International Journal of Agriculture Sciences
ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 12, Issue 24, 2020, pp.-10522-10526.
Available online at https://bioinfopublication.org/pages/jouarchive.php?id=BPJ0000217

Research Article

STATISTICAL EVALUATION OF GROWTH MODELS FOR AREA, PRODUCTION AND PRODUCTIVITY OF TAPIOCA (CASSAVA) IN TAMIL NADU

KUMAR C.*[1], VIJAYABHAMA M.[2], SELVARAJ K.N.[3] AND RAVIKUMAR R.[2]
[1]Department of Physical Science and Information Technology, Tamil Nadu Agricultural University, Coimbatore, 641 003, Tamil Nadu, India
[2]Department of Basic and Social Sciences, Forestry College and Research Institute, Mettupalayam, 641 301, Tamil Nadu Agricultural University, Coimbatore, 641 003, Tamil Nadu, India
[3]Department of Agricultural Economics, Tamil Nadu Agricultural University, Coimbatore, 641 003, Tamil Nadu, India
*Corresponding Author: Email - kumaragri448@gmail.com

Received: December 05, 2020; Revised: December 25, 2020; Accepted: December 26, 2020; Published: December 30, 2020

Abstract: The time-series agricultural production-related data shows many peaks and troughs. Based on the behaviour of the data we can find out the best-fitted model, among the different models. Therefore, the present study was carried out based on the importance of the model selection to find out the growth rate using the various linear and non-linear models, *viz.* linear, logarithmic, inverse, quadratic, cubic, compound, S-curve, Growth function, Power function, Exponential, Logistic, Gompertz and Richards models. Time series data of the area, production and productivity of Tapioca in Tamil Nadu from 1970-71 to 2018-19 were collected from Directorate of Economics and Statistics, Government of Tamil Nadu. The study revealed that the cubic model is the best-fitted model about the area of Tapioca. Similarly, cubic and quadratic models are found to be the best model for the productivity of Tapioca by the obtaining highest R square and lowest RMSE, MAPE and MAE values. And for the production, none of the models was selected as the best model because of the residuals not following the normality.

Keywords: *Growth model, linear and nonlinear model, Time series, Tapioca*

Citation: Kumar C., *et al*., (2020) Statistical Evaluation of Growth Models for Area, Production and Productivity of Tapioca (cassava) in Tamil Nadu. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 12, Issue 24, pp.- 10522-10526.

Academic Editor / Reviewer: Dr Vipul N Kapadia, Dr Pramod Kumar Mishra

Introduction

Tapioca (*Manihot esculenta* Crantz) is considered as the king of tropical tubers. Tapioca and sweet potato are still the major tuber crops of India and cassava production is concentrated mainly in the southern states of Kerala, Tamil Nadu, and Andhra Pradesh. Tapioca is used as a food or feed in Kerala, while it is almost exclusively an industrial crop in the other two states [1]. For the Area, Tamil Nadu stands in the second position with around 40 per cent after Kerala (44%) and terms of productivity Takes 1st place with a productivity of 32 MT/ha.

Estimation of the growth rate is an important factor in agricultural economics and agricultural economics research. If there is one term that has been used in research papers published over the past forty years of the research may be about, particularly in the agricultural economics is about the "compound growth rates computation". However, there are several serious lapses in the methodology being followed for estimating growth rates, and the results drawn are thus not statistically sound. The purpose of the present study is to identify the methodological discrepancies and to suggest more appropriate procedures that can be implemented to achieve results [2].

Yadav and Kalola (2016) [3] evaluated the growth model over productivity data of rice and wheat. They reported cubic as the best model. Rajan and Palanivel (2018) [4] conducted a study on estimating growth rates using 10 growth models for cotton and found the cubic model as the best model. Pal and Mazumdar (2014) [5] forecasted the groundnut production using the non-linear growth models for India. Panwar *et al*., (2012) [6] using the non-linear regression approach using five models for the production of onion and found logistic and Gompertz models are the best models. Kumar *et al*., (2017) [7] calculated the growth and the instability index of Tapioca for the area, production and productivity in Karnataka. Using a semi-log function to find the trends in the area, production and productivity.

