values are used in R software for fitting different nonlinear models, i.e., Monomolecular, Logistic and Gompertz model. The results of the model are given below.

Table 1: Summary statistics for Monomolecular model of chickpea yield (t/ha) in Darbhanga district (Bihar)

Parameters	Estimate	Std. Error	t-value	Pr(>|t|)
K(Carrying capacity)	3.505352	8.196193	0.428	6.71E-01
y0(Initial value at time=0)	0.346818	0.079564	4.359	7.07E-05***
r(Intrinsic growth rate)	0.005684	0.017005	0.334	0.74

It is revealed from the above table that among three parameter y0 are statistically significant at 1% level of significance while r is not.

Table 2: Summary statistics for Gompertz model of chickpea yield (t/ha) in Darbhanga district (Bihar)

Parameters	Estimate	Std. Error	t-value	Pr(>|t|)
K(Carrying capacity)	1.62353	0.75907	2.139	3.77E-02*
y0(Initial value at time=0)	0.35736	0.06876	5.197	4.31E-06***
r(Intrinsic growth rate)	0.02806	0.01837	1.527	0.1334

It is revealed from the above table that among three parameters k and y0 is statistically significant at 5% and 1% level of significance while intrinsic growth rate (r) is not.

Table 3: Summary statistics for Logistic model of chickpea yield (t/ha) in Darbhanga district (Bihar)

Parameters	Estimate	Std. Error	t-value	Pr(>|t|)
K(Carrying capacity)	1.36134	0.36559	3.724	5.25E-04***
y0(Initial value at time=0)	0.36839	0.06169	5.972	2.97E-07***
r(Intrinsic growth rate)	0.04973	0.02023	2.459	0.017679

It is revealed from the above table that among three parameters k and y 0 is statistically significant at 1% level of significance while r is not.

Criteria for selecting the good fitted model

To choose the best fitted model among three different nonlinear model used in study, seven different statistics were compared which are R^2 (coefficient of determination), Residual sum of squared (RSS), Mean absolute percentage error (MAPE), Mean absolute error (MAE), Mean square error (MSE), Root mean square Error (RMSE).

Table 4: Comparison of Monomolecular, Gompertz and Logistic models for chickpea productivity (t/ha) in Darbhanga district (Bihar)

Statistics	Monomolecular	Gompertz	Logistic
R^2	0.95430	0.95949	0.95958
RSS	1.60589	1.42339	1.42008
MAPE	18.25065	17.48445	17.30340
MAE	0.13499	0.12797	0.12680
MSE	0.03149	0.02791	0.02785
RMSE	0.17745	0.16706	0.16687

From table - 4, it can be concluded that Logistic model have highest R^2 than other two models (monomolecular and Gompertz model) and even RSS, MAPE, MAE, MSE values are also low for Logistic model. So, it can be concluded that Logistic model is best fitted for chickpea yield in Darbhanga district of Bihar. Fig.-1 shows the graph of actual v/s fitted yield for Monomlecular model where time in year (t) in x-axis and yield (t/ha) in y-axis.

However, assumptions regarding randomness and normality of residuals are also checked using run test and Shapiro wilk's test respectively and showed in table-5 and table-6.

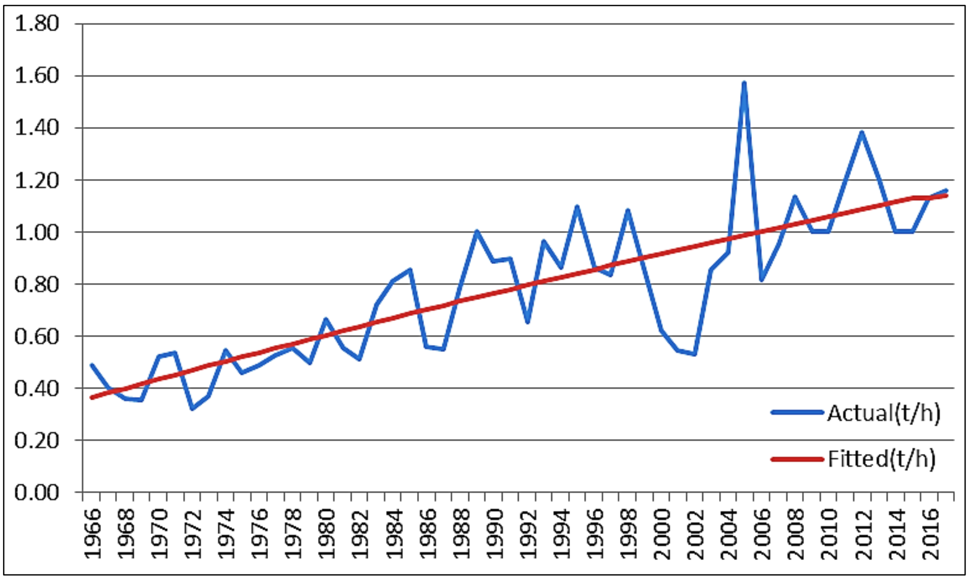

Fig 1: Actual yield vs. Fitted yield (Monomolecular model)

