**Statistical Modeling and Trend Analysis of Soybean Production in Selected States of India**

**ABSTRACT**

In the present paper, the time series analysis of Soybean production in some selected states of India has been carried out by fitting statistical models, viz. linear, exponential, Quadratic and cubic models. The secondary time series data on the production of soybean have been utilized for the analysis. The trend values have been evaluated by fitting the concerned models, and the validity of the models has been tested by using the Chi-square test statistic. Moreover, the coefficient of determination ($R^{2}$), root mean square error (RMSE), and relative mean absolute percentage error (RMAPE) have been computed to reveal the suitability of the concerned models for exploring the trend patterns of soybean production in the concerned states of India.

***Keywords:*** *Time series, Linear model, Exponential model, Quadratic model, Cubic model, Chi-square test, Coefficient of determination.*

1. **Introduction**

Soybean (Glycine max L.) is an annual legume plant of the Fabaceae family. It is economically the most important bean in the world, providing vegetable protein for humans. It improves soil fertility by atmospheric immobilization to the extent of 50-300 kg/ha depending upon the agro-climatic conditions, variety, strains, etc. In India, the soybean is mainly a kharif season crop. The most suitable soil for a soybean crop is loam, pH required from 6.0 to 7.5 and moderate water holding capacity. Soybean is mainly consumed in processed form (soy milk, a whitish liquid suspension, tofu, a curd somewhat resembling cottage cheese, soy sauce, a salty brown liquid, etc.). Some other fermented soybean foods are tempeh, miso, and fermented bean paste. Cultivation of yellow soybean was introduced in India in the early 1970s and since then its cultivation has grown dramatically, placing it at the number one position among the oilseed crops of India. Prashnani M. *et al.* (2024) Globally, soybean is considered as a major component of the human diet, as it is providing vegetable protein for humans. Soybean is popular for its nutritive qualities and health benefits. The soybean is one of the richest sources of protein. It is a staple in the diets of people and animals in numerous parts of the world. It contains about 40 % to 42 % good quality protein.

In India, the leading state in the production of soybean was Maharashtra (6.62 million tons) during the year 2022-23, followed by Madhya Pradesh (5.79 million tons), Rajasthan (1.21 million tons), Karnataka (0.54 million tons), Gujrat (0.37 million tons), and Telangana (0.33 million tons). In India, the overall production of soybean was 14.98 million tons during 2022-23 [Source: Directorate of Economics & Statistics, DAC&FW, Govt. of India].

Many researchers and scientists have performed soybean crop time series analysis and forecasting for various geographic regions around the world. Masuda and Goldsmith (2009) forecasted world soybean production using an exponential smoothing model with a slowing trend. He also presented three scenarios and their implications for increasing supply with decreasing land availability. These scenarios highlight the urgent need for significant investment in yield improvement research for agribusiness policy makers and managers. Ramteke *et al.* (2015) studied the changes in soybean scenario over a period from 2001 to 2013 of Indore district and Madhya Pradesh since 1980 with respect to area, production, and productivity, taking in consideration the changes in climatic variables like rainfall and temperature during the corresponding period. Zeng *et al*. (2016) applied a hybrid phenology detection method that incorporates the “shape-model fitting” concept of the two-step filtering method and a simulation concept of the crop models to detect the critical vegetative stages and reproductive stages of corn (Zea mays L.) and soybeans (Glycine max L.) from MODIS 250-m Wide Dynamic Range Vegetation Index (WDRVI) time-series data and 1000-m Land Surface Temperature (LST) data. Gusso *et al.* (2017) evaluated the reliability of the physiological meaning of the enhanced vegetation index (EVI) data for the development of a remote sensing-based procedure to estimate soybean production prior to crop harvest by using coupled model (CM). Larbi and Green (2018) used to developed models that can aid precision agriculture applications by using resulting data. The former showed a rise and fall trend with daily peaks around 13:00, while the latter showed a decreasing order of correlation with weather variables. Abraham *et al.* (2020) used Artificial Neural Networks (ANN) to predict soybean harvest area, yield, and production in Brazil and compared with classical methods of time series analysis during 1961-2016. Filho *et al.* (2020) applied pre-processing methods in time-series and uses representations combined with textual data to predict the future price of corn and soybeans and indicated that the methods used can be an alternative to improve forecasting performance in regression tasks. Parmar and Devi (2021) explored the trend in area, production and productivity of soybean in Gujarat with the help of time series data pertaining to the period 2010-2019. The results showed that the CGRs of area, production and productivity of soybean were increased significantly over the period of last ten years in Gujarat. Stepanov *et al.* (2022) studied the seasonal time series of the normalized difference vegetation index (NDVT) to obtain an early forecast of soybean yield. This research used the Moderate Resolution Image Spectroradiometer (MODIS), an arable land mask obtained from the VEGA-Science web service, and soybean yield data. Four estimating functions were used to model the NDVI time series. Gaussian, double logistic (DL), and quadratic and cubic polynomials. Han and Ng’ombe (2023) used a Bayesian time series model and data from Statistics Canada and the Alberta Department of Agriculture and Rural Development to examine the time series relationship between hemp, wheat, and soybean acreage. It was also suggested that hemp has the potential to replace soybean in a soybean-wheat dual crop as hemp has key attributes of soybean as a rotation crop (profitability, potential as an energy comp and maintenance of soil fertility). Thomasz *et al.* (2023) founded that in the major number of cases soil water content explains at least 50% percent of the variability in soybean yields, with a maximum of 70% explanatory power in one county, by means of a correlation and regression analysis. Kumar *et al.* (2024) analyzed the growth and trend pattern of wheat production in selected states of India during the period 2011-2020. Thimmegowda *et al.* (2025) compared statistical and machine learning models to assess their ability to predict cotton yield across major producing districts of Karnataka, India, utilizing a long-term dataset spanning from 1990 to 2023 that includes yield and weather factors. Kumar *et al.* (2025) analyzed the trends, pattern of rice production during the period 2011-2020, and fitting some well-known models viz., linear model, exponential model and cubic model in some rice growing states of India. Gowda *et al.* (2025) examined the growth pattern and instability in area, production and yield of rice in some selected states of India during the period 2011-2022.

