

Comparison of Statistical Models for Prediction Area, Production and Yield of Citrus in Gujarat

Prity Kumari^{1*}, D.J. Parmar², Sathish Kumar M.³, A.B. Mahera⁴ and Y.A. Lad⁵

¹Assistant Professor & Head, Department of Basic Science,

College of Horticulture, Anand Agricultural University, Anand (Gujarat), India.

²Associate Professor, Department of Agricultural Statistics,

B.A. College of Agriculture, Anand Agricultural University, Anand (Gujarat), India.

³Ph.D. Research Scholar (Agribusiness Management),

International Agri-Business Management Institute, Anand Agricultural University, Anand (Gujarat), India.

⁴Assistant Professor & Head, Department of Marketing Management,

International Agribusiness Management Institute, Anand Agricultural University, Anand (Gujarat), India.

⁵Associate Professor & Head, Department of HRM & PM,

International Agribusiness Management Institute, Anand Agricultural University, Anand (Gujarat), India.

(Corresponding author: Prity Kumari*)

(Received 17 February 2022, Accepted 26 April, 2022)

(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Present study deals with comparison of different statistical models to predict area, production and yield of citrus in Gujarat state. Secondary data on area, production and productivity of citrus in Gujarat over the period 1991-92 to 2016-17 were collected from Directorate of Horticulture, Government of Gujarat. Autoregressive Moving Average Models (ARIMA) and Exponential smoothing models were used to analyse the data through R Studio (version 3.5.2) software. The study revealed that ARIMA model was superior to explain the area, production and yield of citrus with forecasted values 46.39 ('000' Ha.), 606.55 ('000' MT) & 12.01 (MT/Ha.) respectively. The government will be advised by the study to adopt an appropriate model to predict data and make policy accordingly.

Keywords: Forecasting, area, production, productivity, citrus, exponential smoothing model, ARIMA.

INTRODUCTION

Fruits are significant for a healthy diet since they are rich in vitamins, minerals, and fibre. It improves farmers' socioeconomic conditions and has also become a source of better income for many underprivileged groups. Citrus fruits are one of the healthiest fruits which are nutritional powerhouses.

It is mostly produced in Spain, Brazil, China, the United States, Mexico, and India. Citrus fruit's export from India is a booming enterprise, owing to the country's favourable climate and soil conditions. The mandarin (*Citrus reticulata*), sweet orange (*Citrus sinensis*), and acid lime (*Citrus aurantifolia*) are the most important commercial citrus species in India, accounting for 41, 23 and 23 percent of all citrus fruits produced in the country respectively. After mango and banana, India's citrus business is the country's third largest fruit industry. India is the ninth largest orange producer in the world, accounting for 3% of worldwide orange production. Gujarat is holding third rank in terms of citrus production (National Horticulture Board, 2020-21). Many academics have been utilising statistical models to forecast area, production, and productivity of various crops in recent years.

Dhaikar and Rode (2014) used artificial neural network approach for predicting agricultural crop yield. Guo and Xue (2014) applied artificial neural network for forecasting crop yield and made comparison between spatial and temporal models. Hamjah (2014) used the Box-Jenkins ARIMA model to anticipate significant fruit crop production in Bangladesh. Time series analysis for pineapple production in Bangladesh was computed by Hossain and Abdulla (2015). Hossain *et al.* (2016) used various statistical methods to anticipate banana production in Bangladesh. Dasyam *et al.* (2016) applied statistical modelling to predict area, production and productivity of potato in West Bengal. Kumari *et al.* (2016) used various statistical models to forecast pigeon pea yield in the Vanaransi region. Kumari *et al.* (2017) looked at forecasting models for predicting pigeon pea pod damage in the Varanasi region. Using various statistical models, Rathod and Mishra (2018) estimated mango and banana yield in Karnataka. Aguilar *et al.* (2020) used artificial neural network model for predicting greenhouse tomato yield and aerial dry matter. In Gujarat, Kumar and Kumari (2021) predicted the area, production, and productivity of sapota. Computed a trend study of minor millet area, production, and productivity in India. Unjia *et al.*

(2021) looked into the trend of maize area, production, and productivity in India. Yield forecasting is a way to support policy decisions for fruit crops in order to introduce confidence amongst the farming community for their socio-economic issues. Thus, modeling and forecasting the area, production and yield of citrus fruit crop over the years are of much practical importance. The present study was taken up in order to compare the performance of different statistical models *i.e.*, Autoregressive Moving Average Models (ARIMA) and Exponential smoothing for predicting area, production and yield of citrus in Gujarat. Traditional forecasting methods in India continue to cause problems in predicting crop area, production, and productivity, as well as market price of agricultural commodities. This research has a broad scope in that it aims to develop adequate statistical models for predicting data on area, production, and productivity of various crops, with the proper model assisting in the execution of some forecasting policies.

