Print ISSN : 0972-8813
e-ISSN : 2582-2780

[Vol. 18(3), Sept-Dec, 2020]

Pantnagar Journal of Research

(Formerly International Journal of Basic and Applied Agricultural Research ISSN : 2349-8765)

G.B. Pant University of Agriculture & Technology, Pantnagar

PANTNAGAR JOURNAL OF RESEARCH

Vol. 18(3) September-December, 2020

CONTENTS

Comparison of non-linear models on area, production and productivity of sugarcane crop in Uttar Pradesh

JHADE SUNIL[1] and ABHISHEK SINGH[2]

[1]*Agricultural Statistics, Institute of Agricultural Science, Banaras Hindu University, Varanasi,* [2]*Department of Farm Engineering (Agricultural Statistics), Banaras Hindu University,Varanasi(Uttar Pradesh)*

ABSTRACT: The present study was carried out to bring the realistic nature of modeling variables, using non-linear regression approach on area and production of sugarcane in Uttar Pradesh, using the secondary data of area and production from1980 to 2017. Six different methods of nonlinear regression models as Logistic, Power, Sinusoidal, Richards, Rational and Ratkowsky were used. Levenberg-Marquardt technique was used to estimate the unknown parameters of the nonlinear regression models. To select a best fitted model by using R^2, RMSE, MAE, and residual assumption tests such as Run test, Shapiro-Wilks test were carried out. For area, production and productivity it was found that Rational model had the highest values. i.e. (85 per cent, 83 per cent, 75 per cent) and lowest RMSE (84.4, 9992.7, 3546.4). Hence, Rational model is the most suitable among the fitted nonlinear model which can be used for further trend analysis of sugarcane in Uttar Pradesh.

Key words*:* Area, non-linear model, production, productivity, root mean square error, sugarcane

Sugarcane (*Saccharum officinarum*) is a conventional crop of India. It plays a vital role within the agricultural and industrial economy of the country. It is cultivated in most of the states, but it covers an insignificant share in gross cropped area of the country. Brazil (736 MT), India (352 MT), and China (126 MT) are the highest sugarcane producing countries. Its share within the country's economic growth has become significant. (Anon, 2015). India's share in world sugar production has improved from 5 per cent to 17 per cent over the last 50 years. In Uttar Pradesh sugarcane area 27.94 lakh hectares, production of 138.48 thousand MT and sugarcane productivity is 80.50 tonnes per hectare during 2018-19. It is one of the most important commercial crops of Uttar Pradesh. It is covering less than 3 per cent of the total cropped area within the State, but it utilizes more than sixty per cent of the whole water available for irrigation within the State. Most of the cases relationship among variables in agricultural and biological sciences are 'nonlinear' in nature. A unit increase in the value of independent variable may not result in an equivalent unit increase in the dependent variable in nonlinear models. The main objective of this study was to focus on trends of sugarcane area, production, and productivity in Uttar Pradesh by using appropriate nonlinear growth models analysis model (Dinesh *et al.,* 2018).

MATERIALS AND METHODS

The present study was conducted on 37 years data on area, production, and productivity of Uttar Pradesh state from 1980-2017. Secondary data were collected from the Directorate of Economics and Statistics, Government of Uttar Pradesh. In this study different models are compared *viz.,* Logistic, Power, Sinusoidal, Richards, Rational and Ratkowsky models for estimating the growth of sugarcane crop area, production, and productivity to find the best fit using the statistical methods such as highest coefficient of determination, mean sq. error, and root mean sq. error.

Non-linear regression models

A nonlinear regression model is one in which at least one of the parameters seems nonlinearly. A nonlinear model, which can be transformed into a linear model by some transformation is termed 'intrinsically linear', else it's known as 'intrinsically nonlinear'. Mathematically, in nonlinear models at least one of the derivatives of the expectation function with respect to at least one parameter is a function of parameter(s).

$$y=ab^x x+ e$$

Where'y' is the area, production, and productivity throughout the time x; 'a' represents the carrying capacity; 'b' is function of initial value and 'e' is that the error term.

