I J T A © Serials Publications

National Academy of Agricultural Science (NAAS) Rating : 3.03

Statistical Models for Forecasting Citrus Area and Production in Himachal Pradesh

Ajit Sharma*, V. K. Chaudhary and Shilpa

ABSTRACT: *In this article, an attempt has been made to forecast citrus area and production using statistical time series modeling techniques viz; straight line, second degree parabola, exponential, modified exponential, gompertz and logistic. All the models were tried to the area and production under citrus crop by using time series data for last thirty three years (1980-81 to 2012-13). The models were further validated to forecast the area and production of citrus crop on the basis of R^2, Adjusted R^2 and coefficient of variation (in per cent). From the study, second degree parabola was found to be the best fitted model for citrus area as well as production. On the basis of these results, the present study is very helpful to the policy makers, planners, research workers and orchardists for making the future plans regarding citrus crop and can contributes a lot towards the growth and development of state economy.*

INTRODUCTION

Citrus fruits belong to the family Ructaceae that are grown all over the world and have numerous therapeutic properties like anticancer, anti-tumor and anti-inflammatory. It plays an important role in horticulture and industrial economy of the country. India ranks 6th position in the production of citrus fruit cultivation in the world. It occupies an area of 1042.0 MH in India with annual production of 10090.0 MT, out of which Himachal Pradesh having area 22.809 MH under the citrus and annual production is 412.39 MT during the year 2012-13(NHB, 2014). Himachal Pradesh is one of the important citrus producing state of the country. Citrus occupies a significant position among the commercial crops in Himachal Pradesh. A proper forecast of production of such important commercial crops is very important in an economic system. There is a close association between crop productions with prices. An unexpected decrease in production reduces marketable surplus and income of the farmers and leads to price rise.

Among all fruit crops grown in Himachal Pradesh, citrus is the major contributor to the farmer's economy as well as state's economy. Forecasting models are used to know the real behavior of past studies and predict the future behavior of present studies. Forecasting models are mostly used by the policy makers, planners, administrators and research workers for analyzing the time series data. Its results are further used for making the future planning's and policies for development of State/Nation. It enables us to study the past behavior of the phenomenon under consideration i.e., to determine the type and nature of variations in the data. It helps us to compare the changes in the values of different phenomenon at different time. Validation of prediction model(s) for fruit crops is very essential today. This study is also needed for better understanding and to know the suitability of the fruit crops to the farmers of the state in future, so that farmers and policy makers make a plan in advance for its fruit products. Statistical tests are used to carefully examine prior activities and then use these analyses to make informed predictions about future activities. Regardless of the statistical tests, data are examined in a systematic manner so that decisions can be made with some degree of certainty. The opportunity of using existing data to predict future outcomes is viewed as model building. That is to say, existing data are used to build a model of the future, with a predetermined degree of error built into the model (Thomas, 1998).

A statistical model is an equation or set of equations which represents the relationships among variables. Modelling may lead to less adhoc experimentation, as models sometimes make it easier to design experiments to answer particular questions

* Dr. Y. S. Parmar University of Horticulture and Forestry Nauni, Solan 173230 (HP), *E-mail: ajit.aziz.sharma@gmail.com*

or to discriminate between alternative mechanisms. Modelling provides powerful tools for investigating the dependence and nature of relationship among the variables of interest. The relationship among variables must be determined for the purpose of predicting the values of one or more variables on the basis of observation on other variables. Model building is currently applied in many fields i.e., Agriculture, Biometrics, Econometrics, Education, Meteorology, Industry, Horticulture and Forestry etc. (Prasad, 2010).

From the Statistical outline of Himachal Pradesh, 2014 "Directorate of Economics and Statistics Department Navbhar, Shimla" the secondary data on citrus area and production in Himachal Pradesh for the period of thirty three years i.e., 1980-81 to 2012-13 was utilized for model fitting and data for subsequent periods from 2014-15 to 2020-21 was used for validation. The analysis was carried out by using SPSS package and MS-Excel.