METHODOLOGY

Time series secondary data on area, production and yield of Citrus in Gujarat were collected from Directorate of Horticulture, Govt. of Gujarat from 1991-92 to 2016-17.

Analytical framework: In the present study, time series forecasting models *i.e.*, Exponential smoothing (ES) and Autoregressive integrated moving average (ARIMA), were used to compare their ability for predicting future behavior of area, production and yield of citrus in Gujarat. Analysis was done by RStudio version 3.5.2 software. Time series models used in the present investigation were:

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.

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.

$$\begin{aligned} \text{Forecast equation} & \hat{y}_{t+h/t} = l_t \\ \text{Level equation} & l_t = \alpha y_t + (1 - \alpha)l_{t-1} \end{aligned}$$

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

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

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):

$$\begin{aligned} \text{Forecast equation} & \hat{y}_{t+h/t} = l_t + hb_t \\ \text{Level equation} & l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ \text{Trend equation} & b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \end{aligned}$$

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, $0 < \alpha, \beta \leq 1$

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 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 (Box and Jenkins 1970) 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 to be explained by its past, or lagged values and stochastic error terms. The non-seasonal ARIMA (p, d, q) model can be written as:

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

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

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).

RESULTS

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

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

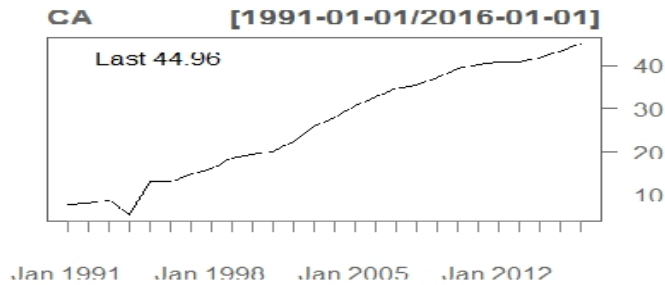


Fig. 4. Area (In '000 Hectare) under citrus in Gujarat.

Table 1: Summary statistics of citrus area.

N.obs.	Min.	Max.	Mean	Median	SEm	Variance	Stdev	Skewness	Kurts
26	5.4	44.96	26.21	26.91	2.53	166.81	12.92	-0.10	-1.54

Exponential smoothing (ES) model: In case of fitting exponential smoothing model, the performance of

ETS(A, A,N) holt linear method was found to be the best out of all. The results were shown in Table 2.

Table 2: ETS (A, A, N) model parameters for area of Citrus.

Model		Parameter estimation	
Holt linear ES		Estimate(S.E.)	
Alpha (Level)		0.56 (0.18)	Sig. <0.01
Beta (Trend)		1e-04 (0.04)	NS
Forecast Value 2017-18 (C.I.)		46.50 (43.05 to 49.96)	
Fit Statistics			
AIC		Box-Ljung testresid_fit_p value	
119.87		0.24	

Table 2 shows that the estimate of alpha was found to be significant only. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.24. The forecasted value of Citrus area in Gujarat for the year 2017-18 was obtained as 46.50('000' Hectares) with confidence interval 43.05to 49.96.

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, 1, 0) with drift was found to be the best. The results were given in Table 3.

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

Model		Parameter estimation	
ARIMA (1,1,0) with drift		Estimate(S.E.)	
Drift		1.50 (0.24)	Sig. <0.01
AR1		-0.36 (0.18)	<0.05
Forecast Value 2017-18 (C.I.)		46.39 (43.06 to 49.72)	
Fit Statistics			
AIC		Box-Ljungtestresid_fit_p value	
101.55		0.92	
Lagrange-Multiplier test:			
order		LM test statistics	
4		17.14	
		p. value	
		<0.01	

Table 3 reveals that the estimates of all parameters were found to be statistically significant. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.92. The forecasted value of citrus area in Gujarat for the year 2017-18 by ARIMA (1, 1, 0) with drift was obtained as 46.39

(‘000’ Hectares) with confidence interval 43.06 to 49.72.

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

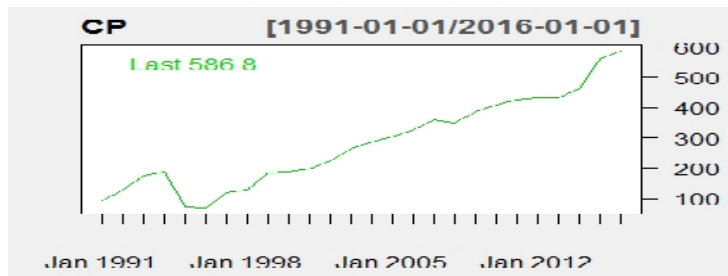


Fig. 2. Production (In ‘000 MT) under citrus in Gujarat.

