Comparison of non-linear models to describe growth of cotton

Sundar Rajan M and Palanivel M

Abstract
The objective of this study were to compare the goodness of fit of six non-linear growth model Monomolecular, Logistic, Gompertz, Richards, Quadratic and Reciprocal growth in India Cotton Area, Production and Productivity data collected during 1980-2013. The models parameters (a, b and c), Coefficient of Determination ($R^2$), Residual Sum of Square (RSS) and Root Mean Square Error (RMSE) results. The ‘Run test’ and ‘Shapiro-Wilk’ test were also used to test the compliance of the error term to the underlying assumptions. Among the six models, under study predicted closely the observed values of top area, production and productivity in the selected nonlinear growth model has been selected for its accuracy of fit according to the highest $R^2$; Lower Residual Sum of Square and Mean Square Error.

Keywords: Cotton (Gossypium hirsutum), Area, Production, Productivity and Non-linear model

Introduction
Cotton is an important cash crop and India cultivates the highest acreage in the world. It provides the basic raw material (cotton fibre) to the cotton textile industry. Cotton also known as ‘White Gold’ dominates India’s cash crops, and makes up 65 per cent of the raw material requirements of the Indian textile. Cotton play a vital role in economy in many cotton producing nations. India is the major cotton producing nation with 27 million bales production which is approximately account for 24% of global cotton production (2014). India has tripled cotton production form 13 million bales to 40 million bales and has doubled its market share of global cotton production form 12% in 2002 to 25% in 2014, representing a quarter of total global cotton production. The major cotton growing states in India are Punjab, Haryana, Rajasthan, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Tamil Nadu and Karnataka.

The main objective of this study was to focus on past and future trends of cotton area production and productivity in India by using appropriate nonlinear growth models analysis model. Many nonlinear theoretical/mechanistic models (Logistic, Gompertz, Monomolecular, Richards, Reciprocal and Quadratic) rather than empirical models. Growth rates analyses are widely employed to study the long-term trends in various agricultural crops (Panse, 1964) [14], Kaloala et al. (1995) reported that for area the second degree (Quadratic) polynomial; for production and productivity the first degree polynomial model were found suitable to fit the trend of tobacco crop grown in Gujrat state for the period of 1950-51 to 1990-91. In model discussed by Gan-Yan tai et al. (1996) [5] worked on evaluation of selected nonlinear regression models in quantifying seedling emergence rate of spring wheat. Prajneshu and Das (1998) [17] carried out a detailed study dealing with modelling of wheat production data at State level in post-Green revolution era. Specifically, several mechanistic nonlinear growth models, viz. monomolecular, logistic, Gompertz, mixed-influence and Richards were applied using Levenberg Marquardt procedure. The heartening feature of a mechanistic model is that the parameters have specific agricultural/ biological interpretation and provide insight into the underlying mechanism. Six major wheat growing States, viz. Punjab, Haryana, Uttar Pradesh, Madhya Pradesh, Rajasthan and Bihar were considered. For each of these States, it was found that logistic model has performed the best. These mechanistic models were employed in several other studies (Seber and Wild, 1989; Iquebal and Sarika, 2013; Chandran K. P and Prajneshu, 2004; Rajarathinam et al., 2010) [24]. Sheila Zambello de Pinho, Ladia Raquelde

Keeping above points in view, the present study was aimed to develop appropriate nonlinear statistical growth models with a view to provide analytical approach to describe the cotton area, production and productivity trends in India. Based on performance of models fits, three best nonlinear growth models were chosen for future projection of cotton crop trends in India. The Growth model approach is simple to understand and apply for projection of future trends, and is competent of curve fitting a whole range of different models. In this trend investigation studies on cotton will provide insight to policy makers, stake holders and researchers in policy making strategies on sustainable cotton production in the future.

