www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(5): 755-759

www.thepharmajournal.com
Received: 20-03-2022
Accepted: 25-04-2022

Evaluation of linear statistical models for predicting area, production and productivity of Sapota in Gujarat

Prity Kumari, DJ Parmar, Sathish Kumar M, AB Mahera and YA Lad

Prity Kumari
Assistant Professor and Head, Department of Basic Science, College of Horticulture, Anand Agricultural University, Anand, Gujarat, India

DJ Parmar
Associate Professor, Department of Agricultural Statistics, B. A. College of Agriculture, Anand Agricultural University, Anand, Gujarat, India

Sathish Kumar M
Ph.D., Research Scholar, Department of Agribusiness Management, International Agri-Business Management Institute, Anand Agricultural University, Anand, Gujarat, India

AB Mahera
Assistant Professor & Head, Department of Marketing Management, International Agribusiness Management Institute, Anand Agricultural University, Anand, Gujarat, India

YA Lad
Associate Professor and Head, Department of HRM & PM, International Agribusiness Management Institute, Anand Agricultural University, Anand, Gujarat, India

Corresponding Author
Prity Kumari
Assistant Professor and Head, Department of Basic Science, College of Horticulture, Anand Agricultural University, Anand, Gujarat, India

Abstract
The current study compares the performance of linear statistical models for predicting area, production, and productivity of Sapota in Gujarat. Time series secondary data were considered for the period of 1991-92 to 2016-17 from Directorate of Horticulture, Government of Gujarat and the data were analysed through R Studio (version 3.5.2) software. In the present investigation, comparison of ARIMA model with exponential smoothing model was done and ARIMA was shown to be the most effective in explaining the area, production, and productivity of sapota, with forecasted value for the year 2017-18 30.52 ('000' Ha.), 335.52 ('000' MT), and 10.86 (MT/Ha.) for respectively.

Keywords: Evaluation, linear, predicting, Sapota, *Achras zapota*

Introduction
Fruits are vital for a balanced diet as those are high in vitamins, minerals, and fiber. India has huge potential for fruit cultivation due to its diverse weather conditions. India is the second-largest fruit producer after China. Sapota (*Achras zapota*), also known as chiku, is mostly grown in India for its fruit, but it is also produced commercially in South-East Mexico, Guatemala, and other countries. In india, Gujarat, Maharashtra, Karnataka, Tamil Nadu, Andhra Pradesh, and Kerala are the main producers of sapota. The fruit has a good amount of digestible sugar (15-20%) as well as protein, fat, fibre, and minerals (Ca, P and Fe.). Sapota production is ranked eighth among all fruits in India. Gujarat holds first position in production of sapota, with 278.87 thousand tonnes, accounting for 32.84 percent of total production during the period of 2020-21 (National Horticulture Board, 2020-21) [3]. This fruit is currently a substantial crop in India, but due to its short shelf life, there are certain issues with keeping the fruits. As a result, there is a lot of scope for further processing this fruit to make value-added goods like jams, jellies, squash, and so on. In 2019-20, exports accounted for barely 0.6 percent of total production. Because of poor post-harvest processes, transportation procedures, a lack of suitable storage facilities, and obsolete handling practices, sapota exports are low in volume (National Horticulture Board, 2019-20) [2]. Therefore there is need to make an appropriate policy reforms in order to manage these hurdles and forecasting could be the one of the way to get an advance information about it. For estimating crop yield, Hamjah (2014) [6] predicted significant fruit crop yield in Bangladesh using the Box- Jenkins ARIMA model. Hossain and Abdulla computed a time series analysis for pineapple production in Bangladesh (2015) [7]. Hossain (2016) [8] forecasted banana production in Bangladesh using a variety of statistical methodologies. In West Bengal, Dasyam *et al*. (2016) [5] used statistical modelling to estimate potato area, production, and productivity. Kumari *et al*., (2016) [10] forecasted pigeon pea yield in the Vanaransi region using a variety of statistical methods. In the Varanasi region, Kumari *et al*. (2017) [11] investigated forecasting models for predicting pigeon pea pod damage. Rathod and Mishra (2018) [13] assessed mango and banana yield in Karnataka using several statistical methods. Kumar and Kumari (2021) forecasted the acreage, output, and productivity of sapota in Gujarat. Kumar *et al*. (2022) [9] analysed the area, production, and productivity of minor millet in India. The trend of maize acreage, production, and productivity in India was investigated by Unjia *et al*., (2021) [14]. Yield forecasting is a method of assisting policy decisions for fruit crops in order to boost farmer confidence in their socioeconomic concerns. Modeling and projecting the area, production and productivity of the sapota fruit crop through time is thus quite useful. Therefore, in this investigation, Autoregressive Integrated Moving Average Model (ARIMA) and Exponential Smoothing models were compared to forecast the future value of area, production and productivity of the sapota in Gujarat.