Table 4: Summary statistics of citrus production.

N. obs.	Min	Max	Mean	Median	SEm	Variance	St. dev.	Skewness	Kurts
26	68.7	586.8	282.89	274.15	29.32	22353.53	149.51	0.31	-1.06

Exponential smoothing (ES) model: In case of fitting exponential smoothing model, the performance of

ETS(A,A,N) holt exponential method was found to be the best out of all. The results were shown in Table 5.

Table 5: ETS (A, A, N) model parameters for production of citrus.

Model		Parameter estimation	
Holt ES		Estimate(S.E.)	
Alpha (Level)		0.98 (0.21)	Sig.
Beta (Trend)		2.1e-05 (0.13)	NS
Forecast Value 2017-18 (C.I.)		606.37 (531.98 to 680.76)	
Fit Statistics			
AIC		Box-Ljung test resid_fit_pvalue	
279.46		0.48	

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 0.48. The forecasted value of Citrus production in Gujarat for the year 2017-18 was obtained as 606.37 (‘000’ MT) with confidence interval 531.98 to 680.76.

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 citrus.

Model		Parameter estimation	
ARIMA (0,1,0) with drift		Estimate(S.E.)	
Drift		19.75 (7.02)	Sig.
Forecast Value 2017-18 (C.I.)		606.55 (536.25 to 676.84)	
Fit Statistics			
AIC		Box-Ljungtestresid_fit_pvalue	
252.91		0.42	
Lagrange-Multiplier test:			
order	LM test statistics	p. value	
4	9.57	0.02	

Table 6 reveals that the estimate of alpha was found to be statistically significant. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.42. The forecasted value of citrus production in Gujarat for the year 2017-18 by ARIMA (0, 1, 0) with drift was obtained

as 606.55(‘000’ MT) with confidence interval 536.25 to 676.84.

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

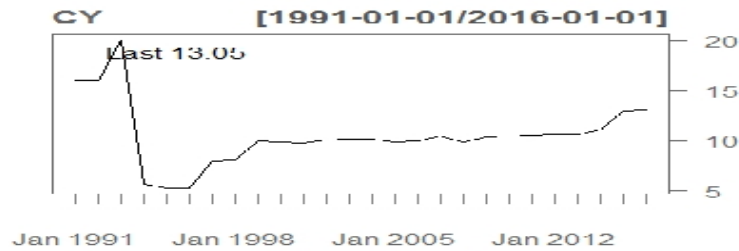


Fig. 3. Yield (In MT/Ha.) under citrus in Gujarat.

Table 7: Summary statistics of citrus yield.

N.obs.	Min	Max	Mean	Median	SEm	Variance	St. dev.	Skewness	Kurtosis
26	5.3	20	10.56	10.18	0.63	10.21	3.19	0.89	1.40

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 yield of Citrus.

Model		Parameter estimation	
Simple ES		Estimate(S.E.)	Sig.
Alpha (Level)		0.74 (0.19)	<0.01
Forecast Value 2017-18 (C.I.)		13.04 (8.00 to 18.09)	
Fit Statistics			
AIC	Box-Ljungtestresid_fit_pvalue		
124.90	0.37		

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 0.37. The forecasted value of citrus yield in Gujarat for the year 2017-18 was obtained as 13.04 ('000' MT) with confidence interval 8.00 to 18.09.

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) with constant was found to be the best. The results were given in Table 9.

Table 9: ARIMA (1, 0, 0) model parameters for yield of citrus.

Model		Parameter estimation	
ARIMA (1,0,0)		Estimate(S.E.)	Sig.
Intercept		10.86 (1.08)	<0.01
AR1		0.52 (0.17)	<0.01
Forecast Value 2017-18 (C.I.)		12.01(6.49 to 17.52)	
Fit Statistics			
AIC	Box-Ljungtestresid_fit_pvalue		
131.82	0.31		
Lagrange-Multiplier test:			
order	LM test statistics		p. value
4	0.01		<0.01

Table 9 reveals that the estimate of both parameters was found to be statistically significant. Also, residual autocorrelation was non-significant as per Box-Ljung test statistics probability value 0.31. The forecasted

value of citrus yield in Gujarat for the year 2017-18 by ARIMA (1, 0, 0) with constant was obtained as 12.01(MT/Ha.) with confidence interval 6.49 to 17.52.