METHODOLOGY

I. Selection of trend and mathematical equations applied

After ensuring the presence of trend in the data for time series analysis following methods were employed:

Time series Models	*Mathematical Equation*
Straight line	$Y_t = a + b\,t$
Second degree parabola	$Y_t = a + b\,t + c\,t^2$
Exponential	$Y_t = a\,b^t$
Modified exponential	$Y_t = a + b\,c^t$
Gompertz	$Y_t = a\,b^{c^\wedge t}$
Logistic	$Y_t = k/1 + e^{(a+bt)}$

(i) Straight line: The straight line trend between the time series values (y_t) and time t be given by the equation:

$$Y_t = a + b\,t$$

where, Y_t = time series values at time t.
t = time period.

a and b are constants.

(ii) Second degree parabola: The second degree parabolic trend between the time series values (y_t) and time t be given by the equation:

$$Y_t = a + b\,t + c\,t^2$$

where, Y_t = time series values at time t.
t = time period.

a, b and c are constants.

(iii) Exponential: The exponential equation is:

$$Y = A\,B^t$$

Where, t is the time period and Y is the observation at time t. The exponential model used to compute the interest so that exponential model also known as compound interest model.

$$Y = A\,B^t$$
$$\log Y = \log A + t \log B$$
$$\text{i.e. } y = a + b\,t$$

where, $y = \log Y$
$a = \log A$ and
$b = \log B$

(iv) Modified exponential: The modified exponential equation is:

$$Y_t = a + b\,c^t$$
$$Y_{t+h} = a + b\,c^{(t+h)}$$
$$Y_{t-h} = a + b\,c^{(t-h)}$$
$$Y_{t+h} - Y_t = b\,c^t(c^h - 1)$$
$$Y_t - Y_{t-h} = b\,c^{t-h}(c^h - 1)$$
$$c^h = (Y_{t+h} - Y_t) / (Y_t - Y_{t-h}) = \text{Constant}$$

From the data three values that are equidistant in time and let these values be y_0, y_1 and y_2. The distance between x_0 and x_1 and between x_1 and x_2 are 'r' years, where r = (n-1) / 2, where n is the number of years i.e. period. The constant values are computed by the following formula.

$$c^r = (y_2 - y_1) / (y_1 - y_0) \text{ so that, } c = \sqrt[r]{C^r}$$
$$b = (y_1 - y_0) / (c^r - 1) \text{ so that, } a = y_0 - b$$

(v) Gompertz: The gompertz equation is:

$$Y_t = a\,b^{c^\wedge t}$$

Where Y_t is the time series value at time t and a, b and c are its parameters.

$$\log Y_t = \log a + c^t \log b$$
$$y_t = A + B\,c^t$$

where, $y_t = \log Y_t$
$A = \log a$ and
$B = \log b$

(vi) Logistic: The logistic equation is:

$$Y_t = k / 1 + (e^{(a+bt)})$$

Where Y_t is the value of time series at time t and a, b and k are constants.

$$1/Y_t = (1/k)\,(1 + e^{(a+bt)})$$
$$= (1/k) + (e^a \,.\, e^{bt}) / k$$
$$= A + bc^t$$

Where, $A = 1/k$ and $B = (1/k)\,e^a$

The logistic curve is fitted by method of three selected points. Three ordinates Y_1, Y_2 and Y_3 are now taken from trend line corresponding to selected equidistant points of time t_1, t_2 and t_3 such that, $t_2 - t_1 = t_3 - t_2$. Then constant values a and b obtained as:

$$k = \frac{y_2^2(Y_1+Y_3)-2\,Y_1\,Y_2\,Y_3}{y_2^2 - Y_1\,Y_3}$$

$$b = \frac{1}{(t_2-t_1)}\log\frac{(k-Y_2)\,Y_1}{(k-Y_1)\,Y_2}$$

$$a = \log\frac{(k-Y_1)}{Y_1} - bt_1$$

II. Estimation: The parameters are estimated by the least square technique appropriate to the time series data.

III. Diagnostic checking: For the adequacy of the model, the best fit model is obtained on the basis of maximum value of R^2 and $\bar{R}^2$ with minimum C.V (%).