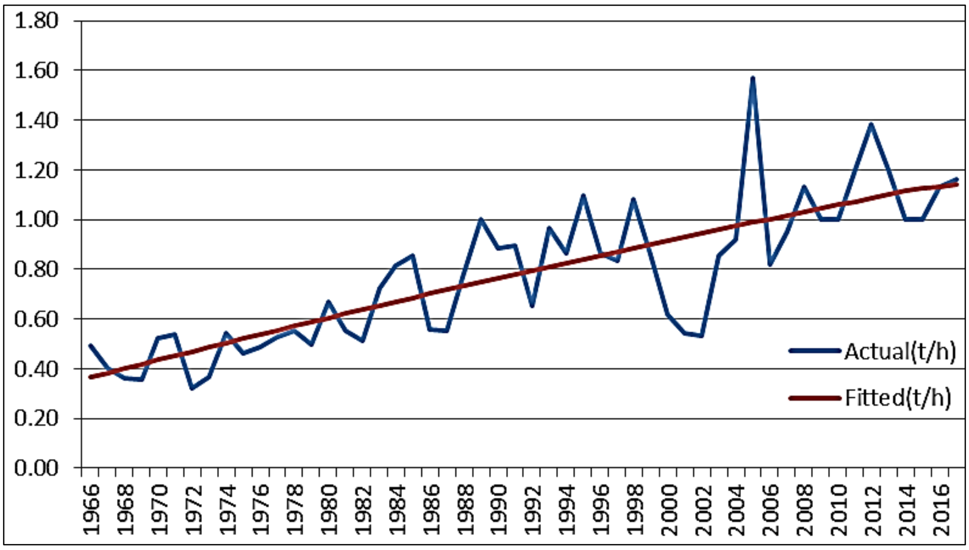

Fig 2: Actual yield vs. Fitted yield (Gompertz model)

Fig. 2 shows the graph of actual v/s fitted yield for Gompertz model where time in year (t) in x-axis and yield (t/ha) in y-axis.

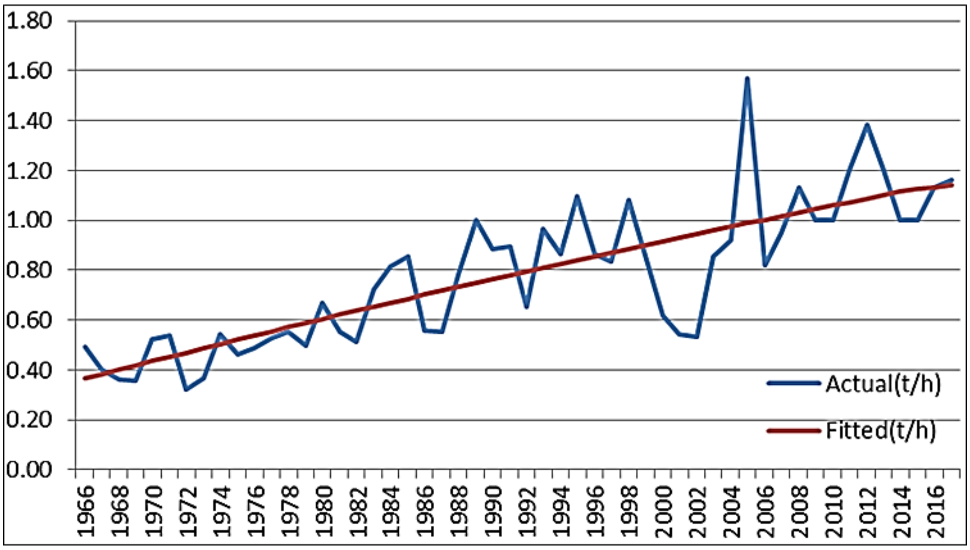

Fig 3: Actual yield vs. Fitted yield (Logistic model)

Fig.-3 shows the graph of actual v/s fitted yield for Logistic model where time in year (t) in x-axis and yield (t/ha) in y-axis.

Examination of Residuals

Although R^2 is considered as effective statistic for adequacy of model (Kvalesh 1985), but sole reliance on it may not reveal important characteristics. So, one should always go through residuals or error examination while fitting any model.

The residuals are the unexplained part of information which can expressed as difference between the observed and predicted values of fitted model. Usually, there are two assumption which are made for residuals:

i) Residuals are random.
ii) Residuals are normally distributed.

So, in the way to select model it must be cleared that residuals should follow the required assumptions.

Run test and Shapiro-wilks (W) test are used for randomness and normality of residuals respectively. Given table shows the result of respective test performed on residuals.