Material and Methods

Source of data

From 1991-92 to 2016-17, time series secondary data on sapota area, production, and productivity in Gujarat were obtained from Directorate of Horticulture, Gujarat.

Analytical framework

Exponential smoothing (ES) and Autoregressive integrated moving average (ARIMA), two time series forecasting models (linear and nonlinear), were employed in this study to examine their ability to predict future behaviour of sapota area, production, and productivity in Gujarat. R Studio version 3.5.2 was used to conduct the analysis.

Exponential Smoothing (ES) model

Smoothing techniques are used to reduce irregularities (random fluctuations) in time series data. One of the most successful univariate time series forecasting technique is the exponential smoothing (ES) to produce a smoothed time series. In this technique, forecasts are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. Exponential smoothing method is classified according to the type of component (trend and seasonality) presented in the time series data. In the present study, based on time series data, only two exponential smoothing methods are used i.e., simple exponential and double exponential smoothing technique.

1. Simple exponential smoothing (SES)

This method is suitable for forecasting data with no trend or seasonal pattern, although the mean of the data may be changing slowly over time. Forecasts are calculated by taking weighted averages of most recent observation and most recent forecast, where the weights decrease exponentially as observations come from further in the past.

Forecast equation $\hat{y}_{t+1/t} = l_t$

Level equation $l_t = \alpha y_t + (1-\alpha) l_{t-1}$

Simple exponential smoothing has a flat forecast function, and therefore for longer forecast horizons,

$\hat{y}_{t+h/t} = \hat{y}_{t+1/t} = l_t$

2. Holt's linear trend(double) exponential smoothing method

Holt (1957) extended simple exponential smoothing to allow forecasting for those data which exhibit trend. This method involves a forecast equation and two smoothing equations (one for the level and one for the trend):

Forecast equation $\hat{y}_{t+1/t} = l_t + h b_t$

Level equation $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$

Trend equation $b_t = \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1}$

Where

y_t, $\hat{y}_t$ are observed and predicted value of series at time t

l_t and b_t are estimate of the level and trend (slope) of the series at time t

α, β are the smoothing parameter for the level and trend, $0 \le \alpha, \beta \le 1$.

Initialisation

The application of every exponential smoothing method requires the initialisation of the smoothing process. For simple exponential smoothing we need to specify an initial value for the level, l_0. Similarly double exponential smoothing involves initial value trend component b_0 also.

Model	Initial Values
Single	$L_0 = y_1$
Double	$L_0 = y_1, b_0 = y_2 - y_1$

In exponential smoothing, the method for obtaining the optimal values of smoothing parameters α and β is an iterative process which is chosen either by trial-and-error method or by some software like *MINITAB*, E Views, SPSS etc. which use an algorithm to select the value of the weights that minimizes mean square error for in-sample forecasts.

Autoregressive Integrated Moving Average (ARIMA) model

ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely-used approaches to time series forecasting, and provide complementary approaches to the problem. While exponential smoothing models were based on a description of trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

ARIMA is one of the most traditional methods of non-stationary time series analysis. Usually time series, showing trend or seasonal patterns are non-stationary in nature. In such cases, differencing and power transformations are often used to remove the trend and to make the series stationary.

Box-Jenkins ARIMA, has been successfully applied in many time series forecasting and is a good tool to develop empirical model which is linear combination of its own past values, past errors (also called shocks or innovations).

ARIMA model allows Y_t tobe explained by its past, or lagged values and stochastic error terms. The non-seasonal ARIMA (p, d, q) model can be written as:

If $w_t = \nabla^d y_t = (1-B)^d y_t$ then

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Where

P: order of the autoregressive part;

d: degree of differencing involved;

q: order of the moving average part.

w_t & ε_t: Differenced data series and white noise

ϕ & θ: Autoregressive and moving average coefficient

The main stages in setting up a Box-Jenkins forecasting model are model identification, estimating the parameters, diagnostic checking of residual and forecasting. (Box and Jenkins 1970) [4].

