AATCC Review

02 April 2025: Received
23 May 2025: Revised
30 May 2025: Accepted
01 July 2025: Available Online

https://aatcc.peerjournals.net/

Original Research Article

Open Access

Evaluation of Predictive models for tomato production and cultivation area in Himachal Pradesh: A Linear and Non- linear approach

Sukhdeep Kaur[1], Ashu Chandel*[1], RK Gupta[1], Geeta Verma[1], Pawan Kumar*[2] and Nandni Vashisht[3]

[1]*Department of Basic Sciences, Dr. Y. S. Parmar University of Horticulture and Forestry Nauni Solan, Himachal Pradesh – 173230, India.*
[2]*Department of Statistics and Data Science, Christ university central campus Bangalore, Karnataka– 560029, India*
[3]*Department of Social Science, College of Horticulture and Forestry, Neri, Hamirpur, Dr Yashwant Singh Parmar university of Horticulture and Forestry, Solan, Himachal, India*

ABSTRACT

This study analyzed the trend in area and production of tomatoes over a time period is important for understanding past behavior and for future planning. Tomato cultivation is highly sensitive to seasonal fluctuations and climatic factors. Therefore, to understand the prior and posterior patterns of tomato cultivation area and production, these statistical models were applied. The statistical study was carried out on different growth models viz. linear, quadratic, cubic, compound, and power for the area and production of tomatoes in Himachal Pradesh for the study period 1995 -2023. The study revealed the cubic and quadratic model was found to best fit the model for area and production, respectively. The highest value of CDVI for the area is 5.40 which indicates higher level of instability in which the variable is more erratic and has less area over time. The increasing annual growth rate for tomato areas is 4.60 percent and 5.90 percent with respect to production of tomatoes over the studied period of time using the compound model. The best-fit statistical models can be used to predict future values with greater accuracy.

Keywords: *Tomato, Model, Growth rate, Cuddy Della Vella Index, Production and Prediction*

Introduction

Tomatoes entered India through the Portuguese in the early 16th century. But they were introduced as a berry around areas of their dominion-modern-day Mumbai and Goa and never made it to the Plate. In India, where fresh fruits of tomatoes are in greater demand around the year, it is grown in an area of 8.49 lakh hectares with a production of 204.25 lakh tonnes and a productivity of 24.05 lakh tonnes per hectare [1]. Tomato is also an important off-season crop of Himachal Pradesh and is grown during summer and rainy seasons as the climatic conditions are congenial for optimum plant growth and yield. The state is the leading producer of fresh market tomatoes to the plains as high temperatures and continuous rains limit its production during the rainy season in such areas. In Himachal Pradesh, it occupies an area of 15,840 hectares with production of 5,32,340 metric tonnes and productivity of 33.60 metric tonnes per hectare [2]. Tomato is considered a 'Protective food or Poor man's orange' as it has some special nutritive value and antioxidant properties due to the presence of lycopene and flavonoids [3].

Methodology

In the present study, the time series data on the area and production of tomatoes was selected from the directorate of Agriculture, Shimla, Himachal Pradesh, for the period 1995 to 2023.

*Corresponding Author: ***Ashu Chandel and Pawan Kumar***

DOI: https://doi.org/10.21276/AATCCReview.2025.13.03.323

The data were analyzed by using R software for trend analysis.

Analytical Framework

Various statistical techniques were used to achieve the objective of the study.

Forecast through Regression models

When we consider time as an independent variable, Various linear and nonlinear regression models were used for trend analysis where the time is considered as an independent variable can be represented by the following equations:

i) Linear $Y_t = a + bt + e_t$
ii) Quadratic $Y_t = a + bt + ct^2 + e_t$
iii) Cubic $Y_t = a + bt + ct^2 + dt^3 + e_t$
iv) Power $Y_t = a \times t^b + e_t$
v) Compound $Y_t = a \times b^t + e_t$
vi) Exponential $Y_t = a \times e^{bt} + e_t$

where;

Y_t = time series values of dependent variable
T = independent variable, time element which takes the value 1, 2,…,n for various years.
A = intercept
b & c = regression coefficients
e_t = error term

Growth Rate

Linear growth rate: $L G R = \frac{b}{\bar{Y}t} \times 100$

The linear growth rate

Where,

b = regression coefficient of the linear model

$\overline{Yt}$ = mean of predicted value by a linear model

Compound growth rate:

The compound growth rates for different variables were computed by fitting the exponential function to the figure of production of tomatoes for the period of 1995 to 2023 in Himachal Pradesh. The ordinary least square method was used to fit the exponential function of the following form $Y = ae^{bt}$ It was converted into a log linear function with the help of logarithmic transformation as under:

$Ln\ Y = Ln\ a + b\ t.$

Where,

Ln = Natural logarithm

Y = Dependent variable (Production or area)

t = Independent variable (time in a year)

Cuddy Della- Valle Index

Variability can be calculated from the simple coefficient of variation. However, it is often overestimated the level of instability in the time series data due to the presence of long-term trends, so the Cuddy Della Vella index was constructed to correct the flaws of the present study.

$$CDVI = CV \times \sqrt{1 - R^2}$$

Where, CV = Coefficient of variation

R^2 Coefficient of determination

Model Adequacy

The choice of the model amongst the available alternatives is judged on the basis of, The goodness of fit is explained by R^2, Adj. R^2 , minimum value of RMSE (Root mean square error), CV (Coefficient of variation), and Theil's Inequality.

$$R^2 = 1 - \left[\frac{\sum(yi - \hat{y}i)^2}{\sum(yi - \bar{y})^2}\right]$$

$$\overline{R}^2 = 1 - \frac{\frac{ESS}{(n-k-1)}}{\frac{TSS}{(n-1)}}$$

$$RMSE = \sqrt{\frac{\sum(yi - \hat{y}i)^2}{n-2}}$$

$$\text{Theil's Inequality (U)} = \sqrt{\frac{\frac{\sum(Pi - Ai)^2}{n}}{\frac{\sum Ai^2}{n} + \frac{\sum pi^2}{n}}}$$

Where,

A_i = Actual change in the dependent variable.

n = Number of observations.

P_i = Predicted change in the dependent variable.

Y_i = Actual value

$\hat{y}$I= Predicted value

Result and Discussion

Tomato area over time

Table 1 shows that the prediction models namely: linear, quadratic, cubic, compound, power, and exponential had statistically significant regression coefficients. The standard error and t- statistic values for the regression coefficient in these models are also shown. Table 2 shows that all these six models were well fitted for the prediction of the area under tomato as R values range from (0.892 to 0.986) which was maximum in the case of the cubic model and the value of R^2 ranged between (0.795 to 0.972) which was found maximum in case of cubic model. The values R^2 ranged between (0.795 to 0.972) which was found maximum in case of cubic model. The values R^2 range from (0.788 to 0.969) and the highest value of R^2 (0.969) was found in the cubic model whereas the lowest value of 0.788 was found in the case of the compound and exponential model. The lowest RMSE value was found to be 550.67 for the cubic model and it was highest (1388.21) in case of the compound and exponential model. The lowest value of the coefficient of variation was found to be 32.31 for a cubic model which shows that the model is more consistent or has less variability. Low values of Theil's inequality coefficient were found and ranged between 0.037 to 0.094 which reflected good power of these models for forecasting the area of tomatoes in Himachal Pradesh.

Table 1: Coefficients, Standard error of coefficients, t- statistic of various models to Predict area under tomato in Himachal Pradesh

Statistical Models		Coefficient	Standard Error	t-Statistic
Linear	a	4083.640	393.741	10.371*
	b	359.829	22.924	15.696*
Quadratic	a	3239.062	591.384	5.477*
	b	523.295	90.866	5.759*
	c	-5.449	2.939	-1.854
Cubic	a	727.986	486.561	1.496
	b	1450.099	138.086	10.501*
	c	-81.389	10.600	-7.678*
	d	1.688	0.232	7.258*
Compound	a	4497.655	337.594	13.323*
	b	1.046	0.005	209.200*
Power	a	2626.761	144.161	18.221*
	b	0.492	0.021	23.264*
Exponential	a	4497.655	337.594	13.323*
	b	0.046	0.004	11.500*

Table 2: R, R^2, $\overline{R}^2$, RMSE, Coefficient of Variation (%) and Theil's Inequality (U) for the prediction models of area under tomato

Statistical Models	R	R^2	$\overline{R}^2$	RMSE	C.V	Theil's inequality (U)
Linear	0.949	0.901	0.898	1032.80	33.55	0.070
Quadratic	0.955	0.913	0.906	970.70	32.52	0.066
Cubic	**0.986**	**0.972**	**0.969**	**550.67**	**32.31**	**0.037**
Compound	0.892	0.795	0.788	1388.21	37.59	0.094
Power	0.976	0.952	0.951	874.63	33.32	0.059
Exponential	0.892	0.795	0.788	1388.21	37.59	0.094