**2. DATA AND METHODOLOGY**

**2.1 Source of Data**

The present paper deals with the analysis of secondary time series data on soybean production pertaining to the period (2005-2022) in some selected states of India. The time series data is obtained through the records of Directorate of Economics & Statistics, DAC&FW, Govt. of India.

**2.2 Terminologies and Notations**

In the present analysis, we have considered three soybean growing states of India, viz. Karnataka (S1), Maharashtra (S2), and Madhya Pradesh (S3). In these states, we observe various trends of soybean production during the concerned period of study.

**2.3 Fitting of Statistical Models to the Data**

In order to analyze the growth and trend patterns of soybean production in the concerned states S1, S2 and S3, we compute the trend values by fitting linear, exponential, Quadratic and cubic models to the time series data on soybean production as follows:

**(a) Linear Model:**

$y\_{t}=a+bt$ … (1)

where $y\_{t}$ denotes the time series value at time $t$. The values of constants ‘$a$’ and ‘$b$’ are obtained on using the principle of least squares by solving the following normal equations:

$\sum\_{}^{}y\_{t}=na+b\sum\_{}^{}t $… (2)

$\sum\_{}^{}ty\_{t}=a\sum\_{}^{}t+b\sum\_{}^{}t^{2}$… (3)

where ‘$n$’ represents the number of observed values.

**(b) Exponential Model:**

$y\_{t}=ae^{bt}$… (4)

Taking natural log on both sides of above equation, we have

$log\_{e} y\_{t}$*=* $log\_{e}$ *a + bt* $log\_{e}$ *e*

i.e., $Y\_{t}=A+bt$ … (5)

where$ Y\_{t}= log\_{e}y\_{t}$ , $ A = log\_{e} a$ , and $log\_{e} e=1$

The normal equations for estimating the values of ‘$A$*’* and ‘$b$*’* are as follows:

$\sum\_{}^{}Y\_{t}=nA+b\sum\_{}^{}t$… (6)

$\sum\_{}^{}tY\_{t}=A\sum\_{}^{}t+b\sum\_{}^{}t^{2}$… (7)

Finally, the value of ‘$a$*’* is obtained on using

$$a=antilog \left(A\right)$$

**(c) Quadratic Model:**

$y\_{t}= a +bt + ct^{2}$… (8)

The values of constants ‘$a$’, ‘$b$’, and ‘$c$’ are obtained on solving the following normal equations**.**

$\sum\_{}^{}y\_{t}=na+b \sum\_{}^{}t+c \sum\_{}^{}t^{2}$… (9)

$\sum\_{}^{}ty\_{t}=a\sum\_{}^{}t+b\sum\_{}^{}t^{2}+c\sum\_{}^{}t^{3}$… (10)

$\sum\_{}^{}t^{2}y\_{t}=a\sum\_{}^{}t^{2}+b\sum\_{}^{}t^{3}+c\sum\_{}^{}t^{4}$… (11)