No comparative study has so far been carried out to find the growth model suitable for Tapioca. Comparison of different growth models will give the best-fitted model. Therefore, in this study, we shall evaluate 13 models and find the best-fitted model for the Tapioca.

Materials and methods

The study mainly focused on secondary data of Tamil Nadu regarding Tapioca area, production and productivity. Time series crop data were collected from the season and crop report, Directorate of Economics and Statistics, Chennai, for the period of 1970-71 to 2018-19.

The linear and non-linear models have different properties and behaviours. The present study examines the trend in the area, production and productivity of Tapioca using different linear and non-linear growth models *viz*, linear, quadratic, cubic, logarithmic, inverse, power, compound, S-curve, exponential, logistic, growth function, Gompertz and Richards function [8-10].

Linear and non-linear model

The trend equations were fitted using various growth models representing variable overtime behaviour. Using the modified Gauss-Newton iterative procedure available in the SPSS software, individual growth parameter estimates were obtained [8].

In this paper, the focus towards the area of Tapioca, production and productivity with help of linear and non-linear statistical models for highest R square and lowest RMSE, MAPE and MAE values to be considered for best fitting.

The growth models discussed here are as follows.

SN	Growth models	Formulae
1	Linear	$Y_t = \alpha + \beta t + e$
2	Logarithmic	$Y_t = \alpha + \beta \ln(t) + e$
3	Inverse	$Y_t = \alpha + \beta / t + e$
4	Quadratic	$Y_t = \alpha + \beta t + \gamma t^2 + e$
5	Cubic	$Y_t = \alpha + \beta t + \gamma t^2 + \delta t^3 + e$
6	Compound	$Y_t = \alpha \beta^t + e$
7	S-curve	$Y_t = \exp(\alpha + \beta / t) + e$
8	Growth function	$Y_t = \exp(\alpha + \beta t) + e$
9	Power function	$Y_t = \alpha t^\beta + e$
10	Exponential function	$Y_t = \alpha \exp \beta t + e$
11	Logistic function	$Y_t = \gamma / 1 + \beta exp^{-\alpha t} + e$
12	Gompertz model	$Y_t = \gamma \exp -\beta \exp -\alpha t + e, (\beta = \ln \gamma / y_{(0)})$
13	Richards model	$Y_t = \frac{\gamma}{\left[\{1 + \beta \exp(-\alpha t)\}^{\frac{1}{\delta}}\right]} + e$

Goodness of fit

The best-fitted model was examined by the goodness of fit of the models. The different goodness of fit measures used is Coefficient of determination (R^2), Root mean square error (RMSE), Mean absolute percentage error (MAPE) and Mean absolute error (MAE) using the following formula:

Goodness of fit	Formulae
Coefficient of determination (R^2)	$R^2 = 1 - \left[\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y}_t)^2}\right]$
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}$
Mean absolute percentage error (MAPE)	$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left\lvert\frac{y_t - \hat{y}_t}{y_t}\right\rvert$
Mean absolute error (MAE)	$MAE = \frac{\sum_{t=1}^{n}\lvert y_t - \hat{y}_t \rvert}{n}$

Where y_t the actual observation of time t. $\hat{y}_t$ is the predicted value for the same period $\bar{y}_t$. Overall sample mean of all the observations. In the present study, the model with the highest R square and lowest RMSE, MAPE and MAE values considered the best-fitted model.

Assumptions of the error term

After fitting out the statistical model, the next important step was to see whether the assumptions made about the error term are correct. This was done by examining the residues. It was obvious from the description that the residuals are the differences between what is observed and what is expected by the model *i.e.*, the amounts unexplained. The three assumptions are

Test for Normality

A Shapiro-Wilk test was used to verify normality. The test statistic w is the ratio of the best variance estimator (based on the square of the linear combination of order statistics) to the normal corrected sum of the predicted variance squares. The 'w' scale is from 0 to 1.

H_o = The residuals follow the normality

H_1 = The residuals do not follow the normality

Run test

The run test was used to determine whether a dataset was random or not. The run test was characterized as a series of increasing or decreasing values. The initial step in the run test was to count the number of runs in the data sequence. The values are then coded based on the median value. The value above the median as positive and the value below the median as negative.

H_o = The sequence produced in a random manner

H_1 = The sequence not produced in a random manner

Test for the independence of residuals

Durbin Watson was used to testing the existence or nonexistence of autocorrelation in residuals.