On the basis of R, R^2, Adj. R^2 root mean square error (RMSE), coefficient of variation (C.V.), and Theil's inequality coefficient (U), The cubic model was found to be the best prediction model to forecast the area of tomato among all the six prediction models because the cubic model has highest value of R (0.986), $\overline{R}^2$ (0.919) and R^2 (0.907) and has the lowest value of RMSE (550.6732), Theil's inequality coefficient (U) (0.037) and low coefficient of variation (32.31%) which was in close affinity with the finding of area, production and productivity of wheat in Gujrat state [4] and trends of area, production and productivity of onion in Madhya Pradesh [5].
It can be clearly observed from the figure the observed and predicted area of tomato is close in the case of the cubic model as compared to the other five models.

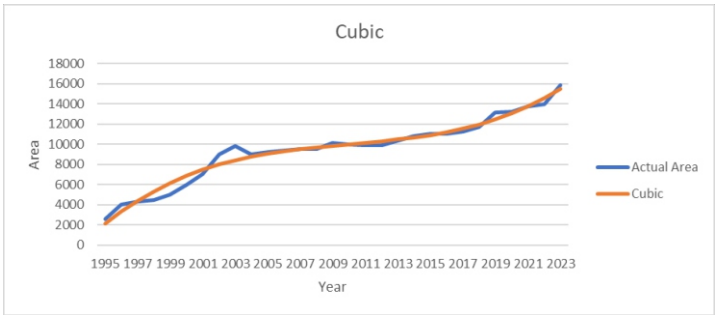

Fig 1: Trend values of actual and predicted values of cubic model for tomato area in Himachal Pradesh

Tomato production over time

Table 3: Coefficients, Standard error of coefficients, t- statistic of various models to Predict tomato production in Himachal Pradesh

Statistical Models		Coefficient	Standard Error	t-Statistic
Linear	a	100667.820	9292.550	10.833*
	b	16852.897	541.033	31.149*
Quadratic	a	58628.506	10295.876	5.694*
	b	24989.539	1581.956	15.797*
	c	-271.221	51.169	-5.300*
Cubic	a	67164.226	14737.926	4.557*
	b	21839.124	4182.631	5.221*
	c	-13.085	321.066	-0.041
	d	-5.736	7.042	-0.292
Compound	a	133749.597	9529.490	14.035*
	b	1.059	0.004	241.065*
Power	a	70462.637	2712.885	25.973*
	b	0.611	0.015	41.200*
Exponential	a	133749.597	9529.490	14.035*
	b	0.059	0.004	13.844*

**indicates significance at 5% level of significance*

Table 4: R, R^2, $\overline{R}^2$, RMSE, Coefficient of Variation (%) and Theil's Inequality (U) for the prediction models of tomato production

Statistical Models	R	R^2	$\overline{R}^2$	RMSE	C.V	Theil's inequality (U)
Linear	0.986	0.973	0.972	24376.53	40.59	0.043
Quadratic	**0.994**	**0.988**	**0.987**	**16679.88**	**39.54**	**0.029**
Cubic	0.993	0.986	0.985	16899.77	40.89	0.030
Compound	0.936	0.877	0.872	60581.91	47.81	0.106
Power	0.992	0.984	0.984	19385.10	40.91	0.034
Exponential	0.936	0.877	0.872	60581.91	47.81	0.106

Table 4 shows that all these six models were fitted for the prediction of tomato production as R values range from (0.936 to 0.994) which was the maximum in the case of quadratic model and the value of R^2 ranged between (0.877 to 0.988) which was found highest in case of the quadratic model whereas it was lowest in case of the compound and exponential model. $\overline{R}^2$ Values range from (0.872 to 0.987) in the highest value was found in the quadratic model whereas the lowest value was found in case of the compound and exponential model. The lowest RMSE value was found to be 16679.88 for the quadratic model and it was highest in the case of the compound model. The lowest value of the coefficient of variation was found to be 39.54 for a quadratic model which shows that the model is more consistency or less variability. Low value of Theil's inequality coefficient was found and ranged between (0.029 to 0.106) which reflected good power of these models for the prediction of tomato production.