**(d) Cubic Model:**

$y\_{t}= a +bt + ct^{2}+ dt$3… (12)

The values of constants ‘$a$’, ‘$b$’, ‘$c$’ and ‘$d$’ are obtained on solving the following normal equations**.**

$\sum\_{}^{}y\_{t}=na+b \sum\_{}^{}t+c \sum\_{}^{}t^{2}+d \sum\_{}^{}t^{3}$… (13)

$\sum\_{}^{}ty\_{t}=a\sum\_{}^{}t+b\sum\_{}^{}t^{2}+c\sum\_{}^{}t^{3}+d\sum\_{}^{}t^{4}$… (14)

$\sum\_{}^{}t^{2}y\_{t}=a\sum\_{}^{}t^{2}+b\sum\_{}^{}t^{3}+c\sum\_{}^{}t^{4}+d\sum\_{}^{}t^{5}$… (15)

$\sum\_{}^{}t^{3}y\_{t}=a\sum\_{}^{}t^{3}+b\sum\_{}^{}t^{4}+c\sum\_{}^{}t^{5}+d\sum\_{}^{}t^{6}$… (16)

**3. DATA ANALYSIS AND RESULTS**

The secondary time series data on soybean production in states S1, S2 and S3 of India is presented in Table 1. The trend values are obtained on fitting linear, exponential, Quadratic and cubic models to the data in the concerned states, and are depicted in Tables 2, 3 and 4, respectively. Moreover, the model equations for linear, exponential, Quadratic and cubic trends in the respective states are elaborated in Table 5.

**Table 1:** Time series data on soybean production in selected states of India

|  |  |
| --- | --- |
| **Year** | **\*Production (in million tons) for the states** |
| **S1** | **S2** | **S3** |
| 2005 | 0.07 | 2.53 | 4.50 |
| 2006 | 0.09 | 2.89 | 4.78 |
| 2007 | 0.10 | 3.98 | 5.48 |
| 2008 | 0.09 | 2.76 | 5.85 |
| 2009 | 0.08 | 2.20 | 6.41 |
| 2010 | 0.15 | 4.32 | 6.67 |
| 2011 | 0.17 | 3.97 | 6.28 |
| 2012 | 0.18 | 4.67 | 7.80 |
| 2013 | 0.27 | 4.75 | 5.24 |
| 2014 | 0.19 | 2.38 | 6.35 |
| 2015 | 0.14 | 2.06 | 4.91 |
| 2016 | 0.32 | 4.59 | 6.65 |
| 2017 | 0.25 | 3.80 | 5.32 |
| 2018 | 0.26 | 4.61 | 6.67 |
| 2019 | 0.38 | 4.83 | 4.89 |
| 2020 | 0.38 | 6.26 | 4.26 |
| 2021 | 0.44 | 5.50 | 5.39 |
| 2022 | 0.54 | 6.62 | 5.79 |

**(\*Source:** Directorate of Economics & Statistics, DAC&FW, Govt. of India)

**Table 2:** Trend values for linear, exponential, Quadratic and cubic models in state S1

|  |  |  |
| --- | --- | --- |
| **Year** **(t)** | **Production****(**$y\_{t}$**)** | **Trend Values** |
| **Linear Model (**$L\_{t}$**)** | **Exponential Model (**$E\_{t}$**)** | **Quadratic Model (**$C\_{t}$**)** | **Cubic Model (**$C\_{t}$**)** |
| 2005 | 0.07 | 0.03 | 0.07 | 0.08 | 0.08 |
| 2006 | 0.09 | 0.05 | 0.08 | 0.09 | 0.10 |
| 2007 | 0.10 | 0.07 | 0.09 | 0.09 | 0.11 |
| 2008 | 0.09 | 0.10 | 0.10 | 0.10 | 0.12 |
| 2009 | 0.08 | 0.12 | 0.12 | 0.11 | 0.13 |
| 2010 | 0.15 | 0.14 | 0.13 | 0.13 | 0.15 |
| 2011 | 0.17 | 0.17 | 0.14 | 0.14 | 0.16 |
| 2012 | 0.18 | 0.19 | 0.16 | 0.16 | 0.17 |
| 2013 | 0.27 | 0.22 | 0.18 | 0.18 | 0.18 |
| 2014 | 0.19 | 0.24 | 0.20 | 0.20 | 0.20 |
| 2015 | 0.14 | 0.26 | 0.22 | 0.23 | 0.22 |
| 2016 | 0.32 | 0.29 | 0.25 | 0.26 | 0.24 |
| 2017 | 0.25 | 0.31 | 0.28 | 0.29 | 0.27 |
| 2018 | 0.26 | 0.33 | 0.31 | 0.33 | 0.30 |
| 2019 | 0.38 | 0.36 | 0.35 | 0.36 | 0.34 |
| 2020 | 0.38 | 0.38 | 0.39 | 0.40 | 0.38 |
| 2021 | 0.44 | 0.41 | 0.44 | 0.44 | 0.43 |
| 2022 | 0.54 | 0.43 | 0.49 | 0.49 | 0.49 |