H_o: $\rho=0$

H_1: $\rho>0$

The DW values are ranges between 0 to 4

2 is not an autocorrelation.

0 to <2 is positive autocorrelation; (common in time series data).

>2 to 4 is a correlation between the variables negative (less common in time series data).

Result and discussion

The statistical software's such as R, Minitab and SPSS were used for the analysis. Various growth models were employed to finding the growth rates of area, production and productivity of Tapioca. The findings of the study are discussed below.

Table-1 *Descriptive statistics of Area, Production and Productivity of Tapioca*

	Area (hectare)	Production (kg)	Productivity (kg/hectare)
Mean	78078.53	2548746.67	30721.16
Minimum	38587.00	466660.00	10715.26
Maximum	140092.00	5912307.00	42900.46
SD	26818.90	1340509.93	7959.83
Kurtosis	-0.43 (Platykurtic)	-0.19 (Platykurtic)	0.85 (Leptokurtic)
Skewness	0.63	0.54	-1.03
CV %	34.35	52.59	25.91

Box plot

Box plot was the tool for the graphical representation of the numerical data by their quartiles. The standardised way of displaying the data with minimum, maximum, median, first quartile and third quartile. Box plot was also useful to find the number of outliers present in the data set. There was no outlier present in the area and production of the Tapioca. In productivity, data were five outliers present because of the lower productivity of the Tapioca [Fig-3].

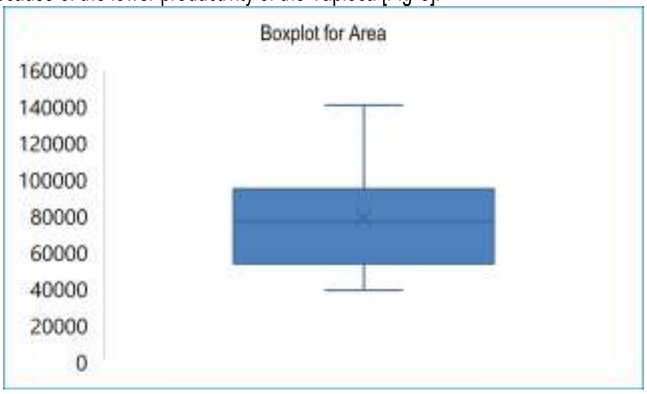

Fig-1 Boxplot of Area

Trends in area

Area of Tapioca in Tamil Nadu ranges from 38587 hectares to 140092 hectares with an average of 78078 hectares [Table-1] and the coefficient of variation is 34%. Kurtosis value is -0.43 it shows that the data was slightly platykurtic. Skewness value was 0.63 means positively and moderately skewed. Different linear and non-linear models were fitted for the area of Tapioca among these models' cubic model with the maximum R^2 (0.785), minimum of RMSE (12304), MAPE (0.1231) and MAE (9237.889) values [Table-2] found the suitable fit for the model when compared to the other models. The parametric values are 59005.35 (a), -4058.42 (b), 328.89 (c) and -4.92 (d). Normality of each model was calculated, the Shapiro-Wilk test was used for the calculation of normality of residuals, and the value is 0.97 with the probability value of 0.22. This value shows that it follows the normality. The randomness of the residuals was tested using the run test with the value of -1.92 with the probability value of 0.06 it showed that the randomness of the residuals. Durbin Watson test (0.99) showed that the errors are independent. The cubic model follows all the above-listed error assumptions also confirmed by the [Table-2].

The equation of the cubic model was

$Y = 59005.35 - 4058.42\,X + 328.89\,X^2 - 4.92\,X^3$ .... (1)

Table-2 *Characteristics of linear and non-linear models for the area of Tapioca in Tamil Nadu*