Material and Methods

The present study has been conducted on area, production and productivity of cotton crop in India for the period 1980-2013. The cotton crop secondary data were collected from Cotton Corporation of India (CCI). In his study we have compared different models viz Logistic, Gompertz, Monomolecular, Richards, Reciprocal and Quadratic model for estimating the growth of cotton crop on area, production, and productivity to find the best fit using the statistical methods such as highest $R^2$, Residual sum of square and Root mean square. Subsequently forecasted the cotton growth using the same statistical ncss software version 10. In this section we describe different growth models that can describe nonlinear curve fitting change patterns. Non -linear estimation is a general fitting procedure that will estimate any relationship between a dependent (or response variable), and a list of independent variables. This work will help for better forecast, and in developing nonlinear growth models for area, production and productivity of cotton for India. This modelling effort will help on one hand, to understand the past performance and on the other to forecast future possibility.

Linear and Non-linear models

A linear model is one which is directly proportional to input. In such a model, all the parameters appear linearly. In contrast to this, nonlinear model is one in which at least one of the parameters appear nonlinearly. For example, following equations:

\[ Y(t) = A + Bt \] ........................ (i)

\[ Y(t) = A + Bt + Ct^2 \] ........................ (ii)

represents a linear model, whereas equation:

\[ Y(t) = Y(0) \exp(-At) \] ........................ (iii)

Represents a nonlinear model. A consequence of nonlinearity is that it has curved solution locus, whereas linear model has straight line solution locus. We consider two types of nonlinear model given by following equation, respectively:

\[ Y(t) = \exp(A + Bt^2) \] ........................ (iv)

And \[ Y(t) = \exp(-Bt) - \exp(-At) \] ........................ (v)

Although, both these models are nonlinear as the parameters A and B appear nonlinearly but they are of essentially different characters. Equ.(iv) can be transformed, by taking natural algorithm, into the form:
\[ \ln Y(t) = A + Br^2 \]

Which is linear in parameters. The model given in Equ(iv) is called intrinsically linear since it can be transformed into linear form. Equ(iii) however, cannot be converted into a form which is linear in parameters. Such a model is said to be intrinsically nonlinear (Draper and Smith, 1981).

**Some important nonlinear growth models**

A mathematical model is an equation or a set of equations which represents the behaviour of a system (France and Thornley, 1984). We now discuss briefly some well-known nonlinear growth models. The following nonlinear growth models (Seber and Wild, 1989) (24) of Logistic, Gompertz, Monomolecular, Richards, Reciprocal and Quadratic etc., have been tried.

**Monomolecular**

This model describe the progress of a growth situation in which it is believed that the rate of growth at any time is proportional to the resources yet to be achieved by the equation. This model is represented by the equation:

\[ Y(t) = \lambda - (\lambda - \beta) \exp(-\alpha t) \]

Where symbols have their usual meaning.

**Logistic Model**

The parameters of this model have a simple physical interpretation. The mathematical form of the model is given by.

\[ Y_i = c \exp(1+b \exp(-at)) + e, b = c \exp(y(0) - 1) \]

Where \( Y_i \) denotes the cotton area during the time \( t \), \( \dot{a} \) denotes the intrinsic growth rate, \( \dot{b} \) denotes the different functions of the initial value \( Y(0) \), \( c \) denotes the carrying capacity of the model, and \( e \) is the error term.

**Gompertz Model**

This model has sigmoid type of behavior and is found quite useful in the biological work. However, unlike logistics model, this is not symmetric about its point of inflexion. This model is given by.

\[ Y_i = c \exp(-b \exp(-at)) + e, b = c \exp[y(0)] \]

**Quadratic Model**

The quadratic or second-order polynomial model results in the familiar parabola.

\[ Y_i = b_0 + b_1 t + b_2 t^2 \]

\( Y \) and \( t \) is area and time period respectively, \( b_0 + b_1 \) is constants to be an estimated. The quadratic model can be used to model a series that takes off” or a series that dampens.