Table 5: Randomness test for the residuals (Run test)

Models	Standard normal value	p-Value
Monomolecular	-0.560	0.965
Gompertz	-0.520	0.603
Logistic	-0.011	0.991

The table-5 shows that z values for each applied model are not significant which reveals that we accept null hypothesis and conclude that residuals are random for all three model.

Table 6: Normality test for the residuals (Shapiro-wilks test)

Models	W (Statistic)	p-Value
Monomolecular	0.936	0.008
Gompertz	0.913	0.001
Logistic	0.909	0.001

On the basis of p-value showed on above table it can be clearly concluded that residuals are normally distributed which support the assumptions made regarding residuals.

Goodness of fit

To check whether there is difference between observed data and expected data or not, goodness of fit is used. Chi-square test is employed to test the goodness of fit for non-linear models used for study. The findings of the chi-square test are shown in table-6.

Table 7: Pearson's Chi-squared test for goodness of fit for non-linear models for chickpea yield (t/ha) data in Darbhanga district (Bihar)

Model	Chi-squared value	Degree of freedom	p-value
Monomolecular	2652	2601	0.238
Gompertz	2600	2550	0.240
Logistic	2652	2601	0.238

It is found that Chi-squared value for all three nonlinear model is non-significant based on above table-7. It concludes that there is no significant difference between observed and predicted values of chickpea yield.

Fitting of model for Chickpea productivity in Darbhanga (Bihar)

All three model are compared based on different statistics for forecasting productivity of chickpea in Darbhanga district of Bihar. It is found that the Logistic model is best suited for forecasting among three model. As per run test it is cleared that residuals is following random distribution while Shapiro wilk test revealed that residuals are not normally distributed.

The parameter estimates for all the models for estimating the productivity of chickpea in Darbhanga district of Bihar, are given below.

For the Monomolecular model

$\hat{Y}$ = k- (k - y0)* exp. (- r *t).

$\hat{Y}$ =3.5054 -(3.5054-0.3438) *exp. (-0.0057*t)

For the Gompertz model

$\hat{Y}$ = k*exp(log(y0/k)*exp(-r *t))

$\hat{Y}$ =1.6235*exp(log(0.3574/1.6235)*exp(-0.0281*t)).

For the Logistic model

$\hat{Y}$= k/(1+(k/y0-1)* exp(-r *t)).

$\hat{Y}$ =1.3613/(1+(1.3613/0.3684-1) *exp(-0.0497*t))

Table 8: Forecasting yield (t/ha) using Logistic model for Darbhanga (Bihar)

Year	Logistic model		
	Actual yield(t/h)	Forecasted yield(t/h)	% FE
2016	1.13	1.12	0.89
2017	1.16	1.12	3.57
2018	--------	1.14	-------------
2019	------	1.15	-------------
2020	------	1.16	-----------
2021	....	1.17	------------
2022	....	1.17	-------------
2023	....	1.18	-------------

In table-8. Chickpea yield (t/ha) for Darbhanga district (Bihar) are forecasted for upcoming years i.e. up to 2023 i.e. 1.18 t/h.

As computed values of Z-statistics for each model is not significant hence null hypothesis of randomness of errors is not rejected at 5% level of significance. Even W-statistics of Shapiro wilk's test is not significant for each model. Hence, it can be concluded that residuals follow normality.

Conclusion

For Darbhanga district (Bihar) three nonlinear models viz., Monomolecular, Logistic and Gompertz model, are fitted for chickpea yield. These models are fitted to data of Chickpea yield with the help of statistical software-R. To obtain best fitted model among three, 6 different statistics viz., R^2, RSS, MSE, MAE, RMSE, MAPE are compared. Based on comparison of 6 statistics and obtained chi-square value, Logistic model is found best fitted model for Darbhanga district (Bihar) and by Logistic model, Chickpea yield is forecasted for upcoming years up to 2023 (1.18 t/h).

References

1. Draper NR, Smith H. Applied Regression rd Analysis. 3 Edn. John Wiley & Sons. U.S.A.; c1998. p. 543-49.
2. Karadavut U, Palta Ç, Kökten K, Bakoğlu A. Comparative study on some non-linear growth models for

describing leaf growth of maize. International Journal of Agriculture & Biology. 2010;12:227-230.
3. Kvålseth TO. Cautionary note about R 2. The American Statistician. 1985;39(4):279-285.
4. Pal S, Mazumdar D. Forecasting groundnut production of India using nonlinear growth models. Journal Crop and Weed. 2015;11:67-70.
5. Ratkowsky DA. Handbook of nonlinear regression models, parcel Dekker. Inc., New York and Basel; c1990.
6. Seber GA, Wild CJ. Nonlinear regression. Hoboken. New Jersey: John Wiley & Sons. 2003;62(63):12-38.
7. Sundar Rajan M, Palanivel M. Estimation of non-linear models for fitting growth curves of cotton in India. International Journal of Statistics and Applied Mathematics. 2017;2(4):86-93.