Model	Parameter Estimates				Model evaluation				Run test	Shapiro-Wilk	Durbin Watson
	a	b	c	d	R Square	RMSE	MAPE	MAE			
Linear	43536.59	1381.68			0.54	17965.75	0.16	13417.46	-2.84(0.00)	0.97(0.26)	0.48
Logarithmic	14501.53	21549.18			0.51	18601.98	0.21	15270.12	-3.64(0.00)	0.95(0.02)	0.45
Inverse	85310.02	-79108.53			0.22	23438.48	0.26	18811.62	-3.04(0.00)	0.93(0.01)	0.33
Quadratic	26358.15	3402.67	-40.42		0.62	16448.5	0.18	13329.19	-3.14(0.00)	0.96(0.10)	0.56
Cubic	59005.35	-4058.42	328.89	-4.92	0.79	12304	0.12	9237.89	-1.92(0.06)	0.97(0.22)	0.99
Compound	46145.03	1.02			0.62	19223.6	0.17	14167.33	-4.03(0.00)	0.98(0.50)	0.42
Power	30282.9	0.30			0.62	18044.27	0.17	13842.29	-3.64(0.00)	0.94(0.02)	0.47
S	11.31	-1.18			0.31	23044.74	0.23	17800.44	-3.10(0.00)	0.93(0.01)	0.30
Growth	10.74	0.02			0.62	19223.6	0.17	14167.33	-4.03(0.00)	0.98(0.50)	0.42
Exponential	46145.03	0.02			0.62	19223.6	0.17	14167.33	-4.03(0.00)	0.98(0.50)	0.42
Logistic	106887.59	2.27	0.08		0.61	16540.83	0.17	13080.22	-3.75(0.00)	0.98(0.50)	0.56
Gompertz	111234.35	0.22	0.06		0.60	16816.94	0.17	13256.20	-3.17(0.00)	0.98(0.44)	0.54
Richards	102993.00	652	19.0	708	0.68	15064.1	0.14	11384.06	-2.29(0.02)	0.98(0.59)	0.67

Table-3 *Characteristics of linear and non-linear models for the production of tapioca in Tamil Nadu*

Model	Parameter Estimates				Model evaluation				Run test	Shapiro-Wilk	Durbin Watson
	a	B	c	d	R Square	RMSE	MAPE	MAE			
Linear	877100.05	66865.87			0.51	930637	0.32	687867.6	-4.61(0.00)	0.97(0.32)	0.45
Logarithmic	-751585.6	1118634.9			0.55	891135.8	0.34	690747.9	-3.58(0.00)	0.95(0.03)	0.50
Inverse	2945742.1	-4342907.8			0.27	1136888	0.55	914631.9	-3.64(0.00)	0.95(0.00)	0.38
Quadratic	-380806.17	214854.83	-2959.78		0.67	765575.8	0.29	603675.3	-3.69(0.00)	0.94(0.02)	0.65
Cubic	1006691.6	-102239.68	12735.9	-209.28	0.79	608989.2	0.22	459108.9	-2.02(0.04)	0.95(0.04)	1.02
Compound	922296.98	1.03			0.61	1118692	0.35	829771.5	-4.60(0.00)	0.96(0.08)	0.32
Power	344867.65	0.62			0.79	892125.9	0.24	631154.2	-3.72(0.00)	0.96(0.10)	0.49
S	14.84	-2.77			0.50	1090899	0.37	813144.7	-4.19(0.00)	0.90(0.00)	0.33
Growth	13.73	0.03			0.61	1118692	0.35	829771.5	-4.60(0.00)	0.96(0.08)	0.32
Exponential	922296.98	0.03			0.61	1118692	0.35	829771.5	-4.60(0.00)	0.96(0.08)	0.32
Logistic	3643816.1	9.27	0.15		0.65	788722	0.23	576268.2	-2.02(0.04)	0.96(0.13)	0.62
Gompertz	3709943.1	1.07	0.10		0.63	806480.3	0.25	602271.9	-3.17(0.00)	0.96(0.13)	0.59
Richards	3589476.3	15.05	0.55	8.92	0.67	761677.3	0.23	560128.7	-2.89(0.00)	0.96(0.07)	0.66

Table-4 *Characteristics of linear and non-linear models for the Productivity of tapioca in Tamil Nadu*

