Indian Journal of Agricultural Sciences **82** (12): 1051–4, December 2012

Nonlinear regression techniques for analysis of onion (*Allium cepa*) production in India

SANJEEV PANWAR[1], ANIL KUMAR[2], K N SINGH[3], MD SAMIR FAROOQI[4] and ABHISHEK RATHORE[5]

Indian Agricultural Statistics Research Institute, Pusa, New Delhi 110 012

Received: 16 May 2012; Revised accepted: 2 August 2012

ABSTRACT

India occupies second position in the production of onion in the world. Keeping in view the importance of this crop, the present study has been undertaken to find out the growth in production of onion and discusses the application of nonlinear models, viz. Gompertz, Logistic, MMF, Richards and Weibull models, which measure the growth. Time series data on onion production in major growing states; viz. Andhra Pradesh, Gujarat, Karnataka, Maharashtra, Uttar Pradesh, and all India from 1990–91 to 2009–10 has been utilized for the present study. From a realistic point of view, the relationships among variables in agricultural and horticultural sciences are non-linear in nature. Non-linear models are very popularly used to estimate the trend in various fields such as population studies and animal growth where growth is not symmetrical about the point of inflection. The results showed that Logistic and Gompertz models faired marginally better than Weibull and MMF models.

Key words: Forecasting, Goodness of fit, Growth models, Optimization, Run test, Theil statistic

India occupies first position in the production of cauliflower, second in onion and third in cabbage. Keeping in view the importance of onion crop, both from production as well as consumption point of view, the present investigation has been carried out. The objectives of the study are to examine the growth in the production of onion crop and to fit various models and find out the best fit model.

Statistical modeling essentially consists of constructing a model, represented by a set of equations to describe the input-output relationship among the variables of interest. From a realistic point of view, such relationships among variables in agricultural and horticultural sciences are non-linear in nature. In such models, a unit increase in the value of independent variable(s) may not result in an equivalent unit increase in the dependent variables. Using the best fitted model for each state, forecasting is made for five periods in advance (Out-of-Sample forecast)

MATERIALS AND METHODS

For the present study, data on onion production in India from 1990–91 to 2009–10 are utilized for the study. Data were collected from secondary sources, viz. National Horticulture Board (NHB), Department of Horticulture.

In the present study, five non-linear models have been selected such as Gompertz, Logistic, MMF, Richards and Weibull models are employed. This might help in improving the existing models and for better forecasting of the production of crop. This also throws light on the likely instability in the forecast.

Gompertz Model

$$Y = a \exp(-\exp(b - cX)) + e \qquad (1)$$

Logistic Model

$$Y = \frac{a}{1 + \exp(b - cX)} + e \qquad (2)$$

Richards Model

$$Y = \frac{a}{[1 + \exp(b - cX)]^{1/d}} + e \qquad (3)$$

Morgan-Mercer-Flodin Model

$$Y = \frac{bc + aX^d}{c + X^d} + e \qquad (4)$$

Weibull Model

$$Y = a - b \exp(-cX^d) + e \qquad (5)$$

There is a need to *goodness of fit* statistics which will

[1]Scientist(SS) (e mail: scientist1775@yahoo.co.in), [2]Senior Scientist, [3]Principal Scientist, [4]Scientist(SS), IASRI, Library Avenue, Pusa, New Delhi 110 012; [5]Senior Scientist, ICRISAT, Hyderabad

help us to test the goodness of fit vis-à-vis comparison among different competing models. Here, above mentioned models are compared using statistics such as R^2, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Theil Statistic.

Optimization Method: Levenberg-Marquardt method is utilized for fitting all models in this study. To start the iterative procedure, initial estimates of the parameters of the models are required. Many sets of initial values were tried to ensure global convergence. The iterative procedure was stopped when the reduction between successive residual sums of squares was found to be negligibly small.

Three main methods are available to obtain estimates of the parameters of a nonlinear regression model. These are (i) Linearization (or Taylor Series) method, (ii) Steepest-descent method and (iii) Levenberg-Marquardt's technique. However, in all these methods the following steps are carried out.

Step 1 Starting with a good initial guess of the unknown parameters; a sequence of q's which hopefully converge to q is computed.

Step 2 Error sum of squares expressed as

$$S(\theta) = \sum_{t=1}^{N} [Y_r - F_i(\theta)]^2$$ is minimized with respect to the current value of θ. The new estimates are obtained.

Step 3 By feeding the recently obtained estimates as the initial guess for the next iteration, objective function S(θ) is minimized again to obtain fresh estimates. This procedure is continued till the successive iteration yielded parameter estimate value is close to each other.