Research Results

In this study, area, production and productivity (yield) of sapota crop were analyzed by Exponential smoothing model and ARIMA model. The empirical findings of sapota fruit crop are as follow:

Forecasting of area for sapota
Fig. 1 illustrate chart series of area dataset for sapotafrom 1991-92 to 2016-17. Also, the characteristics (basic statistics) of the data set used were presented in the Table 1.

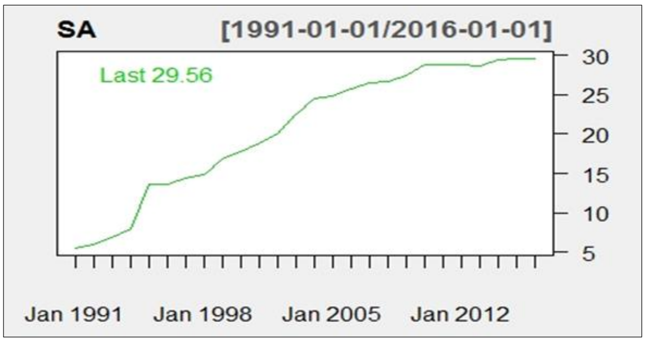

Fig 1: Area (In ' 000 Hectare) under sapotain Gujarat

Table 1: Summary statistics of sapota area

n.obs.	Min	Max	Mean	Median	Sem	Variance	Stdev	Skewness	Kurts
26	5.5	29.56	20.71	23.56	1.60	66.60	8.16	-0.53	-1.17

Exponential smoothing (ES) model: In case of fitting exponential smoothing model, the performance of ETS (A, A_d, N) Additive damped trend method was found to be the best out of all. The results were shown in Table 2.

Table 2: ETS (A, A_d, N) model parameters for area of sapota

Model	Parameter estimation	
Additive damped trend	Estimate (S.E.)	Sig.
Alpha (Level)	0.61 (0.12)	<0.01
Beta (Trend)	1e-04 (0.13)	NS
Phi (Trend damping factor)	0.94 (0.008)	<0.01
Forecast Value 2017-18 (C.I.)	30.23 (27.94 to 32.52)	
Fit Statistics		
AIC	Box-Ljungtestresid_fit_pvalue	
99.25	0.46	

Table 2 shows that the estimate of alpha and phi were found to be significant only. Also, residual autocorrelation was non-significant as per Box-Ljungtest statistics probability value 0.46. The forecasted value of sapota area in Gujarat for the year 2017-18 was obtained as 30.23(‘000’ Hectares) with confidence interval 27.94 to 35.52.

Autoregressive Integrated Moving Average (ARIMA) Model: In case of fitting ARIMA model, out of various ARIMA models with different value of p, d and q, the performance of ARIMA (0,1,0) with drift was found to be the best. The results were given in Table 3.

Table 3: ARIMA (0,1,0) model parameters for area of sapota

Model	Parameter estimation	
ARIMA (0,1,0) with drift	Estimate (S.E.)	Sig.
Drift	0.96 (0.23)	<0.01
Forecast Value 2017-18 (C.I.)	30.52 (28.16 to 32.88)	
Fit Statistics		
AIC	Box-Ljungtestresid_fit_pvalue	
83.19	0.70	
Lagrange-Multiplier test		
order	LM test statistics	p.value
4	21.86	<0.01

Table 3 reveals that the estimate of drift (constant)was found to be statistically significant. Also, residual autocorrelation was non-significant as per Box-Ljungtest statistics probability value 0.70. The forecasted value of sapota area in Gujarat for the year 2017-18 by ARIMA (0,1,0) with drift was obtained as 30.52(‘000’ Hectares) with confidence interval 28.16 to 32.88.

Forecasting of production for sapota
Fig. 2 illustrate chart series of production dataset for sapotafrom 1991-92 to 2016-17. Also, the characteristics (basic statistics) of the data set used were presented in the Table 4.

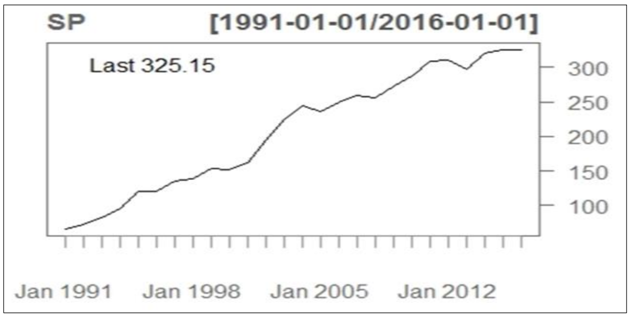

Fig 2: Production (In ‘000 MT) under sapota in Gujarat

Table 4: Summary statistics of sapota production

N.ob	Min	Max	Mean	Median	Sem	Variance	Stdev	Skewness	Kurtosis
26	66	325.15	208.15	229.85	17.05	7558.89	86.94	-0.16	-1.49

Exponential smoothing (ES) model
In case of fitting exponential smoothing model, the performance of ETS (M, A, N) holt exponential method was found to be the best out of all. The results were shown in Table 5.