Model	Parameter Estimates				Model evaluation				Run test	Shapiro-Wilk	Durbin Watson
	a	b	c	d	R Square	RMSE	MAPE	MAE			
Linear	21562.63	366.34			0.43	5935.05	0.2	4721.38	-4.49(0.00)	0.94(0.02)	0.55
Logarithmic	8893.95	7398.25			0.68	4450.33	0.13	3348.19	-3.04(0.00)	0.98(0.38)	0.99
Inverse	33914.88	-34937.49			0.49	5638.48	0.19	4363.29	-3.17(0.00)	0.99(0.86)	0.81
Quadratic	9820.53	1747.76	-27.63		0.83	3290.70	0.09	2567.04	-1.13(0.26)	0.98(0.55)	1.74
Cubic	9256.07	1876.76	-34.01	0.09	0.83	3285.28	0.09	2593.02	-1.68(0.09)	0.98(0.71)	1.75
Compound	19986.86	1.02			0.43	6625.55	0.22	5587.16	-5.71(0.00)	0.93(0.00)	0.44
Power	11388.18	0.32			0.72	4972.24	0.14	3908.30	-4.19(0.00)	0.95(0.06)	0.78
S	10.43	-1.59			0.57	5038.17	0.16	3944.15	-2.60(0.01)	0.99(0.87)	0.82
Growth	9.90	0.02			0.43	6625.55	0.22	5587.16	-5.71(0.00)	0.93(0.00)	0.44
Exponential	19986.86	0.02			0.43	6625.55	0.22	5587.16	-5.71(0.00)	0.93(0.00)	0.44
Logistic	35129.35	2.86	0.21		0.78	3735.75	0.11	2940.80	-2.60(0.01)	0.98(0.77)	1.35
Gompertz	35286.16	0.45	0.17		0.77	3742.53	0.11	2931.85	-1.97(0.05)	0.99(0.79)	1.35
Richards	35178.05	0.35	0.19	0.62	0.78	3733.26	0.11	2936.76	-1.97(0.05)	0.99(0.80)	1.35

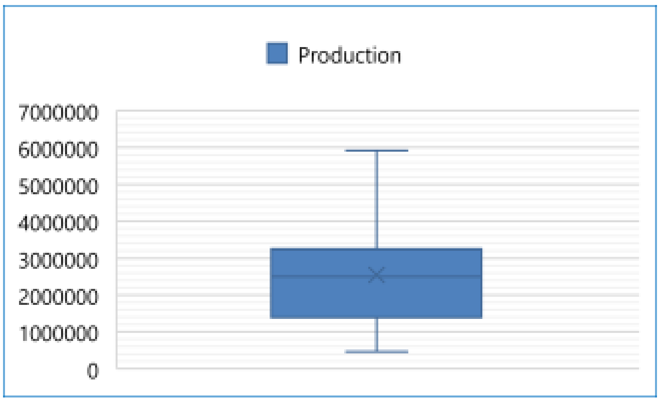

Fig-2 Boxplot of Production

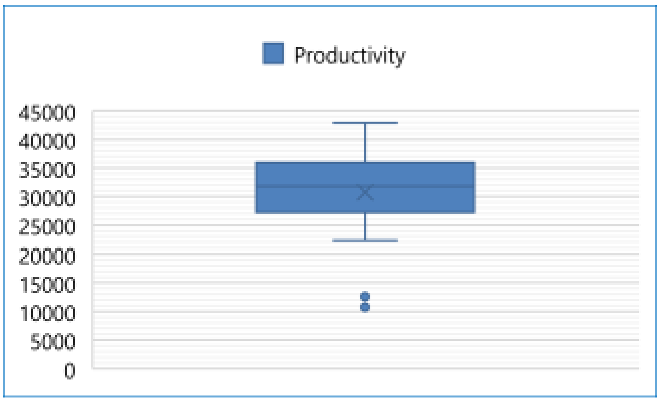

Fig-3 Boxplot of Productivity

Trends in Production

The production of Tapioca ranges from 4,66,660 kilograms to 59,12,307 kilograms with an average of 25,48,746.67 Kgs and the coefficient of variation is 52.59 percentage. The skewness and kurtosis values are 0.54 and -0.19. The value indicates that the time series data was slightly platykurtic and positively skewed.
Different linear and non-linear models were fitted for the production of Tapioca in Tamil Nadu. All the 13 models were compared, and the Shapiro Wilks test and the Run test showed that data does not follow the normality ($p<0.05$) and not randomly distributed. So, none of the models was selected as the best model for the production of the Tapioca.

