Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 77(3) 2023 249–255

Nonlinear Statistical Modelling of Area and Production of Apple Crop of Kashmir Valley

Uzma Majeed, Nageena Nazir, S.A. Mir, Immad A. Shah and Ishfaq A. Bhat

Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, J&K, India

Received 13 February 2023; Revised 18 June 2023; Accepted 08 August 2023

SUMMARY

The current study was conducted to compare the performance of five nonlinear growth models, namely, Monomolecular, Logistics, Gompertz, Richards, and Weibull for studying the area and production of the apple crop in Kashmir. Long-term data for the past 45 years pertaining to the area and production of apple was obtained from the Directorate of Horticulture. Run test and Shapiro-Wilk tests were employed to examine the assumptions of 'independence' and 'normality' of error terms, respectively. Levenberg-Marquardt (LM) iterative method was used to estimate the parameters. The best model was chosen based on various model adequacy test criteria including MAE, RMSE, MSE, and R^2. Therefore, Richards and Weibull models were found to be the best fit for the area and production of apple, respectively. The area and production of the apple crop were forecasted for a few years using the best fit models.

Keywords: Nonlinear models; Apple; Levenberg-Marquardt(LM); Parameters; Forecasting.

1. INTRODUCTION

Apple is amongst the most widely cultivated fruits globally, with numerous cultivars including the well-known Red Delicious, Golden Delicious, Fuji, Jonagold, etc.(wu *et al.* 2016). It has the flexibility to cope with many ecological settings while yielding quality output, which has earned it a prominent position in fruit production (Uzar *et al.* 2019). It has been reported with elevated nutraceutical attributes, consumed raw or processed(Jeločnik *et al.* 2019). However, its production technology is sophisticated, labor and resource-intensive, necessitating huge expenditures with substantial cumulated output. The territory of Jammu and Kashmir is highly renowned across the nation for its horticulture crop production According to FAO 2020, the apple area was 4,904,305 ha with an output of 86,142,197 tonnes globally, where Asia accounts for 62.1% of production., both in term of quality and quantity. Apple, pear, plum, apricot, peach, almond, cherry, and other horticultural crops are produced extensively, where apple cultivation has abundantly colored the tranquility of Kashmir's landscape. It is a dominant crop that generates the most marketable surplus and contributes to more than 70% of total apple production in India, with 10.25 metric tons of production per hectare (Tajamul and Sunita 2017). The area and production of apple in 2019-2020 were 164854 ha and 2026472 metric tonnes, respectively (*Directorate of Economics and Statistics Government of J&K)*. The major contributing districts in apple production includes Baramulla, Shopian, Pulwama, Budgam, Anantnag, and Kulgamhectare (Tajamuland Sunita 2017). Pertaining to the potential of apple, forecasting or prediction of its area and production has achieved a massive cognizance nowadays, is beneficial to farmers, policymakers, planners, etc,. and can be utilized for uplifting the economic growth.

Modeling allows you to investigate the nature of relation among variables of the study. To anticipate the values of one or more variables based on data from other variables, the relation among variables must be evaluated. Model development is being employed in numerous disciplines including agriculture, horticulture, forestry, meteorology, industry, and

Corresponding author: Immad A. Shah
E-mail address: immad.shah@icmr.gov.in

others (Prasad, 2010). Linear regression models have typically been employed to assess crop growth patterns, however, it is conceivable that these models would not serve satisfactorily fit well. Therefore, it is widely acknowledged that any statistical investigation wherein principles from a body of knowledge are seriously contemplated is probable to result in a nonlinear model(Seber and Wild, 2003). Such models are useful for comprehending the intricate inter-relationship between the parameters. A variety of nonlinear growth models were employed to thoroughly investigate the state-wise wheat productivity patterns during 1973-74 to 1996-97 in four dominant wheat-growing states (Punjab, Haryana, Uttar Pradesh, and Rajasthan) as well as at the national level (Prajneshu and Das, 2000). Additionally, the area, production, and productivity of sugarcane and sesame crops in Andhra Pradesh have also been forecasted using various linear, nonlinear, and time series models (Greeshma, *et al.* 2017, Priyanka, 2020. Moreover, growth in the area, production and productivity of cotton crop in India has been investigated using several nonlinear growth models (Sundar,. and Palanivel, 2017). Therefore, keeping the above aspects under consideration; the present study was conducted to forecast the area and production of apple using nonlinear modeling in Kashmir valley. The significance of this research could be considered and used in making strategies concerning the development of apple production.