The above model could be a nonlinear regression model because the derivatives of 'y' with respect to 'a' and 'b' are both functions of 'a' and/or 'b'. Parameters in a nonlinear model can also be estimated by the method of least squares. However, due to the difficulty in the procedure of computation, the common practice is to work with the log transformed model.The original structure of the error term got disturbed due to transformation. Proceeding further to carryout residual analysis for the residuals generated by the transformed model, will result in an incorrect conclusion. As a remedy to these drawbacks, nonlinear regression model procedures are already developed in computer intensive tools to find solution for the parameters (Venugopalan and Shamasundaran, 2003).

To get estimates of the unknown parameters of a nonlinear regression model, Levenberg-Marquardt technique was used. during this methodology, the subsequent steps are carried out.

Step (I): Start with a good initial guess of the unknown parameters, a sequence of 'θ's that hopefully converge to θ is computed.

Step (ii): Objective function expressed as

$$s(\theta)= \Sigma_{x=1}^{n} [Y_x - F_i(\theta)]^2$$

is reducedwith respect to the current value of θ and the new estimates are obtained.

Step (III): The final step is sustained until the successive iteration yielded parameter estimate values are close to each other.

Some important nonlinear growth models

A mathematical model is an equation that represents the behavior of a system (France and Thornley, 1984). We now discuss some well-known nonlinear growth models. The following nonlinear growth models are Logistic, power, Sinusoidal, Richards, Rational and Ratkowsky, are tried. (Priya Krishna,and Bajpai,2011)

1) Logistic model: The parameters of this model has a simple interpretation. The mathematical form of the model is given by.

$$Y= \frac{a}{1+b\exp\{-cX\}} +e$$

Where Y is the area, production, and productivity throughout the time X; a, b, and c are the parameters, and 'e' is the error term. The parameter 'a' represents the carrying capacity, parameter 'c' is the intrinsic growth rate, and the 'b' represents different functions of the initial value Y (0), and 'd' is the added parameter.

$$Y=aX^b+e$$

2) Power: Another non-linear regression model is the power regression model, which is based on the subsequent equation

3) Sinusoidal model: A sinusoidal function is a function in sine or cosine (trigonometric function). The mathematical form of the model is given by.

$$Y=a+b\cos(cX-d)+e$$

The parameter 'a' represents the carrying capacity, parameter 'c' is the intrinsic growth rate, and the parameter 'b' represents different functions of the initial value Y (0), 'd' is the added parameter, and 'e' is the error term.

4) Rational: Rational function models are moderately easy to handle computationally. A main advantage of this model is its ability to compute starting values using a linear least squares fit. To do this, *p* points are chosen from the data set, with *p* denoting the number of parameters in the rational model.

$$Y= \frac{a+bX}{1+cX+dX^2} +e$$

5) Richards: It is a four-parameter growth model. In 1959 Richards proposed this growth model. It is represented by:

$$Y= \frac{a}{(1+e^{b-cX})^{1/d}} +e$$

Where the value of 'd' lies between the range $-1 < d < 0$. Richards model is a generalization of logistic (when d = 0). 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 Prajneshu, 2004).

6) Ratkowsky: The mathematical form of the model is given by (Ratkowsky,1968).

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

Choice of initial values of the parameters for several models

There are several ways to determine starting parameter values for nonlinear models. The most appropriate method for making the initial guesses is by the use of previous information. All the iterative procedures require initial values θ_{r0} (r = 1, 2, 3…, k) of the parameter θ_r. In this study, the Curve expert Ver.2.6.5 software package is used to estimate the initial values.

Model adequacy checking

The following measures of goodness of fit have been used to justify the adequacy of the model developed. The coefficient of determination R^2 which is the most commonly used method is given by

$$1)\ R^2=\frac{RSS}{rss}=1-\frac{\Sigma_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{\Sigma_{i=1}^{n}(Y_i-\bar{Y}_i)^2},\ 0 \le R^2 \le 1$$

2) Mean Absolute Error (MAE) = $\Sigma_{i=1}^{n} |Y_i - \hat{Y}_i|/n$

3) Mean Squared Error (MSE) = $\frac{\Sigma_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{(n)}$

4) Root Mean Square Error (RMSE) = $\sqrt{\frac{\Sigma_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n}}$

Here 'n' denotes the total number of observed values and 'p' denotes the number of model parameters

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 is random
(ii) Errors are normally distributed.

Shapiro-Wilk test used to test whether the residuals are normally distributed or not. If the fitted model is correct, the residuals should follow the assumptions.