Diagnosis of Residuals: Having fitted the nonlinear statistical model, the next important step is to see whether the assumptions made regarding the error term are valid or not. This is being done by examining residuals. The residuals are the difference between the observed values X(t) and the fitted or predicted value values $\hat{X}(t)$. It is evident from the definition that residuals are the differences between what is actually observed, and what is predicted by the fitted model, i.e. the amount which is unexplained by the model. If the fitted model is correct, the residuals should exhibit tendencies that tend to confirm the assumptions made, or at least, should not exhibit denial of the assumptions. Test for randomness of residuals 'Run' test and for normality of residuals 'Shapiro-Wilks' (W) test is used for this study.

Forecasting: After identifying the best model through examination of residuals and comparison among the models, the Out-of-sample forecast, using best nonlinear model for each state, is made for five periods in advance that is from 2011 to 2015.

RESULTS AND DISCUSSION

Gompertz, Logistic and Weibull model fits to the Andhra Pradesh onion production data, estimated parameters and statistics are shown in Table 1. There is no evidence to reject the assumptions of randomness and normality of the residuals for these models. It is found that there is significant difference in the value of MSE of all five models, viz. Logistic (24.32), Weibull (3642.19) and Gompertz (4219). Weibull is ignored because parameters of both model about zero. Hence low MSE show that Logistic is better fit to describe the Andhra Pradesh onion production data well. Theil statistic proves that the selected forecasting model, i e Logistic model (Fig 1) is better than the Naive method/OSAF.

$$Y = \frac{89.70}{1 + \exp(1.75 - 0.14X)} + e$$

Gompertz, Logistic, MMF and Weibull model fits well

Table 1 Comparison of different models fitted for data on All India onion production

Parameters/ statistics	(AP) Logistic	(Gujarat) Gompertz	(Karnataka) Logistic	(Maharashtra) Gompertz	(UP) Logistic	(All India) Logistic
a	89.71	177.0	629.90	754.230	64.00	771.0
b	1.7550	5.4160	1.45	2.26	0.25	9.87
c	0.1434	0.4745	0.18	0.01	0.002	0.26
Examination of residuals						
Runs statistic	–1.21	–0.40	–2.03	–2.03	0.41	-2.04
Shapiro-Wilk statistics	0.98	0.72	0.75	0.91	0.99	0.91
Goodness of fit statistics						
R^2	0.92	0.73	0.66	0.78	0.73	0.89
MSE	24.32	1 238.04	31 677	2 122.37	20.33	15.57
RMSE	4.93	35.08	177.98	46.12	4.50	3.94
MAE	3.25	27.77	111.67	30.53	3.39	3.55
MAPE	8.62	5.18	26	16.02	12.38	8.86
Theil statistics	0.73	0.06	0.41	6.22	0.76	0.06

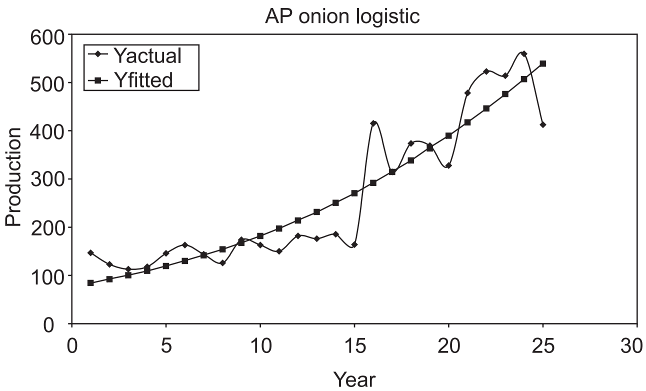

Fig 1 Plot of fitted Logistic, and Actual data of Andhra Pradesh onion Production.

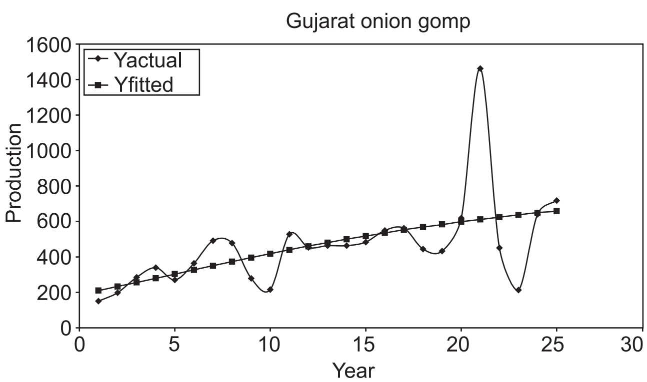

Fig 2 Plot of fitted Gompertz, and Actual data of Gujarat onion Production.