Description: Apple growers packing apples.

2. MATERIALS AND METHODS

The current study was carried out with the aim of analyzing the growth rates of area and production of apple crop using five nonlinear models viz., Monomolecular model, Logistic model, Gompertz model, Richards model, and Weibull model. For this study, data for the last 45 years on the area and production of apple was collected from the Directorate of Horticulture. In the first part of the analysis, the Run test and Shapiro-Wilk tests were used to check the randomness and normality of residuals respectively.

In the second part, regression analysis using various nonlinear models was carried out for the area and production of the apple crop.

In the third part, model selection tests were carried out for the area and production of apple crop based on the results obtained from regression analysis.MAE, RMSE, MSE, and R^2were used to examine and select the best model. Based on the selected model, the forecasting was done for a period of seven years.

2.1 Trend Models

2.1.1 Nonlinear Models

A model in which at least one of the parameters appears nonlinearly is termed as a nonlinear model. For example

$$Y_t = aX_t^b + \varepsilon_i\ ,\ 1 \leq i \leq n$$

$$\varepsilon_i \sim N(0,\sigma^2)$$

It is a nonlinear model because the derivatives of Y_t with respect to a and b are both functions of a and/or b. Ratkowsky (1990) goes into great detail regarding the nonlinear model family.

The nonlinear growth models include:

2.1.1.1 Monomolecular model

This model depicts the progression of a growth situation in which the growth rate is considered to be proportionate to the resources yet to be achieved.

$$Y_t = c-(c-b)e^{-at} + \varepsilon_i\ ,1 \leq i \leq n$$

$$\varepsilon_i \sim N(0,\sigma^2)$$

2.1.1.2 Logistic model

This model is given by the equation

$$Y_t = c \div (1+be^{-at}) + \varepsilon_i\ ,1 \leq i \leq n$$

$$\varepsilon_i \sim N(0,\sigma^2)$$

The graph is elongated S-shaped and symmetrical about its point of inflection.

2.1.1.3 Gompertz model

This model is also a sigmoid type. However, it is not symmetric about its point of inflection. It is given by the equation

$$Y_t = ce^{-e^{b-at}} + \varepsilon_i\ ,\ 1 \le i \le n$$

$$\varepsilon_i \sim N(0,\sigma^2)$$

2.1.1.4 Weibull model

This model is given by

$$Y_t = c - be^{-at^d} + \varepsilon_i\ ,\ 1 \le i \le n$$

$$\varepsilon_i \sim N(0,\sigma^2)$$

2.1.1.5 Richards Model (Richard 1959)

This model is given by

$$Y_t = c \div (1+e^{b-at})^{1/d} + \varepsilon_i\ ,1 \le i \le n$$

In all these models

t is the independent variable i.e., time

Y_t is the value of the response variable at time t

c is the carrying capacity

b represents the different functions of the initial value Y_0

a is the intrinsic growth rate

d is an added parameter in the Richards model

ε is the error term

Evidently, the Richards model has 4 parameters.

a. Model Diagnostics

2.2.1 Residual Analysis

In regression model, two important assumptions are made:

I. Independence of residuals and
II. Normality of residuals

2.2.1.1 Test for the independence of residuals (Run test) (Waldand Wolfowitz, 1940)

We test

Ho: Residuals are independent

against

H1: Residuals are not independent.

A residual is substituted with a '+' or '-' sign depending on the sign of its value in this test. Assume that m represents the number of '+' and n represents the number of '-'. The number of runs determines the outcome of the test (r). A run is a collection of symbols of one type divided by symbols of another type (Siegel *et al.* 1988). A large number of runs approximates to the sampling distribution of the normal distribution

For large samples the required test statistic is

$$Z = (r + h - \mu)/\sigma \sim N(0,1)$$

where

h=+0.5 if r<μ,

h=-0.5 if r>μ.