Shapiro-Wilk test (Test for Normality)

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

$$W = \frac{(\Sigma_{i=1}^{n} a_i x_i)^2}{\Sigma_{i=1}^{n} (x_i x_i)^2}$$

Where, $x_{(i)}$ is the i^{th} order statistic, i.e., the smallest number in the sample; The constants ai are given by

$$(a_1, a_2, ...a_n) = \frac{m^T v^{-1}}{\sqrt{(m^T v^{-1} v^{-1} m)}}$$

Where, $m^T = (m_1, m_2, ...m_n)^r$ and $m_1, m_2, ...m_n$ are the expected values of the order statistics of (iid) Independent and Identically-Distributed random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics. Then values a_1, coefficients are tabulated by Shapiro and Wilk,1965.

Run test (Test for Randomness)

A run is defined as sequence of symbols of one kind separated by symbols of another kind. The residuals are replaced by 'plus' or 'minus' and accordingly they are positive or negative. The run test is based on number of runs. The test statistics is,

$$z = \frac{r - \mu_r^{-1}}{\sigma r} \sim SND\,(0,1)$$

Where

$$\mu_r = \frac{2n_1 n_2}{n_1 + n_2} + 1 \qquad \sigma r = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_1 - 1)}$$

With n_1 and n_2 denote the number of positive and negative values in the series respectively.

The runs test rejects the null hypothesis if

$$|z| > z_{1-\frac{\alpha}{2}}$$

Table 1: Estimate of the parameters along with model adequacy of fitted nonlinear models for area under sugarcane (1980-2017)

Estimated parameters	Nonlinear Models					
	Logistic	Power	Sinusodial	Rational	Richards	Ratkowsky
a	100393.81	0.000017	-67766.09*	-110.948*	3481.1	2271.21
	(172603.07)	(0.0001)	(25324611)	(311.390)	(7423.21)	(82.030)
b	0.990**	19.098	69946.33	0.057*	26.44	140.25**
	(0.001)	(1.715)	(25324760)	(0.158)	(5047.1)	(33.214)
c	0.0072*		0.0074**	-0.0012**	0.01	0.07
	(0.015)		(0.651)	(0.0004)	(2400.2)	(0.017)
d			162.29	0.00025*	0.0081	
			(1313.74)	(0.0006)	(2550.1)	
R-Square	0.774	0.775	0.851	0.852	0.814	0.850
MAE	81.047	80.8941	65.442	65.227	71.58	65.693
MSE	10496.682	10442.260	7128.080	7126.38	8961.17	7206.330
RMSE	102.453	102.187	84.427	84.417	93.98	84.890
S-W Test (p value)	0.128	0.201	0.112	0.200	0.164	0.157
Run test (p value)	0.058	0.058	0.045*	0.097	0.045	0.054

* Significant at 5% level; ** Significant at 1% level, MAE: Mean Absolute Error; MSE: Mean Square Error; RMSE: Root Mean Square Error, Values in parentheses indicate standard errors

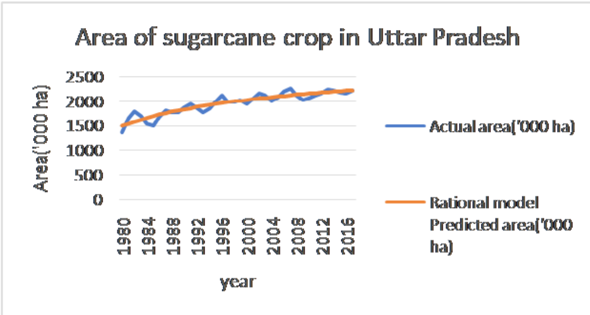

Fig 1: Actual and rational model predicted for sugarcane cultivated area in Uttar Pradesh

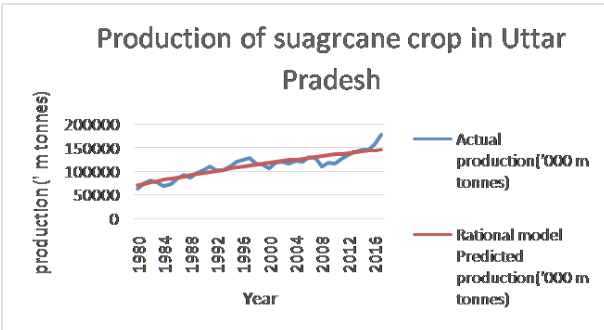

Fig 2: Actual and rational model predicted for sugarcane production in Uttar Pradesh