Trends in productivity

The productivity of Tapioca ranges from 10715.26 kg/ha to 42900.46kg/ha with the mean and standard deviation of 30721.16 kg/ha and 7959.83 kg/ha with the coefficient of variation of 28 percent. The skewness and kurtosis values were -1.03 and 0.85. These values showed that the data was negatively skewed and slightly leptokurtic [Table-1]. All the 13 different linear and non-linear models fitted for the productivity of Tapioca in Tamil Nadu are listed in [Table-4]. The best-fitted model was selected based on the R^2, MAPE, RMSE and MAE values. In the productivity of Tapioca, two models were selected as the best model. The Cubic model was selected as the best model based on the R^2 (0.83) and RMSE (3285.28) values; the quadratic model was selected as the best model based on the MAPE (0.09) and MAE (2567.04) values. These two models were given smaller variations in the goodness of fit measures. Major variation caused by the R^2 and RMSE values cubic model considered as the best-fitted model for the productivity of Tapioca in Tamil Nadu. The parameter values are a (9256.07), b (1876.76), c (-34.01) and d (0.09).
The equation of the cubic model given below
Y=9256.07+1876.76X-34.01 X^2+0.09 X^3………………… (2)

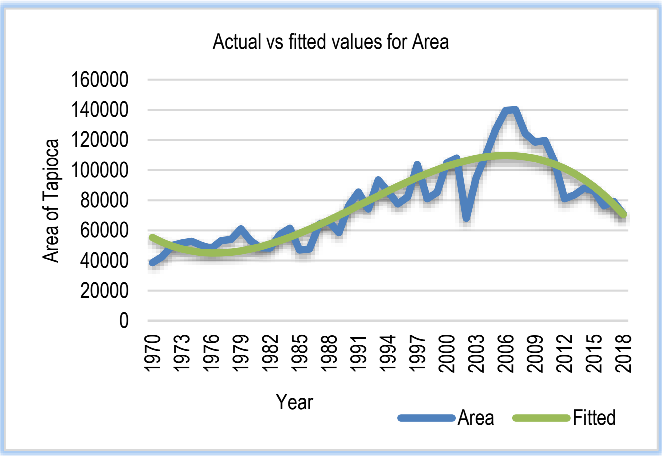

Fig-4 Plot fit of cubic and actual data of Productivity of Tapioca in Tamil Nadu

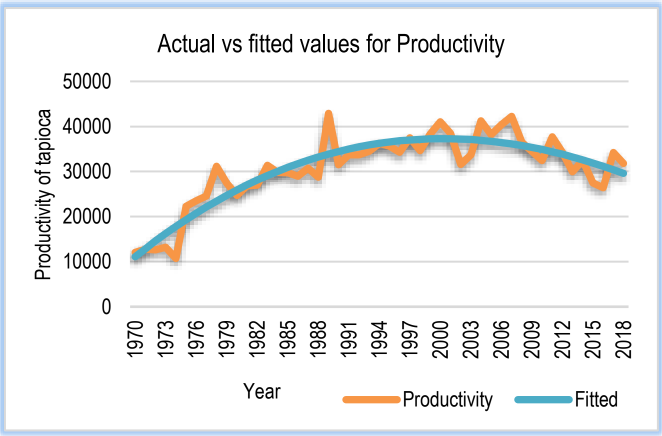

Fig-5 Plot fit of cubic and actual data of Productivity of Tapioca in Tamil Nadu

Future prediction

Using the best-fitted models, we can propose long-term plans concerning the area and productivity of Tapioca. The cubic model was used for the future prediction of values from 2019-20 to 2022-23.

Table-5 *Future prediction values of Area and Productivity*

Year	Area (ha)	Productivity (kg/ha)
2019-20	62794.83	28702.05
2020-21	54279.77	27794.81
2021-22	44915.70	26845.60
2022-23	34673.07	25854.92

Moreover, the predicted result showed a decreasing trend in area and productivity of Tapioca. This may be due to the prevailing unremunerative price of the Tapioca, which could have resulted in the decreasing trend over the years.