Ho is rejected at level α if $|Z| > Z_{\alpha/2}$, where

$$Z_a = P\{Z > Z_a\} = \alpha$$

2.2.1.2 Test for normality of residuals (Shapiro-Wilk test) (Shapiro and Wilk, 1965)

We test

Ho: Residuals are normally distributed

against

H1: Residuals are not normally distributed.

The required test statistic W is defined as

$$W = S^2 / b$$

where

$$S^2 = \sum a(k)\{x_{(n+1-k)} - x_{(k)}\}$$

$$b = \sum (X_i - \bar{X})^2$$

k takes the values

k=1,2,…,n/2, when n is even

k=1,2,…,(n-1)/2, when n is odd

and $x_{(k)}$ is the kth order statistic of the set of residuals. The values of coefficients a(k) for different values of n and k are obtained from tables(Shapiro and Wilk, 1965). If the calculated value of W is less than tabulated value at level α, thenH_0 is rejected otherwise accepted.

2.2.1.2 Test for autocorrelation of residuals (Durbin-Watson test)

Durbin-Watson test is used to test the presence or absence of autocorrelation in residuals. Durbin-Watson is the ratio of the distance between the errors to their overall variance. The test statistic is

Where $e_i = y_i - \hat{y}_i$ and y_i and $\hat{y}_i$ are, respectively, the observed and predicted values of the response variable for individual i. Thus, DW is equal to 2 minus two times the correlation of e_t and e_{t-1}.

Durbin-Watson is used both as diagnostic for autocorrelation and as estimate of ρ. DW statistic is a correlation and thus depends on values of independent variables as $-1 \leq \rho \leq +1$ thus $0 \leq DW \leq 4$.

3. RESULTS AND DISCUSSION

The results of nonlinear statistical models developed and statistical analysis of the area and production of apple overtime period of 45 years (1974-75 to 2019-2020) using nonlinear growth models are individually presented for studying the area and production of the apple crop in Kashmir. They are also used for developing a growth model which can be used to describe and also to forecast the area and production.

Tables 1 shows the results of the residual assumption tests for apple area and production. Tables 1 shows the random distribution of residuals because the run test statistic and Shapiro-Wilk test statistic values for the fitted models were well within the critical region of 1.96, indicating nonsignificant findings for all of the models considered. Further, the results revealed that for all the models, the number of runs and the Shapiro-Wilk test statistic was found to be non-significant (p-value > 0.05) at 5 per cent level significance indicating that the assumptions of randomness and normal distribution of residual were satisfied. The normal probability plots (Fig. 1 and Fig. 2) also revealed the same results. Moreover, the Durbin-Watson statistic recorded the value of closer to 2, which indicated that there was no serial correlation among the residuals and were independent. This further improves the statistical adequacy of the fitted models.

Table 2(a) shows that Weibull fit explains the area of the apple crop over time. The parameter estimates of Weibull model are, a carrying capacity(intercept) of 210.02and a growth rate of 0.109, depicting that the growth rate for every year is found to be 10.9percent. Table 2(b) reveals that Richards fit explains the production of the apple crop over time. The parameter estimates of the Richards model are, a carrying capacity(intercept) of 218.674 and a growth rate of 0.311 depicting that growth rate for every year is found to be 31.1percent. All the coefficients are significant.

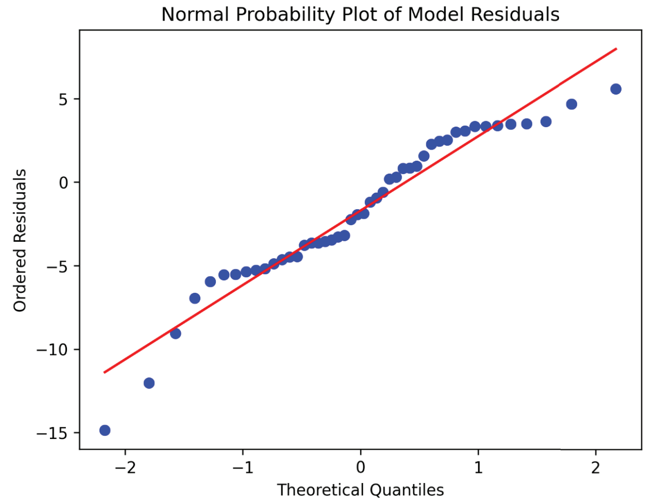

Fig. 1. Normal Probability plot of model residuals for apple crop area in Kashmir