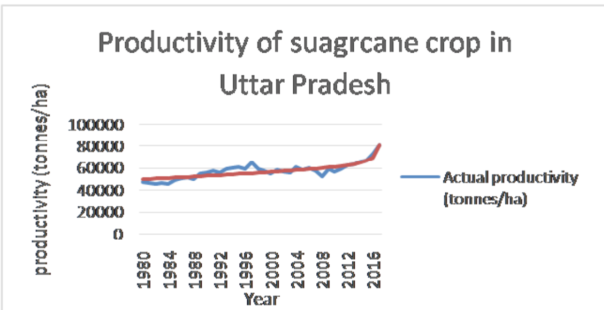

Fig 3: Actual and rational model predicted for sugarcane productivity in Uttar Pradesh

RESULTS AND DISCUSSION

The Levenberg-Marquardt procedure is the best iteration procedure in the methodology, which was used for solving nonlinear normal equations. Power, Logistic, Rational, Sinusoidal, Richards, and Ratkowsky models were used for studying area, production, and productivity of sugarcane in Uttar Pradesh. The results are discussed in the followings.

Trends in Area based on Non-linear Models

The result given (Table 1) for the area under the cultivation of sugarcane crop in Uttar Pradesh revealed that among the different non-linear models fitted, the highest R-square (85.29%) was observed in the case of Rational function with the lowest value of RMSE (84.41) and MAE (65.22) in comparison to that of other non-linear models. The next best nonlinear model was the sinusoidal model with R-square (85.10%). Based on the performance of these models, it was found that Rational model is found to be best fit. In the rational model actual and predicted trend value fitted (fig.1).Among the non-linear models, the rational model was found suitable to fit the trends in cultivated area of sugarcane crop.

Trends in Production based on Non-linear models

Results study shows that among the six models used for the production of sugarcane in Uttar Pradesh, (Table 2) revealed that, among the different nonlinear models fitted, the maximum R-square value of 83.3 per cent was observed in the rational model with the minimum RMSE (9992.7) and minimum MAE (7322.4) values on comparison with all other nonlinear models. The next best nonlinear model was the sinusoidal model with 83.10 per cent of R-square value. The p value of Shapiro-Wilk test statistic (0.098) and the Run test statistic (0.830) to test for assumptions, indicate that the residuals of the rational model were normal and random respectively Among the non-linear models, the rational model was found suitable to fit the trends (Fig.2) in cultivated sugarcane crop estimated model.

Trends in Productivity based on Non-linear Models

In case of productivity of sugarcane, the maximum R-square (75.8%) was observed in the model of rational function with minimum value of RMSE (3546.4) and MAE (2814.5) in comparison to that of other non-linear models. The next best nonlinear model was the power model with 67.10 per cent of R-square value. The p value of Shapiro-Wilk test statistic (0.178) and the Run test statistic (0.145) to test for assumptions indicates that the residuals of the rational model were normal and random respectively Among the non-linear models. All the estimated value of the parameter in this model were significant at 5% level of significance. Among the sugarcane nonlinear models fitted to the productivity of sugarcane crop the rational model was found to be the selected to fit (Fig 3).

CONCLUSION

Sugarcane is one of the important cash crops in Uttar Pradesh. The data on sugarcane area, production, and

Table 2: Estimate of the parameters along with model adequacy of fitted nonlinear models for sugarcane production in UP (1980-2017)