(i) **R^2:** It is one of the indices to measure the goodness of fit of the time series models, it is also known as coefficient of determination.

$$R^2 = 1 - \frac{SSE}{SST}$$

Where, SSE = $\Sigma(y_i - \hat{y}_i)^2$ and

SST = $\Sigma(y_i - \bar{y})^2$

(ii) $\bar{R}^2$: This statistic used the R^2 statistic, $\bar{R}^2$ is based on the residual d.f and defined as the number of response values 'n' and number of fitted coefficients 'm' estimated from the response value.

Mathematically, V = n - m

V, indicates the number of independent piece of information involving the n data points that are required to calculate the SS. The $\bar{R}^2$ statistic is generally best indicator of the fit quality. When we compare the two models that are nested that is a series of model each of which add previous models.

$$\bar{R}^2 = 1 - \frac{SSE(n-1)}{SST(v)}$$

Adjusted R^2 statistic can take all values ≤ 1 closer to 1 indicating better fit, negative values can occur when the model contains terms that do not help to predict the response.

(iii) Coefficient of variation: It is the ratio of the S.D to the mean. Usually it is expressed in percentage. The formula for C.V is:

$$C.V = \frac{S.D}{\bar{Y}} \times 100$$

Where, S.D is the standard deviation and $\bar{Y}$ is the mean.

A series or a set of values having lesser coefficient of variation as compared to the other is more consistent.

RESULTS AND DISCUSSION

1. **Area of citrus:** Actual and estimated area of citrus crop computed by fitting of six equations namely: straight line, second degree parabola, exponential, modified exponential, gompertz and logistic.

Table 1
R^2, Adjusted R^2 and C.V (%)for the prediction models of citrus area.

Models	*R^2*	*$\bar{R}^2$*	*C.V (%)*
Straight line	0.001	0.001	29.97
Second degree parabola	0.773	0.757	14.53
Exponential	0.001	0.001	30.33
Modified exponential	0.060	0.040	25.85
Gompertz	0.001	0.0001	18.93
Logistic	0.005	0.001	30.63

Table 2
Coefficients of prediction models of citrus area.

Models	*a*	*b*	*c*
Straight line	29212.08	8.19 (185.48)*	-
Second degree parabola	10747.07	3470.78 (371.64)	-111.70
Exponential	27683.54	1.00 (0.007)	-
Modified exponential	31513.48	91.14 (56.38)	1.003
Gompertz	30199.52	1.0053 (0.07)	1.005
Logistic	0.0000078	0.989 (0.029)	-

* Values in parenthesis indicate the S.E of coefficient b.

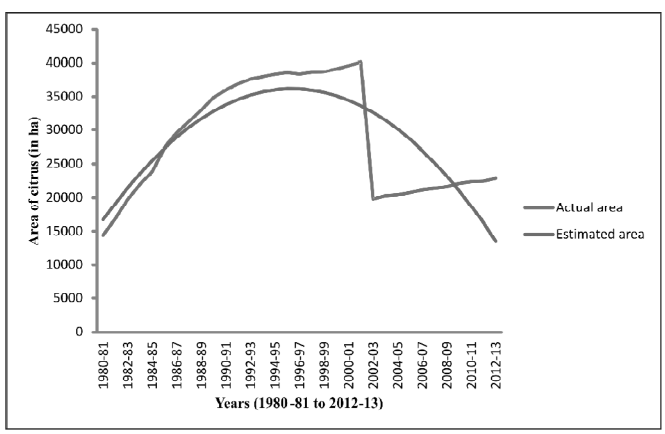

Figure 1: Actual and estimated area of citrus crop by second degree parabola

On the basis of statistic R^2, Adjusted R^2 and C.V values (Table 1) revealed that second degree parabola model fitted well. The present findings were in close agreement with Makinde *et al.* (2011), Sreekant (2008), Gaddour *et al.* (2008) and Lamien *et al.* (2007).