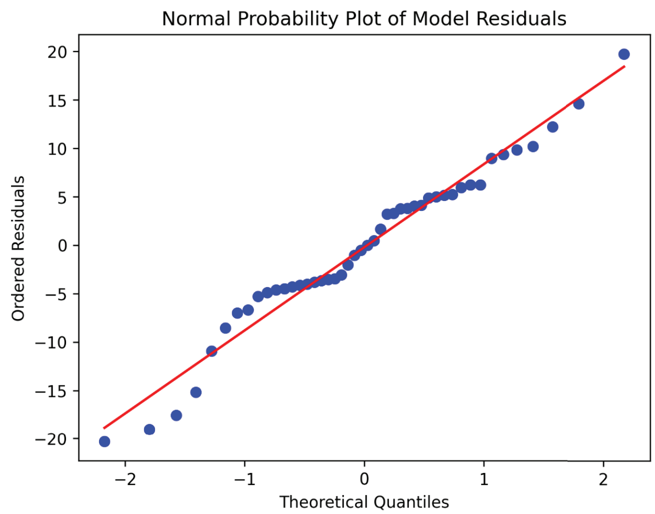

Fig. 2. Normal Probability plot of model residuals for apple crop production in Kashmir

Table 3 indicates the results of the model adequacy test used to identify the best fit for the apple area and production. For apple area, in the case of the Weibull fit, the Mean Absolute Error (3.84), Root Mean Square Error (6.08) and Mean Square Error (36.98) are all lower, and R^2(85.6%)is the highest. Hence the best fit that can be used to forecast the area under the apple crop is the Weibull fit. Further the table also indicates the model adequacy test, based on assessed components to identify the best fit for apple production. In the case of the Richards model, the Mean Absolute Error (9.27), Root Mean Square Error (14.51) and Mean Square Error (210.67) are all lower and R^2(88.9%) is the highest. Hence the best fit that can be used to forecast the production of the apple crop is Richards fit.

Table 1. Test for randomness and normality of residuals for apple area and production

TESTS	MONOMOLECULAR MODEL	LOGISTIC MODEL	GOMPERTZ MODEL	RICHARDS MODEL	WEIBULL MODEL
Run test (Area)	Parameters did not converge MONOMOLECULAR	Results in negative estimates of carrying capacity LOGISTIC	Parameters did not converge GOMPERTZ	-5.642^{NS} (0.561)	-4.621^{NS} (0.502)
Shapiro-wilk test (Area)				0.841^{NS} (0.291)	0.875^{NS} (0.285)
D-W statistic (Area)				1.543	1.680
Runtest (Production)				-0.139^{NS} (0.432)	Parameters did not converge
Shapiro-wilk test (Production)				0.905^{NS} (0.162)	
D-W statistic (Production)				1.795	1.502

Figures in the parentheses indicate p-values; NS: Not Significant

Table 2(a). Regression analysis for area under apple crop

PARAMETERS	MONOMOLECULAR MODEL	LOGISTIC MODEL	GOMPERTZ MODEL	RICHARDS MODEL	WEIBULL MODEL
Carrying capacity/ Intercept (c)	Parameters did not converge	Results in negative estimates of carrying capacity	Parameters did not converge	234.545** (33.039)	210.020** (12.508)
Function of initial value (b)				7.095** (2.152)	13.681** (5.340)
Intrinsic growth rate (a)				0.242** (0.045)	0.109** (0.001)
Added parameter (d)				2.342** (0.484)	3.130** (0.405)

**Significant at 5% Figures in the parenthesis indicate the standard error

Table 2(b) Regression analysis for production of apple crop

PARAMETERS	MONOMOLECULAR MODEL	LOGISTIC MODEL	GOMPERTZ MODEL	RICHARDS MODEL	WEIBULL MODEL
Carrying capacity/Intercept(c)	Parameters did not converge	Results in negative estimates of carrying capacity	Parameters did not converge	218.674** (13.958)	Parameters did not converge
Function of initial value (b)				12.481** (4.547)	
Intrinsic growth rate (a)				0.311** (0.084)	
Added parameter (d)				7.950** (3.739)	