Table 10: Performance of models for citrus.

Forecasting model for Citrus		Area	Production	Yield
ES	Model	Holt	Holt	Simple
	AIC	119.87	279.46	124.90
	Forecast	46.50	606.37	13.04
	C.I.	43.05 to 49.96	531.98 to 680.76	8.00 to 18.09
ARIMA	Model	ARIMA (1,1,0) with drift	ARIMA (0,1,0) with drift	ARIMA (1,0,0)
	AIC	101.55	252.91	131.82
	Forecast	46.39	606.55	12.01
	C.I.	43.06 to 49.72	536.25 to 676.84	6.49 o 17.52

*Highlighted forecasted values have the least AIC.

Table 10 shows the performance of different models for predicting area, production and yield of citrus. Area, production and yield of citrus was best explained by ARIMA model with forecasted values 46.39 ('000' Ha.), 606.55 ('000' MT) & 12.01 (MT/Ha.) respectively.

CONCLUSION

The present investigation was carried out to develop forecast and predict area, production and yield of citrus by Exponential smoothing (ES) and Autoregressive integrated moving average (ARIMA). The performance of these two models were compared and selected on the basis of best fit statistics measure. Out of these two, ARIMA model was best to explain the area, production and yield of citrus with forecasted values for 2017-18, 46.39 ('000' Ha.), 606.55 ('000' MT) & 12.01 (MT/Ha.) respectively.

Acknowledgement. Author thankful to College of Horticulture, Anand Agricultural University, Anand-388110, Gujarat, India.

Conflict of interest. None.

REFERENCES

- Aguilar, K. L., Mendoza, A. B., Morales, S. G. and Maldonado, A. J. (2020). Artificial neural network modelling of greenhouse tomato yield and aerial dry matter. *Agriculture*, 10(2): 2-14.
- Anonymous (2015). Directorate of Horticulture, Agriculture, Farmers Welfare and Co-Operation Department, Government of Gujarat.
- Anonymous (2020-21). Annual Report, National Horticulture Board, Ministry of Agriculture and Farmers Welfare, Government of India.
- Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.
- Dasyam, R., Bhattacharyya, B. and Mishra, P. (2016). Statistical modelling to area, production and yield of potato in west Bengal. *International Journal of Agriculture Sciences*, 8(53): 2782-2787.
- Dhaikar, S. S. and Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 2(1): 683-686.
- Guo, W. W. and Xue, H. (2014). Crop yield forecasting using artificial neural networks. A comparison between spatial and temporal models. *Mathematical problems in Engineering*, 8(3): 4-10.
- Hamjah, M. A. (2014). Forecasting major fruit crops productions in Bangladesh using Box-Jenkins ARIMA model. *Journal of Economics and Sustainable Development*, 5(7): 96-107.
- Hossain, M. M. and Abdulla, F. (2015). A Time Series Analysis for the Pineapple Production in Bangladesh. *Jahangirnagar University Journal of Science*, 38(2): 49-59.
- Hossain, M. M., Abdulla, F. and Majumder, A. K. (2016). Forecasting of Banana Production in Bangladesh. *American Journal of Agricultural and Biological Sciences*, 11(2): 93-99.
- Kumar, M. S. and Kumari, P. (2021). Artificial neural network model for predicting area, production and productivity of sapota in Gujarat. *International Journal of Agricultural Sciences*, 13(10): 10909-10912.
- Kumari, P., Mishra, G. C. and Srivastava, C. P. (2016). Statistical models for forecasting pigeon pea yield in Varanasi region. *Journal of Agrometeorology*, 18(2): 306-310.
- Kumari, P., Mishra, G. C. and Srivastava, C. P. (2017). Forecasting models for predicting damage of pigeon pea in Varanasi region. *Journal of Agrometeorology*, 19(3): 265-269.
- Rathod, S. and Mishra, C. G. (2018). Statistical models for forecasting Mango and Banana yield of Karnataka, India. *Journal of Agricultural Science and Technology*, 20(3): 803-816.
- Unjia, Y. B., Lad, Y. A., Kumar, M. S. and Mahera, A. B. (2021). Trend analysis of area, production and productivity of maize in India. *International Journal of Agricultural Sciences*, 13(9): 10880-10882.

How to cite this article: Prity Kumari, D.J. Parmar, Sathish Kumar M., A.B. Mahera and Y.A. Lad (2022). Comparison of Statistical Models for Prediction Area, Production and Yield of Citrus in Gujarat. *Biological Forum – An International Journal*, 14(2): 690-695.