**Richards Model**

It is a four parameter growth model. It is proposed by Richards (1959) and is represented by:

\[ y(t) = \lambda \left[ 1 + \beta \exp(-\alpha t) \right]^{1/ \delta} \]

The upper sign within the brackets is applicable when \( \delta \) is positive and lower sign when \( \alpha \) lies in the range \(-1 < \delta < 0\). Richards model is a generalization of logistic (when \( \delta = 0 \)) Gompertz (when \( \delta = 0 \)) and monomolecular (when \( \delta = -1 \)). In such cases, the solution is to either study more advanced single-species growth models, such as Richard’s model and mixed influence model, or apply non-parametric regression procedures (Chandran and Prajeshu, 2004) [3].

**Reciprocal Model**

This model, known as the reciprocal or Shinozaki and Kira model, is mentioned in Ratkowsky (1989, page 89) and Seber (1989, page 362) [24], and is given by

\[ y = 1 / (a + bx) \]

So we regress the reciprocal of the dependent variable upon the untransformed independent variables.

The data were analysed using the non-linear as well as the nonparametric regression models and the conclusion is given based on the best model to study the trends and growth rate of cotton crop area, production and productivity in India.
Model Evaluation

The following measure of goodness of fit vis-a-vis comparison among different competing model developed. The goodness of fit is examined by using the co-efficient of determination\(^2\) appear in the literature. One of the most frequent mistakes occurs when the fits of a linear and a non-linear model are compared by using the same \( R^2 \) expression but different variables. The \( R^2 \) is generally interpreted as a measure of goodness of fit of even the original nonlinear model, which is incorrect. Scott and Wild (1991) have given a real example where two models are identical for all practical purpose and yet have very different values of \( R^2 \) calculated on the transformed scales. Kvalseth (1985) has emphasized that, although \( R^2 \) given by

\[
R^2 = 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2
\]

This is most appropriate for nonlinear statistical model. It would be used as the coefficient of determination for fit. The potential range of values of this \( R^2 \) is well defined with end points corresponding to perfect fit and complete lack of fit, such as \( 0 < R^2 < 1 \). where \( R^2 = 1 \) corresponds to perfect fit and \( R^2 \geq 0 \) for any reasonable model specification. For nonlinear models its value can be negative, if the selected model fit worse than the mean.

Residual sum of square (RSS)

The RSS is defined as

\[
RSS = \sum_{i=1}^{n} (y_i - (\alpha + \beta_i))^2
\]

Residual sum of square (RSS) is also known as the sum of square residuals (SSR) or sum of squared error (SSE) of prediction. It is an amount of the difference between data and the estimated model. Where \( \alpha \) and \( \beta_i \) is the estimated value of the slope coefficient.

\[
\text{Mean square error (MSE)} = \frac{\sum (y_i - \hat{y}_i)^2}{n - p}
\]

The smaller the value of MSE the better is the model.

Examination of residuals

As a measure of goodness of fit the residual analysis of the models is carried out. The main assumptions made in the models are

(i) Error are random
(ii) Errors are normally distributed.

To test whether the residuals are normally distributed Shapiro-Wilk test has been performed. If the fitted model is correct, the residuals should exhibit tendencies that tend to confirm or at least should not exhibit a denial of the assumptions. The procedure for carrying out the Shapiro-Wilk test has been explained in brief.

Shapiro-Wilk test (Test for Normality)

Commonly used to test the normality of the given observation. The Null hypothesis is given by,

\( H_0 : \) The residuals are normally distributed against
\( H_1 : \) The residuals are not normally distributed

The test statistic is given by

\[
W = \frac{S^2}{b}
\]

Where,

\[
S^2 = \sum a(k) \left\{x_{(9n+1-k)} - x_{(k)}\right\}
\]

\[
b = \sum (x_i - \bar{x})^2
\]

In the above the parameter \( k \) takes the values \( k = \{1, 2, \ldots, n/2\} \) when \( n \) is even, \( k = \{1, 2, \ldots, n - 1/2\} \) when \( n \) is odd and \( x_{(k)} \) is the \( k \)th order statistic of the set of residuals. \( H_0 \) is rejected at level \( \alpha \) if \( W \) is less than the tabulated value.