**Table 3:** Trend values for linear, exponential, Quadratic and cubic models in state S2

|  |  |  |
| --- | --- | --- |
| **Year****(t)** | **Production****(**$y\_{t}$**)** | **Trend Values** |
| **Linear Model (**$L\_{t}$**)** | **Exponential Model (**$E\_{t}$**)** | **Quadratic Model (**$C\_{t}$**)** | **Cubic Model (**$C\_{t}$**)** |
| 2005 | 2.53 | 2.53 | 2.65 | 3.17 | 2.56 |
| 2006 | 2.89 | 2.71 | 2.77 | 3.12 | 2.95 |
| 2007 | 3.98 | 2.89 | 2.89 | 3.10 | 3.22 |
| 2008 | 2.76 | 3.06 | 3.02 | 3.11 | 3.41 |
| 2009 | 2.20 | 3.24 | 3.15 | 3.15 | 3.53 |
| 2010 | 4.32 | 3.42 | 3.28 | 3.21 | 3.59 |
| 2011 | 3.97 | 3.60 | 3.43 | 3.30 | 3.62 |
| 2012 | 4.67 | 3.77 | 3.58 | 3.43 | 3.63 |
| 2013 | 4.75 | 3.95 | 3.73 | 3.58 | 3.65 |
| 2014 | 2.38 | 4.13 | 3.90 | 3.75 | 3.68 |
| 2015 | 2.06 | 4.31 | 4.07 | 3.96 | 3.75 |
| 2016 | 4.59 | 4.48 | 4.24 | 4.19 | 3.88 |
| 2017 | 3.80 | 4.66 | 4.43 | 4.45 | 4.07 |
| 2018 | 4.61 | 4.84 | 4.62 | 4.74 | 4.36 |
| 2019 | 4.83 | 5.02 | 4.82 | 5.06 | 4.76 |
| 2020 | 6.26 | 5.19 | 5.03 | 5.41 | 5.29 |
| 2021 | 5.50 | 5.37 | 5.25 | 5.78 | 5.96 |
| 2022 | 6.62 | 5.55 | 5.48 | 6.19 | 6.79 |

**Table 4:** Trend values for linear, exponential, Quadratic and cubic models in state S3

|  |  |  |
| --- | --- | --- |
| **Year** **(t)** | **Production****(**$y\_{t}$**)** | **Trend Values** |
| **Linear Model (**$L\_{t}$**)** | **Exponential Model (**$E\_{t}$**)** | **Quadratic Model (**$C\_{t}$**)** | **Cubic Model (**$C\_{t}$**)** |
| 2005 | 4.50 | 5.81 | 5.71 | 4.92 | 4.23 |
| 2006 | 4.78 | 5.80 | 5.71 | 5.22 | 5.02 |
| 2007 | 5.48 | 5.79 | 5.70 | 5.49 | 5.63 |
| 2008 | 5.85 | 5.78 | 5.70 | 5.72 | 6.06 |
| 2009 | 6.41 | 5.77 | 5.69 | 5.90 | 6.34 |
| 2010 | 6.67 | 5.77 | 5.69 | 6.05 | 6.48 |
| 2011 | 6.28 | 5.76 | 5.68 | 6.16 | 6.52 |
| 2012 | 7.80 | 5.75 | 5.67 | 6.23 | 6.47 |
| 2013 | 5.24 | 5.74 | 5.67 | 6.26 | 6.35 |
| 2014 | 6.35 | 5.73 | 5.66 | 6.25 | 6.17 |
| 2015 | 4.91 | 5.72 | 5.66 | 6.21 | 5.97 |
| 2016 | 6.65 | 5.71 | 5.65 | 6.12 | 5.76 |
| 2017 | 5.32 | 5.71 | 5.65 | 5.99 | 5.56 |
| 2018 | 6.67 | 5.70 | 5.64 | 5.83 | 5.40 |
| 2019 | 4.89 | 5.69 | 5.63 | 5.62 | 5.29 |
| 2020 | 4.26 | 5.68 | 5.63 | 5.38 | 5.25 |
| 2021 | 5.39 | 5.67 | 5.62 | 5.10 | 5.30 |
| 2022 | 5.79 | 5.66 | 5.62 | 4.78 | 5.47 |

In Tables 2, 3 and 4, the term ‘$y\_{t}$’ denotes the observed value of soybean production (in million tons) for the year ‘$t$’ $(t=2005, 2012, … , 2022)$. Moreover, ‘$L\_{t}$’ denotes the linear trend value of soybean production for the year ‘$t$’. In a similar manner, ‘$E\_{t}$’ denotes the exponential trend value of soybean production, ‘$Q\_{t}$’ denotes the quadratic trend value of soybean production and ‘$C\_{t}$’ denotes the cubic trend value of soybean production.