to the Gujarat onion production data (Table 1). Run test indicate randomness of the residuals and Shapiro-Wilk statistics are low, about 0.72, that shows, errors are not independently distributed which poses considerable doubt on the model. But we find that there is much difference in the value of MSE of Gujarat states all five models, viz. Gompertz (1238), Logistic (49228), MMF (4219) and Weibull (55153). So Gompertz or Logistic models fit Gujarat onion production data well. Estimated Theil statistics (0.56) show that Gompertz or model (Fig 2) is better than the Naive method.

$$Y = 177 \exp(-\exp(5.41-(0.47\ X))) + e$$

Parameter estimates of Gompertz, Logistic and Weibull for Karnataka onion production data, parameters converged, is presented in Table 1. Run test indicate that all are less than 1.96, it shows residuals are normally distributed and Shapiro–Wilk statistics are low which shows not normally distributed and poses doubt to fitting selected models. But Logistic model (Fig 3) have low MSE 31677 fits Karnataka onion production data well and it is found to be better than Naive method as proved by Theil statistic and equation of fitted logistic model is,

$$Y = \frac{695.10}{1 + \exp(0.57 - 0.11X)} + e$$

Five non-linear models were tried on the Maharashtra

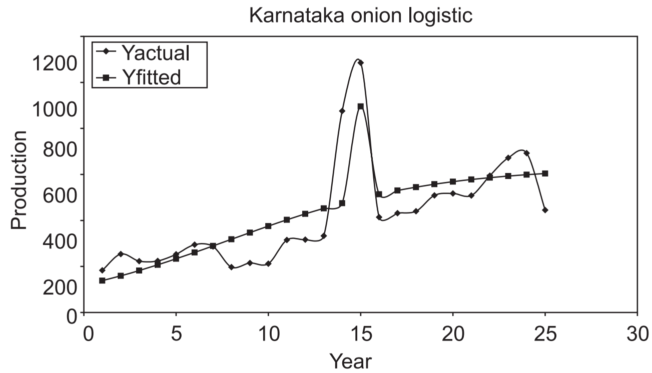

Fig 3 Plot of fitted Logistic, and Actual data of Karnataka onion Production.

onion production data. The estimation details of four models viz. Gompertz, Logistic and Weibull, were given in the Table 1. Two models, viz. Richards and MMF did not converge. The Run test indicates randomness of residuals at 5% level of significance and the Shapiro-Wilk statistics shows that the residuals are normally distributed for all models except Weibull. The best-fitted models based on estimated MSE, are in the order: Logistic (55362), Gompertz (2127.37) and Weibull (57966) model. The Gompertz model (Fig 4) is found to be better fit to describe Maharashtra onion area data and they are found to be slightly better than Naive method (Theil statistics).

$$Y = 754.23 \exp(-\exp(2.26 - (0.01\ X))) + e$$

For Uttar Pradesh onion production data, three parameters models, viz. Gompertz and Logistic models converged and estimated parameters and statistics shown in Table 1. None of the four-parameter model is converging. The best-fitted models are in the order: Gompertz (797-MSE) and Logistic (20.33-MSE). The Logistic model (Fig 5) is found to be better fit to describe Uttar Pradesh onion production data well and found slightly better than Naive method as shown by Theil statistic.

$$Y = \frac{64.012}{1 + \exp(0.25 - 0.002)} + e$$

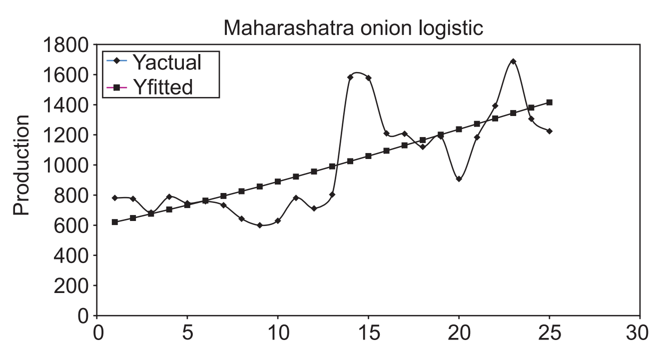

Fig 4 Plot of fitted Gmpertz, and Actual data of Maharashtra onion Production.