Run test (Test for Randomness)

The run test is based on number of runs (r), where a run is defined as sequence of symbols of one kind separated by symbols of another kind. The residuals are replaced by ‘+’ or ‘-‘accordingly they are positive or negative. The null hypothesis \( H_0 \) is given as

\( H_0 : \) The residuals are random against

~89~
\( H_0: \) Residuals are not random.

Let ‘m’ be the number of positive signs, ‘n’ be the number of runs.

The mean is given as \( \mu = \frac{2mn}{(m+n)+1} \) and Variance is given as \( \sigma^2 = \frac{2mn}{(m+n)[(m+n)-1]}. \)

For large sample, the test statistic is
\[ Z = \frac{(r - h - \mu)}{\sigma} \sim N(0,1) \]

Where, \( H^0 \) is rejected at level \( \alpha \) if \( \frac{|z|}{\alpha} > Z_{\alpha/2} \)

Where, \( Z_{\alpha} = P(Z > Z_{\alpha}) = \alpha \).

**Result and Discussion**

Results of the study shows that among the six models considered for India, converged (Table 1, 2 and 3). Further, the table reveals that difference between each model could explain area, production and productivity of cotton crop data, based on the goodness of fit statistics presented same tables given below.

**Trends in Area based on Non-linear Models**

The result presented (Table 1) for the area under the cultivation of cotton crop revealed that among the different non-linear models fitted, the highest R-square (83%) was observed in case of Richards function with lowest value of SSE(566.43) and RMS (9.60) in comparison to that of other non-linear models. Based on the performance of these models, it found that Richards’s model is found to be best fit. All the parameters were found to be significant. Statistical tests for normality are more precise since actual probabilities are calculated. The above (Table 4) presents the results from the tests of normality are run. However the Shapiro-Wilks, test value 0.97 with p-value 0.17 (test for normality) was found to be non-significant the residuals due to this model. Moreover, run-test statistic (test for random

**Table 1:** The results on non-linear growth models for cotton cultivated area

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Quadratic</th>
<th>Reciprocal</th>
<th>Monomolecular</th>
<th>Logistic</th>
<th>Gompertz</th>
<th>Richards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimate</td>
<td>Asym Error</td>
<td>Parameter Estimate</td>
<td>Asym Error</td>
<td>Parameter Estimate</td>
<td>Asym Error</td>
<td>Parameter Estimate</td>
</tr>
<tr>
<td>a</td>
<td>41172.96</td>
<td>5280.23</td>
<td>12.30</td>
<td>0.07</td>
<td>1549369.66</td>
<td>149790653.11</td>
</tr>
<tr>
<td>b</td>
<td>-41.89</td>
<td>5.33</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>8996.12</td>
</tr>
<tr>
<td>c</td>
<td>0.01</td>
<td>0.00</td>
<td>1950.21</td>
<td>5.03</td>
<td>0.00</td>
<td>2506.12</td>
</tr>
<tr>
<td>d</td>
<td>2072.20</td>
<td>16438681.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Goodness of fit**

| R-Square | 0.82 | 0.82 | 0.64 | 0.12 | 0.82 | 0.83 |
| SSE | 598.20 | 628.18 | 1223.23 | 3000.78 | 625.53 | 366.43 |
| RMS | 9.97 | 10.3 | 20.39 | 30.01 | 10.43 | 9.60 |
value -1.39 which gives p=0.16 (Table 4) were found to be non-significant indicating that the residuals were correlated due to this model selection criteria. The Richards model actual and predicted trend value fitted shown fig.1, among the non-linear models, the richards model was found suitable to fit the trends in cultivated area of cotton crop.