**Significant at 5% Figures in the parenthesis indicate the standard error

There is a dearth of data on the amount of apple area and production expected in Kashmir in the future years. To remedy this gap in the literature, the current research attempts to anticipate apple area and production amounts for the years 2020–2026 using long-term data from 1974 to 2019, with an increase in apple area and output expected in 2026. This trend may be essential for the long-term viability of domestic consumption and exports in Kashmir's apple production.

Table 3. Model adequacy tests for area and production of apple crop

COMPONENTS	Weibull (Area)	Richards (Production)
Mean absolute error (MAE)	3.84	9.27
Root mean square error (RMSE)	6.08	**14.51**
Mean square error (MSE)	36.98	**210.67**
R^2(%)	85.60	**88.90**

Forecasting area, production and productivity of apple crop

Table 4 shows the forecasted values of apple crop area and production for the period of seven years (2020-2027) of area and production of apple crop from the previous 45 years (1974-2019). The best fitted models i.e. Weibull model for apple area and Richards model for apple production are used for forecasting.

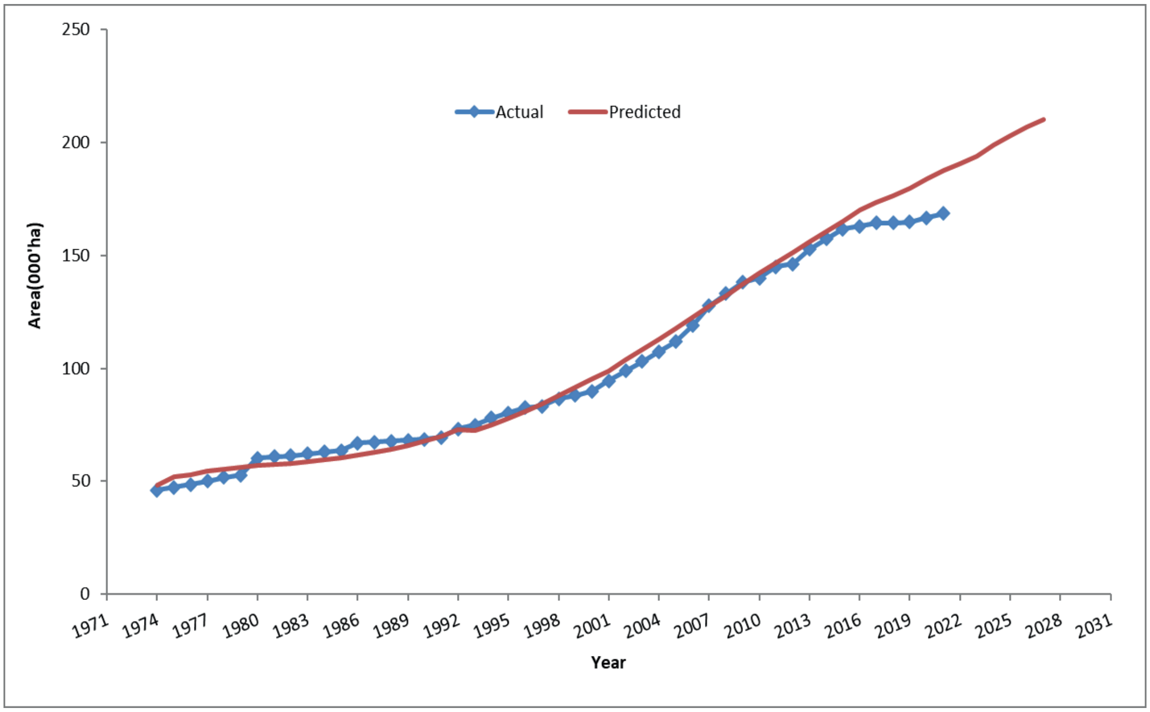

Fig. 3. Graph of actual, estimated and forecasted values of apple area (Weibull model)