Table 5: ETS (M, A, N) model parameters for production of sapota

Model	Parameter estimation	
Holt ES	Estimate (S.E.)	Sig.
Alpha (Level)	0.90 (0.21)	<0.01
Beta (Trend)	1e-04 (0.07)	NS
Forecast Value 2017-18 (C.I.)	337.44 (296.28 to 378.60)	
Fit Statistics		
AIC	Box-Ljungtestresid_fit_pvalue	
218.21	0.38	

Table 5 shows that the estimate of alpha was found to be significant. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.38. The forecasted value of sapota production in Gujarat for the year 2017-18 was obtained as 337.44 ('000' MT) with confidence interval 296.28 to 378.60.

Autoregressive Integrated Moving Average (ARIMA) Model

In case of fitting ARIMA model, out of various ARIMA models with different value of p, d and q, the performance of ARIMA (0,1,0) with drift was found to be the best. The results were given in Table 6.

Table 6: ARIMA (0,1,0) with drift model parameters for production of sapota

Model	**Parameter estimation**	
ARIMA (0,1,0) with drift	Estimate (S.E.)	Sig.
Drift	10.36 (2.31)	<0.01
Forecast Value 2017-18 (C.I.)	335.52 (312.45 to 358.57)	
Fit Statistics		
AIC	Box-Ljungtestresid_fit_pvalue	
197.19	0.39	
Lagrange-Multiplier test:		
order	LM test statistics	p.value
4	0.81	NS

Table 6 reveals that the estimate of drift (constant) as found to be statistically significant. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.81. The forecasted value of sapota production in Gujarat for the year 2017-18 by ARIMA (0,1,0) with drift was obtained as 335.52('000' MT) with confidence interval 312.45 to 358.57.

Forecasting of productivity for sapota

Fig. 3 illustrate chart series of productivity dataset for sapotafrom 1991-92 to 2016-17. Also, the characteristics (basic statistics) of the data set used were presented in the Table 7.

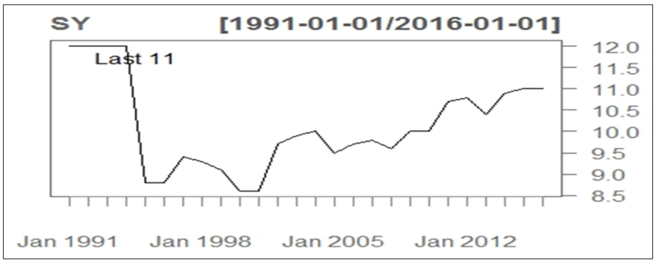

Fig 3: Productivity (In MT/Ha.) under sapota in Gujarat

Table 7: Summary statistics of sapota productivity

N.obs.	Min.	Max.	Mean	Median	Sem	Variance	Stdev	Skewness	Kurtosis
26	8.6	12	10.14	9.95	0.21	1.15	1.07	0.41	-0.97

Exponential smoothing (ES) model

In case of fitting exponential smoothing model, the performance of ETS (M, N, N) simple exponential smoothing method was found to be the best out of all. The results were shown in Table 8.

Table 8: ETS (A, N, N) model parameters for productivity of sapota

Model	**Parameter estimation**	
Simple exponential smoothing	Estimate (S.E.)	Sig.
Alpha (Level)	0.93 (0.20)	<0.01
Forecast Value 2017-18 (C.I.)	11.00 (9.53 to 12.46)	
Fit Statistics		
AIC	Box-Ljungtestresid_fit_pvalue	
69.00	0.22	

Table 8 shows that the estimate of alpha was found to be significant. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.22. The forecasted value of sapota productivity in Gujarat for the year 2017-18 was obtained as 11.00 ('000' MT) with confidence interval 9.53 to 12.46.

Autoregressive Integrated Moving Average (ARIMA) Model

In case of fitting ARIMA model, out of various ARIMA models with different value of p, d and q, the performance of ARIMA (1,0,0) was found to be the best. The results were given in Table 9.