\[ Y = (2.93) \exp\((-0.08) ((C_1) - (0.0075))) \ (r^2 = 83) \]

**Trends in Production based on Non-linear models**

It has observed of the study shows that among the six models considered for cotton production, the parameters of only logistic, quadratic and reciprocal were converged (Table 2). Further, the table reveals that Richards model could explain production data better than quadratic and gompertz model, based on the goodness of fit statistics. The highest R-square (83%) was observed in the model of Gompertz function with lowest value of SSE (566.43) and RMS (9.60) in comparison to that of other non-linear models. All the parameters were found to be significant. However, the above (Table 4) presents the results from the tests of normality are run. The Shapiro-Wilk test (test for normality) values were significant indicating that the residuals due to this model were independently normally distributed. However, the run-test (test for randomness) value were found to be significant indicating that the residuals were correlated. Among the non-linear models, the Richards model was found suitable to fit the trends shown (fig.2) in cultivated production of cotton crop estimated model,

\[ Y = (709.78) + (0.67) ((C_1) + (0.0001) ((C_1)^2) \ (2) \ (r^2 = 83) \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Quadratic</th>
<th>Reciprocal</th>
<th>Monomolecular</th>
<th>Logistic</th>
<th>Gompertz</th>
<th>Richards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Asymptotic</td>
<td>Parameter</td>
<td>Asymptotic</td>
<td>Parameter</td>
<td>Asymptotic</td>
<td>Parameter</td>
</tr>
<tr>
<td>Error</td>
<td>Error</td>
<td>Error</td>
<td>Error</td>
<td>Error</td>
<td>Error</td>
<td>Error</td>
</tr>
<tr>
<td>a</td>
<td>41172.96</td>
<td>5280.23</td>
<td>12.30</td>
<td>0.07</td>
<td>1549366.66</td>
<td>1497930653.11</td>
</tr>
<tr>
<td>b</td>
<td>-41.89</td>
<td>5.33</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>8996.12</td>
</tr>
<tr>
<td>c</td>
<td>0.01</td>
<td>0.00</td>
<td>1950.21</td>
<td>5.03</td>
<td>0.00</td>
<td>12861.2</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Goodness of fit**

<table>
<thead>
<tr>
<th>R-Square</th>
<th>SSE</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.82</td>
<td>598.20</td>
<td>9.97</td>
</tr>
<tr>
<td>0.82</td>
<td>628.18</td>
<td>10.3</td>
</tr>
<tr>
<td>0.64</td>
<td>1223.23</td>
<td>20.39</td>
</tr>
</tbody>
</table>

**Table 2:** The results on non-linear growth models for cotton production

![Fig 2: Actual and predicted for cotton production](image)

**Trends in Productivity based on Non-linear Models**

Furthermore, they are better able to represent the entirety of the developmental process different models fitted and (Table 3) the maximum \( R^2 \) (0.92) was observed in the model of reciprocal function with minimum value of SSE(166179) and RMS(4053) in comparison to that of other non-linear models. The above (Table 4) presents the results from the tests of normality are run. However the Shapiro-Wilk test statistic value 0.91 with p=0.00 was found to be significant indicating that the residuals due to this model were independently normally distributed. Finally run test (randomness) value of -5.72 which gives p=0.00 values were found to be non-significant indicating that the residuals fulfilled model selection criteria. All the estimated value of the parameter in this model were found to be within the 95% confidence interval indicating that the parameter were significant at 5% level of significance. Among the cotton nonlinear models fitted to the productivity of cotton crop the following reciprocal model was found to be the selected to fit shown as (fig 3) the estimated model,

\[ Y = 1 / ((0.173 + (-8.570 - 0.05) \ (C_1) \ (2)) \ (R^2 = 86) \]
Table 3: The results on non-linear growth models for cotton productivity