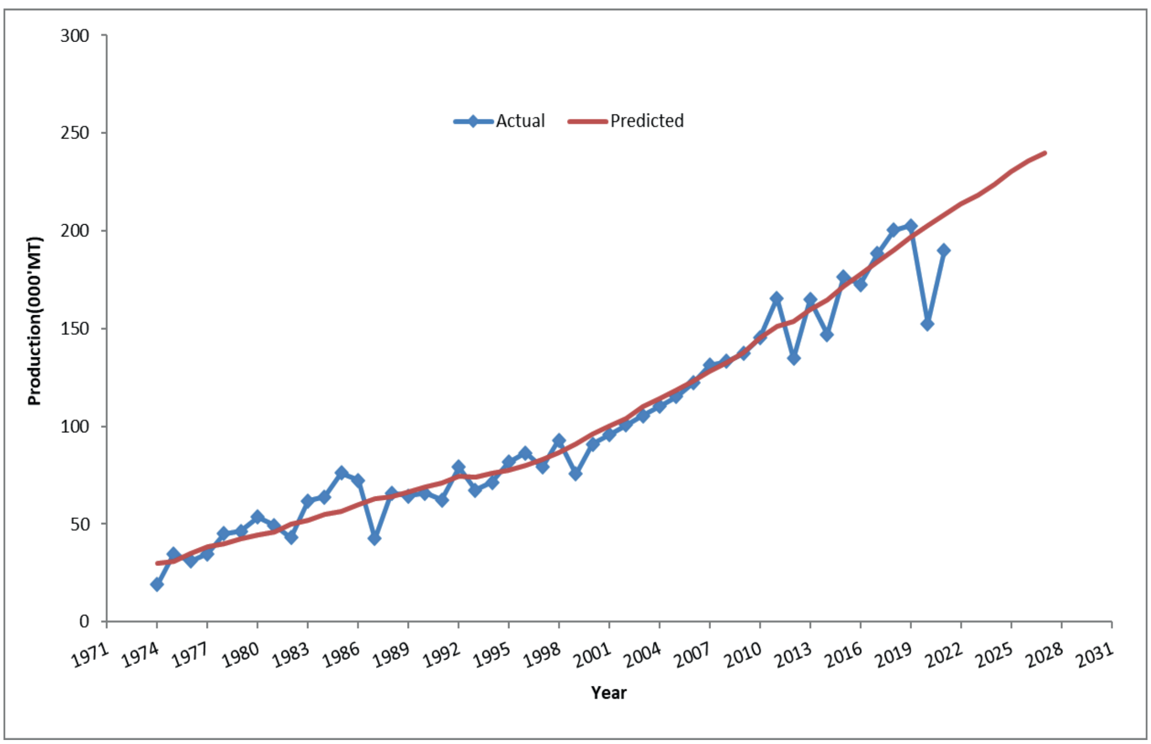

Fig. 4. Graph of actual, estimated and forecasted values of apple production (Richards model)

Table 4. Forecasted values of area and production of apple crop considering the actual recorded values of 45 years using the best fitted model

Year	Area (000' ha) Weibull Model	Production (0000` MT) Richards Model
2020-21	176.42	190.21
2021-22	179.67	196.65
2022-23	184.01	202.54
2023-24	187.67	208.19
2024-25	190.57	213.72
2025-26	194.09	218.33
2026-27	199.06	224.03

4. CONCLUSION

Nonlinear models show promising results when applied to data on the apple crop of Kashmir. The nonlinear models viz. Monomolecular, Logistic, Gompertz, Richards, and Weibull models are applied for the area and production of the apple crop in Kashmir. The best nonlinear models for the given series are chosen based on the performance of these fits. Attempts are made to identify the model that best describes this data set in the first instance. The area and production of the apple crop are clearly variable, as evident from the analysis. The value of the Run-test statistic and Shapiro-Wilk test are also presented. It is observed that the Weibull fit produced satisfactory results for the apple area data. Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error are lesser and highest R^2 (85.6%) in the case of Weibull fit. It can be deduced that the Weibull model is the best fit for forecasting the area under the apple crop.