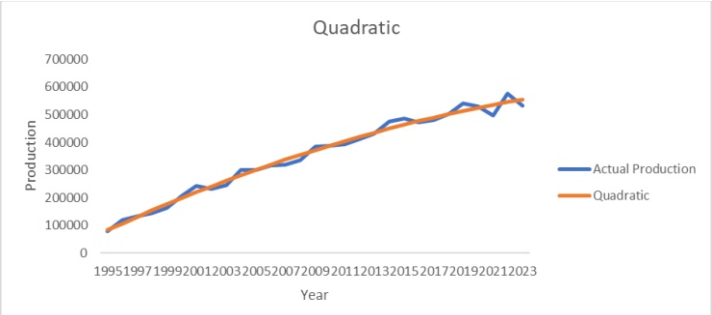

Fig 2: Trend values of actual and predicted values of quadratic model for tomato production in Himachal Pradesh

On the basis of R, R^2, $\overline{R}^2$, RMSE, Coefficient of variation, and Theil's inequality (U) quadratic model was found to be the best prediction model to forecast the production of tomatoes among all six prediction models because the quadratic model has the highest values of R (0.994), R^2 (0.988) and $\overline{R}^2$ (0.987) and has the lowest value for RMSE (16679.88), Thiel's inequality coefficient (U) (0.029) and coefficient of variation (39.54%) which has in close affinity with the finding of area, production, and productivity of fruit crops in Nagaland state [6] and trends of area, production, and productivity of turmeric in Assam [7].
It can be clearly observed from figure 2 that the observed and predicted value of tomato production is closed in case of the quadratic model as compared to other five models.

Table 5: Trends in area and Production of tomato during 1995 – 2023

	Average	CDVI	LGR	CGR
Area (ha)	9481.06	5.40	3.79	4.76
Production (MT)	353461.30	4.33	4.60	5.90

Table 5 represents the average tomato area and production for the study period is 9481.06 ha and 353461.30 MT, respectively. The highest value of CDVI for the area is 5.40 which indicates a higher level of instability in which the variable is more erratic. The increasing annual growth rate for tomato area is 3.79 and 4.60 percent and 4.76 and 5.90 percent with respect to the production of tomatoes over the studied period of time using linear and compound models.

Conclusion

The present study regression parameter for estimating of cultivated tomato crop area and production in Himachal Pradesh. The appropriateness of the model was judged by the magnitude and sign of the parameter estimates, goodness of fit (Goodness of fit is explained by R^2, Adj. R^2, the minimum value of RMSE (Root mean square error), CV (Coefficient of variation), and Theil's Inequality. The developed of regression model for the tomato area was cubic and production was quadratic regression models was found to best fit Himachal Pradesh for the study period with the increased trend. These regression models were used for the prediction of the area and production of tomatoes in Himachal Pradesh. The average tomato area and production for the study period is 9481.06 ha and 353461.30 MT, respectively. The highest value of CDVI for area is 5.40 which indicates a higher level of instability in which the variable is more erratic and less area over time. The increasing annual growth rate for tomato area is 4.60 percent and 5.90 percent with respect to production of tomatoes over the studied period of time using the compound model.

References

1. Anonymous. 2023. Area and Production of Horticulture crops 2023. Ministry of Agriculture & Farmer welfare. http://pib.gov.in

2. Anonymous. 2023. Area, Production and Productivity of crops 2023. Department of Agriculture Himachal Pradesh. http://agriculture.hp.gov.in

3. Sepat N K, Sepat S R, Sepat S and Kumar A. 2013. Energy use efficiency and cost analysis of tomato under greenhouse and open field production system at Nubra valley of Jammu and Kashmir. *International journal of Environmental sciences.* 3: 1233- 1241

4. Singh L N, Draji V B and Singh Y S. 2015. Area, Production and Productivity of Wheat (*Triticum aestivum*) in Gujarat state: forecasting by using ARIMA models. *International Journal of Bio- resource and stress Management*. 7: 1093-1098

5. Niranjan H K and Chouhan R S. 2016. Past and future trends using ARIMA model in area, production and productivity of onion in Madhya Pradesh. *International Journal of Agricultural Sciences and Research.* 6: 189-94

6. Sharma A. 2012. Trends of Area, Production and Productivity of Fruit crops in Nagaland state of North Eastern Hill Region of India. *Economic Affairs.* 57: 216 – 78

7. Choudhury K and Kalita D C. 2018. Trends of area, production and productivity of turmeric in Assam. *Journal of Hill Agriculture.* 9: 322-24