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Quadratic</th>
<th>Reciprocal</th>
<th>Monomolecular</th>
<th>Logistic</th>
<th>Gompertz</th>
<th>Richards</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter Estimate</td>
<td>Asymptotic Error</td>
<td>Parameter Estimate</td>
<td>Asymptotic Error</td>
<td>Parameter Estimate</td>
<td>Asymptotic Error</td>
</tr>
<tr>
<td>a</td>
<td>495523.83</td>
<td>69030.39</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>b</td>
<td>-505.35</td>
<td>69.7</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>c</td>
<td>0.13</td>
<td>0.02</td>
<td>1943.11</td>
<td>6.01</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td>2099.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Goodness of fit

<table>
<thead>
<tr>
<th>R-Square</th>
<th>SSE</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87</td>
<td>102240</td>
<td>1704</td>
</tr>
<tr>
<td>0.87</td>
<td>94258</td>
<td>1545</td>
</tr>
</tbody>
</table>

Table 4: Selected model using for residual form the fitting normality test

<table>
<thead>
<tr>
<th>Normality Test</th>
<th>Residual from the fitted model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area</td>
</tr>
<tr>
<td></td>
<td>Statistic</td>
</tr>
<tr>
<td>Shapiro-Wilk</td>
<td>0.97</td>
</tr>
<tr>
<td>Randomness test</td>
<td>Run test</td>
</tr>
</tbody>
</table>

Table 5: Forecasting Area, Production and Productivity

<table>
<thead>
<tr>
<th>Year</th>
<th>Area (Million hec)</th>
<th>Production (Million bales)</th>
<th>Productivity (Kg/he)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>10.34</td>
<td>29.52</td>
<td>643</td>
</tr>
<tr>
<td>2014</td>
<td>10.47</td>
<td>30.89</td>
<td>727</td>
</tr>
<tr>
<td>2015</td>
<td>10.60</td>
<td>32.32</td>
<td>837</td>
</tr>
<tr>
<td>2016</td>
<td>10.74</td>
<td>33.82</td>
<td>985</td>
</tr>
<tr>
<td>2017</td>
<td>10.88</td>
<td>35.39</td>
<td>1198</td>
</tr>
<tr>
<td>2018</td>
<td>11.02</td>
<td>37.03</td>
<td>1528</td>
</tr>
<tr>
<td>2019</td>
<td>11.17</td>
<td>38.75</td>
<td>2110</td>
</tr>
<tr>
<td>2020</td>
<td>11.32</td>
<td>40.55</td>
<td>3405</td>
</tr>
</tbody>
</table>

Conclusion
In conclusion, nonlinear models shows promising results when applied to data on cotton crop of India. The nonlinear models viz., monomolecular, logistic, gompertz, quadratic, richards and reciprocal model were applied for cotton crop in India. Based on performance of these fits, best non-linear models were chosen for the selected series. In the first instance, attempts were made to identify the model that best described this data set. It is evident, from the analysis, area, production and productivity of cotton is fluctuating and all six nonlinear models are suitable to fit. The initialization of parameter done results are compared using statistics such as, $R^2$, residual sum of square and mean square error accuracy than their source model (Table 1, 2 and 3). Run-test statistic ($|z|$ value and the normality test for Shapiro-Wilk test also presented. The area of anlysed for the selected for Gompertz model for the best fit statistics and the shows an parameter equation model and the production measures for selected the best fit as quadratic model comparison on logistic and reciprocal shows the equation(2). And also productivity selected for reciprocal as the goodness of fit comparing other models. The results with the shapiro-wilk test normality and run test randomness criteria and the
correlation characteristic are summarized in Table 5. A graphical representation of the adequacy of the fitted models is presented in Fig. 1, 2 and 3. Furthermore, nonlinear estimation techniques may contribute to determining of the economic information and marketing strategies in plant-based enterprises.

References