In the case of apple production data, Richards fit produced satisfactory results. It has the lowest Mean Absolute Error, Root Mean Square Error and Mean Absolute Percentage Error, and the highest R^2 (88.9%) making it the best fit that can be used to forecast the production of the apple crop. A visual representation of the adequacy of the fitted models is presented in Fig.3 and 4.

ACKNOWLEDGEMENTS

The authors are grateful to the reviewers for their useful suggestions that helped in improving the quality of the manuscript.

REFERENCES

Economic Survey. (2019-2020). *Directorate of Economics and Statistics Government of J&K,* 2019-20.

FAO. Global apple production, portal of the FAO, Rome, Italy, (2020). retrieved at: www.fao.org/faostat/en/#data/QC, 15 May .

Greeshma, R., Bhave, M.H.V., Kumar, P.S.(2017). Application of growth models for area, production and productivity trends of sugarcane crop for coastal Andhra region of Andhra Pradesh. *International Journal of Agricultural Science and Research*, **7**(1), 7-14.

Jeločnik, M., Subić, J., Kovačević, V., (2019). Competitiveness of apple processing. *Ekonomika,*, **65**(4), 41-51.

N. Priyanka Evangilin, (2020). Statistical Model for forecasting area, production and productivity of Sesame Crop (Sesamum indicum L.) in Andhra Pradesh, India. *International Journal of Current Microbiology and Applied Sciences*, **9**(7), 1156-1166.

Panwar, S., Singh, K.N., Kumar, A., Sarkar, S.K., Paul, R., and Sivaramane, N.,(2014). Forecasting of growth rates of wheat yield of Uttar Pradesh through non-linear growth models. *Indian Journal Agriculture Science*, **84**, 856-859.

Prajneshu and Das, P.K.(2000). Growth models for describing state-wise wheat productivity. *Indian Journal of Agricultural Research*, **34**(3),179-181.

Prasad, H. (2010). Statistical modeling technique on export of fruit crops in India. *Indian Journal of Agricultural Sciences*, **74**(5), 619-625.

Ratkowsky, D.A.(1990). Handbook of nonlinear regression models. Marcell Dekker.

Richards, F.J.,A (1959) flexible growth function for empirical use. *Journal of Experimental Botany*, **10**, 290-300.

Seber, GA.F., Wild, C.2003.*Nonlinear Regression*, John Wiley, New York.

Shapiro, S.S. and Wilk, M.B., (1965). An analysis of variance test for normality. *Biometrika,* **52**, 229-240.

Siegel, S. and Centellen, (1988). N. J., Nonparametric Statistics for Behavioural Science, *Mcgraw-Hill*.

Singh, D.P., Gautam, S.S., Saxena, R.R.,(2017). Wheat production modelling in Madhya Pradesh using nonlinear statistical growth models. *American International Journal of Research in Science, Technology, Engineering & Mathematics,* **20**(1), 73-76.

Sundar, R.M. and Palanivel, M., (2017). Comparison of nonlinear models to describe growth of cotton. *International Journal of Statistics and Applied Mathematics*, **2**(4), 86-93.

Tajamul, R., and Sunita, S. (2017). A study on area, production and productivity of Apples in J&K from 2006-07 to 2015-16. *International Journal of Scientific Research and Management*, **5**(7), 6513-6519.

Ufuk, K.,, Palta, C., Kokten, K. and Bakoglu, A., (2010). Comparative study on some non-linear growth models for describing leaf growth of maize. *International Journal of Agriculture and Biology*, **12**(2), 227-230.

Užar,, D., Tekić, D., Mutavdžić, B.,(2019). Analiza i predviđanje proizvodnje jabuke u Republici Srbiji i Bosni i Hercegovini. *Ekonomija, teorija i praksa*, **12**(4), 1-10.

Wald, A. and Wolfowitz, J.,(1940). "On a test whether two samples are from the same population," *Ann. Math Statist.,* **11**, 147-162.

Wu., X., Wu., B., Sun., J., Ya., N.,(2016). Classification of Apple Varieties using Near Infrared Reflectance Spectroscopy and Fuzzy Discriminant C-Means Clustering Model. *Journal of Food Process Engineering*, **40** (2), 1-7 doi: 10.1111/jfpe.12355.