Estimated parameters	Nonlinear Models					
	Logistic	Power	Sinusodial	Rational	Richards	Ratkowsky
a	17251.6	0.00162	-2072557	-931762*	661826.63	192586.1
	(12451.2)	(0.0002)	(1743465)	(624856)	(19219.8)	(4127.14)
b	0.981**	37.498**	1003171	538839**	10.278	8.847*
	(0.001)	(2.894)	(1743559)	(31874.2)	(838.45)	(0.0005)
c	0.002*		0.003	-1249.81*	0.010**	0.044
	(0.00058)		(1.266)	(83820.08)	(0.017)	(0.0001)
d			154.094	0.0665*	0.00292	
			(2654.34)	(4458.6)	(0.023)	
R-Square	0.813	0.826	0.831	0.833	0.824	0.814
MAE	8130.458	8116.642	7346.831	7322.427	7896.425	8122.197
MSE	110047945.4	109827438.8	100916934.5	99854932.71	102768971	110126193
RMSE	10490.373	10479.858	10045.742	9992.744	10137.54	10487.69
S-W Test (p value)	0.092	0.094	0.122	0.098	0.123	0.090
Run test (p value)	0.041*	0.041*	0.045*	0.830	0.054	0.083

* Significant at 5% level; ** Significant at 1% level, MAE: Mean Absolute Error; MSE: Mean Square Error; RMSE: Root Mean Square Error, Values in parentheses indicate standard errors

Table 3: Estimate of the parameters along with model adequacy of fitted nonlinear models for sugarcane productivity in UP (1980-2017)

Estimated parameters	Nonlinear Models					
	Logistic	Power	Sinusodial	Rational	Richards	Ratkowsky
a	1729.160	0.0009	775989.73	4090.691*	5357433.4	5200983.13
	(3717.907)	(0.0001)	(2264555)	(1096.747)	(2491422)	(1252040)
b	0.991	18.407	786346.59	-2.027**	3.211	23.188
	(0.001)	(2.149)	(2273820)	(0.541)	(18589.0)	(194.136)
c	-0.198		0.002	-0.001*	0.002	0.009
	(0.0064)		(2.647)	(0.000)	(4.504)	(0.026)
d			81.924	0.00002*	0.051	
			(4697.59)	(0.000)	(487.1)	
R-Square	0.66	0.671	0.647	0.758	0.647	0.634
MAE	3279.328	3277.500	3286.077	2814.543	3286.077	3292.55
MSE	18365897.9	18369539.6	18357092.9	12577447.9	18357092.9	18343831.3
RMSE	4285.545	4285.970	4284.517	3546.469	4284.517	4282.969
S-W Test (p value)	0.129	0.131	0.144	0.178	0.134	0.130
Run test (p value)	0.178	0.152	0.045*	0.145	0.149	0.054

* Significant at 5% level; ** Significant at 1% level, MAE: Mean Absolute Error; MSE: Mean Square Error; RMSE: Root Mean Square Error, Values in parentheses indicate standard errors

productivity has shown an increasing trend over years. It was observed that nonlinear models are more appropriate to visualize the temporal trend of area, production and productivity of sugarcane in Uttar Pradesh. The nonlinear models *viz.*, Power, Logistic, Rational, Sinusodial, Richards and Ratkowsky models were used for sugarcane crop in Uttar Pradesh. Based on performance of these fits, best non-linear models were chosen for the selected series. Rational model were the most suitable fitted models which clearly explained the trend of area, production and productivity of sugarcane in Uttar Pradesh.

REFERENCES

Anonymous (2015). Food and Agriculture Organization of United Nations statistics 2015. Food and Agriculture Organization of United Nations, Rome, Italy. http://www.fao.org/faostat/en/#data.

Chandran, K.P. and Prajneshu (2004). Computation of growth rates in agriculture: Nonparametric regression approach. *Journal of the Indian Society of Agricultural Statistics*,57:382-392.

Dinesh Kumar, P., Bishvajit B. and Manjunath, V. (2018).Nonlinear Modeling of Area and Production of Sugarcane in Tamil Nadu, *India. Int.J.Curr.Microbiol.App.Sci*., 7(10): 3136- 3146.

Priya Krishna, S.R. and Bajpai. (2011). Computation of compound growth rates for sugarcane using non-linear growth models. *Ind. J. of Sugarcane Tech*., 26(2): 33-39.

Ratkowsky D.A. (1986). Statistical properties of alternative parameterization of the von Bartalanffy growth curve. *Can. J. Fish. Aquat. Sci*.,43:742-747.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3- 4): 591-611.

Venugopalan, R. and Shamasundaran, K.S. (2003). Nonlinear Regression: A realistic modeling approach in Horticultural crops research. *J.Ind. Soc.Ag. Statistics*, 56(1):1-6.

Received: December 12, 2020
Accepted: December 25, 2020