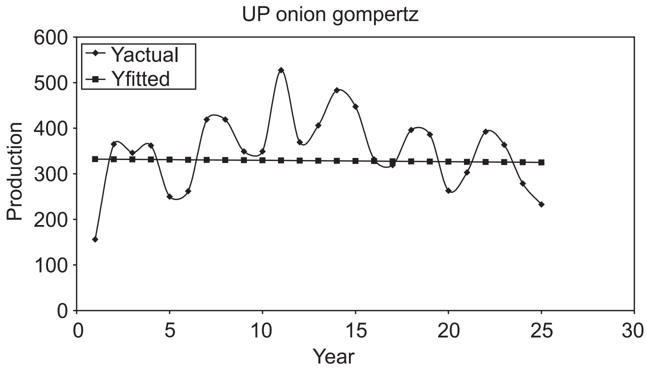

Fig. 5 Plot of fitted Logistic, and Actual data of Uttar Pradesh onion Production.

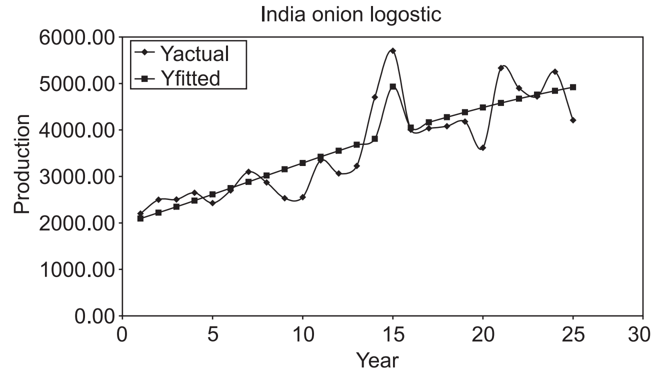

Fig. 6 Plot of fitted Logistic, and Actual data of all India onion Production.

Table 2 Out-of-Sample forecast, using Best fitted nonlinear model for selected states of onion production (million tonnes)

State	Best model	2011	2012	2013	2014	2015
Andhra Pradesh	Logistic	698.27	704.1	721.11	749.32	788.67
Gujarat	Gompertz	1700.5	1 760.5	1 808.2	1 894.9	1 915.1
Karnataka	Logistic	2 350.6	2 402.5	2 509.5	2 584.01	2 617.4
Maharashtra	Gompertz	3 002.5	3 069.2	3 150.3	3 205.2	3 561.6
Uttar Pradesh	Logistic	389.2	397.6	410.6	489.8	523.5
India	Logistic	12 990.4	13 001.1	13 500.6	14 060.4	15 100.6

Only two non-linear models, viz. Gompertz and Logistic models fit for all India onion production data, estimated parameters and statistics were presented in Table-1. Run test and Shapiro-Wilk test there is no evidence to reject the assumptions of randomness and normality of the residuals for these models they are independently and normally distributed. It is found that there is significant difference in the value of MSE of all models is in orders: Logistic (350639) and Gompertz (354274) so Logistic model is better than all selected models because in Weibull model one parameter is near to zero but estimated Theil statistic prove that Naive method is not better than the best selected Logistic model (Fig 6).

$$Y = \frac{5703}{1 + \exp(0.690 - 0.9X)} + e$$

It was evident, from the analysis, production of onion is fluctuating and Gompertz, Logistic and Weibull nonlinear models are suitable to fit. Models having least MSE are found better fit over all selected models. The initialization of parameters is done using Ratkowsky method and the results are compared using statistics such as 'R^2', MSE, RMSE, MAE, MAPE, and Theil statistic. On the basis of goodness of fit statistics it can be observed that logistic is best fitted models for Andhra Pradesh, Karnataka and UP states and Gompertz model is better than other models for Gujarat and Maharashtra states. For all India onion production data it is found that Logistic model is best fitted model. Out-of-Sample forecast, using Best Fitted Nonlinear Model for selected states of onion production (MT) (Table 2). The results showed that Logistic and Gompertz fitted faired marginally better than Weibull and MMF models.

REFERENCES

Draper N R and Smith H. 1998. *Applied Regression Analysis,* 3rd ed. Wiley, NY.

Gallant A R. 1987. *Nonlinear Statistical Models*. Wiley, New York.

Ralston M L and Jennrich R I. 1978. DUD, a derivative-free algorithm for nonlinear least squares. *Technometrics* **20**: 7–14.

Seber G A F and Wild C J. 1989. *Nonlinear Regression*. Wiley, New York.

Prajneshu. 1998. A nonlinear statistical model for aphid population growth. *Journal of Indian Society of Agricultural Statistics.*

Chandran K P and Prajneshu. 2004. Computation of growth rates in agriculture: Nonparametric regression approach. *Journal of Indian Society of Agricultural Statistics* **57**: 382–92.