Table 9: ARIMA (1,0,0) model parameters for productivity of sapota

Model	**Parameter estimation**	
ARIMA (1,0,0)	Estimate (S.E.)	Sig.
Intercept	10.42(0.56)	<0.01
AR1	0.77 (0.12)	<0.01
Forecast Value 2017-18 (C.I.)	10.86(9.45 to 12.28)	
Fit Statistics		
AIC	Box-Ljungtestresid_fit_pvalue	
61.82	0.54	
Lagrange-Multiplier test:		
order	LM test statistics	p.value
4	37.33	<0.01

Table 9 reveals that the estimates of both parameters were found to be statistically significant. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.54. The forecasted value of sapota productivity in Gujarat for the year 2017-18 by ARIMA (1,0,0) was obtained as 10.86(MT/Ha.) with confidence interval 9.45 to 12.28.

Table 10: Performance of models for sapota

Forecasting model for Sapota		**Area**	**Production**	**Productivity**
ES	Model	Additive damped trend	Holt	Simple
	AIC	99.25	218.21	69.00
	Forecast	30.23	337.44	11.00
	C.I.	27.94 to 32.52	296.28 to 378.60	9.53 to 12.46
ARIMA	Model	ARIMA (0,1,0) with drift	ARIMA (0,1,0) with drift	ARIMA (1,0,0)
	AIC	83.19	197.19	61.82
	Forecast	30.52	335.52	10.86
	C.I.	28.16 to 32.88	312.45 to 358.57	9.45 to 12.28

1. Highlighted forecasted values have the least AIC

Table 10 shows the performance of models for predicting area, production and productivity of sapota. Area, production and productivity of sapota was best explained by ARIMA model with forecasted value for 2017-18, 30.52 (‘000’ Ha.), 335.52 (‘000’ MT) and 10.86 (MT/Ha.) respectively.

Conclusion

Present investigation focuses on comparison of Exponential smoothing (ES) with Autoregressive integrated moving average to build best statistical models for projecting citrus area, production, and productivity (ARIMA). Based on the lowest error measures i.e. AIC, ARIMA was found to produce better result as compared to exponential models with. next year forecast as 30.52 ('000' Ha.), 335.52 ('000' MT), and 10.86 (MT/Ha.) correspondingly for area, production, and productivity of Sapota.

References

1. Anonymous. Directorate of Horticulture, Agriculture, Farmers Welfare and Co-Operation Department, Government of Gujarat.
2. Anonymous, Annual Report, National Horticulture Board, Ministry of Agriculture and Farmers Welfare, Government of India, 2019-20.
3. Anonymous, Annual Report. National Horticulture Board, Ministry of Agriculture and Farmers Welfare, Government of India, 2020-21.
4. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970.
5. Dasyam R, Bhattacharyya B, Mishra P. Statistical modelling to area, production and yield of potato in west Bengal. International Journal of Agriculture Sciences. 2016;8(53):2782-2787.
6. Hamjah MA. Forecasting major fruit crops productions in Bangladesh using Box-Jenkins ARIMA model. Journal of Economics and Sustainable Development. 2014;5(7):96-107.
7. Hossain MM, Abdulla F. A Time Series Analysis for the Pineapple Production in Bangladesh. Jahangirnagar University Journal of Science. 2015;38(2):49-59.
8. Hossain MM, Abdulla F, Majumder AK. Forecasting of Banana Production in Bangladesh. American Journal of Agricultural and Biological Sciences. 2016;11(2):93-99.
9. Kumar MS, Lad YA, Mahera AB. Trend analysis of area, production and productivity of minor millets in India. Biological Forum- An International Journal. 2022;14(2):14-18.
10. Kumari P, Mishra GC, Srivastava CP. Statistical models for forecasting pigeon pea yield in Varanasi region. Journal of Agrometeorology. 2016;18(2):306-310.
11. Kumari P, Mishra GC, Srivastava CP. Forecasting models for predicting damage of pigeon pea in Varanasi region. Journal of Agrometeorology. 2017;19(3):265-269.
12. Kumar MS, Kumari P. Artificial neural network model for predicting area, production and productivity of sapota in Gujarat. International Journal of Agricultural Sciences. 2021;13(10):10909-10912.
13. Rathod S, Mishra CG. Statistical models for forecasting Mango and Banana yield of Karnataka, India. Journal of Agricultural Science and Technology. 2018;20(3):803-816.
14. Unjia YB, Lad YA, Kumar MS, Mahera AB. Trend analysis of area, production and productivity of maize in India. International Journal of Agricultural Sciences. 2021;13(9):10880-10882.