2. Production of citrus crop

Table 3
R^2, Adjusted R^2 C.V (%) for the prediction models of citrus production

Models	R^2	$\bar{R}^2$	C.V (%)
Straight line	0.566	0.551	34.19
Second degree parabola	0.681	0.657	29.88
Exponential	0.525	0.508	32.58
Modified exponential	0.362	0.349	39.16
Gompertz	0.268	0.214	41.88
Logistic	0.529	0.512	37.24

Table 4
Coefficients of prediction models of citrus production

Models	*a*	*b*	*c*
Straight line	3.03	0.67 (0.112)*	-
Second degree parabola	9.56	-0.54 (0.405)	0.03
Exponential	5.42	1.05 (0.009)	-
Modified exponential	8.73	0.004 (0.001)	1.34
Gompertz	8.01	1.00 (0.69)	1.25
Logistic	0.24	0.88 (0.019)	-

* Values in parenthesis indicate the S.E of coefficient b.

On the basis of statistic R^2, Adjusted R^2 and C.V values (Table 3) second degree parabola (0.681, 0.657 and 29.88) respectively were found to be best fitted.

It is observed forecasting of citrus crop obtained by fitting of different prediction models for the year 2020-21 area and production under the citrus crop will be 29141.73 (ha) and 35.96 (100 MT) respectively as per second degree parabola model which is best fitted to the area and production of citrus crop.

Table 5
Forecasting citrus area and production of Himachal Pradesh for the year 2020-21

Models	*Area (ha)*	*Production (100 MT)*
Straight line	29540.40	29.83
Second degree parabola	**29141.73**	**35.96**
Exponential	27683.54	38.15
Modified exponential	31616.22	494.30
Gompertz	30380.01	8.01
Logistic	31138.16	81.68

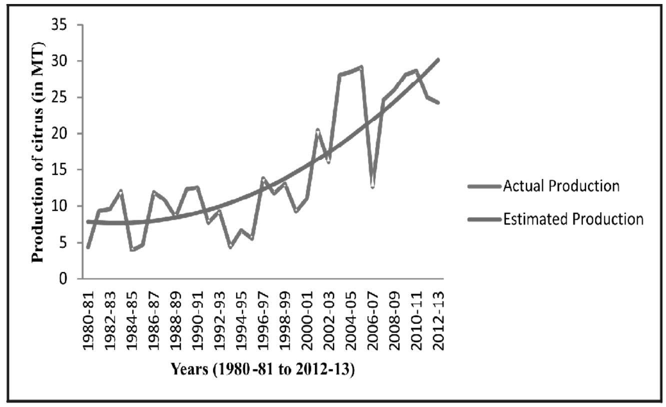

Figure 2: Actual and estimated production of citrus crop by second degree parabola

REFERENCES

Anonymous. (2014), Status of citrus area and production in India & HP. National Horticultural Board, Gurgaon.

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

Thomas, W. Macfarland. (1998), Regression, prediction and model building. *Journal of Agriculture Sciences*. **74**(5): 418-423.

Gaddour, A. and Najari, S. (2008), Adjustment of the kid's growth curve in pure goat breeds and crosses under southern Tunisian conditions. *Journal of Applied Animal Research*. **34**(2): 117-120.

Lamien, I. N., Tigabu, M., Guinko, S. and Oden, P. C. (2007), Variation in Dendromatric and fruiting characters of *Vitellari Paradoxa* population and multivariate models for estimation of fruit yield. *Agro Forestry Systems* **69**(1): 1-11.

Makinde, A. A., Afolayan, S. O. and Olanlyan, A. A. (2011), Effect of climate on citrus yield in rain forest savanna transitional zone of Nigeria. *Journal of Agriculture and Biological Sciences* **2**: 10-13.

Sreekant, R. D. (2008), Empirical prediction model for regional cashew production in India. *Journal of Plantation Crops*. **36**(3): 508-511.