Conclusion

The present investigation was made to evaluate the growth rate of the area, production and productivity of Tapioca in Tamil Nadu. Due to some outliers present in the data and the skewness value was about more than ±1. The productivity data showed positive kurtosis value with the leptokurtic remaining area and production data with the negative values with platykurtic. 13 different linear and non-linear models have been employed for the study period of 1970-71 to 2018-19. Concerning the prediction of area cubic model selected as the best-fitted model with the highest R^2 and lowest RMSE, MAPE, and values. Regarding the production, none of the models was selected as the best-fitted model due to the presence of the errors, the normality and randomness criteria were not fulfilled in any of the models. Also, for the productivity cubic model was selected as the best model based on the R^2 and RMSE values while Quadratic model selected as the best-fitted model based on the MAPE and MAE values.
Using the best-fitted model future prediction has been carried out from 2019-20 to 2022-23. The predicted values showed that the decreasing trend was prevailing in the area and productivity of Tapioca. It could be attributed to the pricing policies and preference for alternate crops.

Application of research: Study shows the annual growth rate of the area, production and productivity were calculated, and the values were 1.246, 3.263 and 1.992 percentage. Production got a higher annual growth rate when compared to the area and productivity. Because of the combined effect of the area and productivity of the tapioca, production showed the maximum annual growth rate.

Research Category: Agricultural Statistics

Abbreviations: RMSE – Root mean square error, p value - Probability value
MAPE – Mean absolute percentage error, ha - Hectare
MAE - Mean absolute error, SD - Standard deviation
CV - Coefficient of variation, SW test - Shapiro-Wilks test
DW test - Durbin Watson test, R^2 - Coefficient of determination
H_o - Null hypothesis, H_1 - Alternative hypothesis

Acknowledgement / Funding: Authors are thankful to Department of Physical Science and Information Technology, Tamil Nadu Agricultural University, Coimbatore, 641 003, Tamil Nadu, India. Authors are also thankful to Department of Basic and Social Sciences, Forestry College and Research Institute, Mettupalayam, 641 301, Tamil Nadu Agricultural University, Coimbatore, 641 003, Tamil Nadu, India

**** Research Guide or Chairperson of research: Dr M Vijaya Bhama**
University: Tamil Nadu Agricultural University, Coimbatore, 641003, India
Research project name or number: MSc Thesis

Author Contributions: All authors equally contributed

Author statement: All authors read, reviewed, agreed and approved the final manuscript. Note-All authors agreed that- Written informed consent was obtained from all participants prior to publish / enrolment

Study area / Sample Collection: Tapioca area, Tamil Nadu. Time series crop data were collected from the season and crop report, Directorate of Economics and Statistics, Chennai, for the period of 1970-71 to 2018-19.

Cultivar / Variety / Breed name: Tapioca (*Manihot esculenta* Crantz)

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
Ethical Committee Approval Number: Nil

References

[1] Rawel H.M. and Kroll J. (2003) *Deutsche Lebensmittel-Rundschau*, 99(3), 102-108.
[2] Prajneshu and Chandran K.P. (2005) *Agricultural Economics Research Review*, 18, 317-324.
[3] Yadav R.L. and Kalola A.D. (2016) *Indian Journal of Economics and Development*, 4, 4.
[4] Rajan M.S. and Palanivel M. (2018) *International Journal of Statistical Distributions and Applications*, 4(1), 1.
[5] Pal S. and Mazumdar D. (2015) *Journal of Crop and Weed*, 11, 67-70.
[6] Panwar S., Kumar A., Singh K.N., Farooqi S. and Rathore A. (2012) *Indian Journal of Agricultural Sciences*, 82(12), 1051-1054.
[7] Kumar N.S., Joseph B. and Jaslam P.K.M. (2017) *International Journal of Advance Research, Ideas and Innovations in Technology*, 4(1), 446-448.
[8] Topal M., Ozdemir M., Aksakal V.E.C.İ.H.İ., Yildiz N.E.C.A.T.İ. and Dogru U. (2004) *Small Ruminant Research*, 55(1-3), 229-232.
[9] Legg J., Somado E.A., Barker I., Beach L., Ceballos H., Cuellar W., Elkhoury W., Gerling D., Helsen J., Hershey C. and Jarvis A. (2014) *Food Security*, 6(2), 231- 248.
[10] Singh M. and Supriya K. (2017) *Int. J. Curr. Microbiol. App. Sci*, 6(7), 2295-2301.