**Table 5:** Model equations for linear, exponential, Quadratic and cubic trends in selected states of India

|  |  |  |  |
| --- | --- | --- | --- |
| **States** | **S1** | **S2** | **S3** |
| Linear Model | $$y\_{t}=-47.504 + 0.023t$$ | $$y\_{t}= -353.45+ 0.18t$$ | $$y\_{t}= 22.653 - 0.0084t$$ |
| Exponential Model | $$y\_{t}=7E-99e^{0.1114t}$$ | $$y\_{t}=1.7E-37e^{0.0427t}$$ | $$y\_{t}=41.712e^{-0.001t}$$ |
| Quadratic Model | $$y\_{t}= -5320.6+$$$$0.024t+0.00132t^{2}$$ | $$y\_{t}= -57546.2+$$$$0.12t+0.014t^{2}$$ | $$y\_{t}=79524.6-$$$$0.0084t-0.0196t^{2}$$ |
| Cubic Model | $$y\_{t}= -822224.3+$$$$0.08t+0.0013t^{2}+0.0001t^{3}$$ | $$y\_{t}=-2456479+$$$$0.0321t+0.0141t^{2}+0.003t^{3}$$ | $y\_{t}= -27855051-0.1729t-0.0196t^{2}$ +0.0034$t^{3}$ |

In order to illustrate the relative influence of linear, exponential, Quadratic and cubic trend values on the observed values of soybean production for the states S1, S2 and S3, the graphical plots are obtained and demonstrated in Figs. 1 to 12.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Fig. 1:** Trend values for Linear Model in state S1 | **Fig. 2:** Trend values for Linear Model in state S2 | **Fig. 3:** Trend values for Linear Model in state S3 |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Fig. 4:** Trend values for Exponential Model in state S1 | **Fig. 5:** Trend values for Exponential Model in state S2 | **Fig. 6:** Trend values for Exponential Model in state S3 |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Fig. 7:** Trend values for Quadratic Model in state S1 | **Fig. 8:** Trend values for Quadratic Model in state S2 | **Fig. 9:** Trend values for Quadratic Model in state S3 |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Fig. 10:** Trend values for Cubic Model in state S1 | **Fig. 11:** Trend values for Cubic Model in state S2 | **Fig. 12:** Trend values for Cubic Model in state S3 |

In order to test the suitability of various fitted models, we have computed the coefficient of determination ($R^{2})$, Root Mean Square Error (RMSE) and Relative Mean Absolute Percentage Error (RMAPE) for the selected states, by using the following formulae:

$$R^{2}=1-\frac{\sum\_{t=1}^{n}\left(y\_{t}-ŷ\_{t}\right)^{2}}{\sum\_{t=1}^{n}\left(y\_{t}-ӯ \right)^{2}}$$

$$RMSE=\sqrt{\frac{1}{n}\sum\_{t=1}^{n}\left(y\_{t}-\hat{y}\_{t}\right)^{2}}$$

and

$$RMAPE=\frac{1}{n}\sum\_{t=1}^{n}\left|\frac{y\_{t}-\hat{y}\_{t}}{y\_{t}}\right|×100$$

where $y\_{t}$ denotes the observed value of soybean production ($Y$), and $ӯ $is the mean value of the variable $Y$. Also, $ŷ\_{t}$ is the trend value of the variable $Y$, which is obtained on fitting the respective statistical model (such as linear or exponential or quadratic or cubic model, as the case may be) to the variable $Y$.

The values of $R^{2}$, RMSE and RMAPE for the concerned states are obtained on fitting linear, exponential, quadratic and cubic models, and presented in Table 6.

**Table 6:** Model evaluation for soybean production in selected states of India

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| States | Models | $$R^{2}$$ | RMSE | RMAPE |
| S1 | Linear | 0.84 | 0.053 | 24.21 |
| Exponential | 0.90 | 0.041 | 16.39 |
| Quadratic | 0.90 | 0.043 | 17.45 |
| Cubic | 0.91 | 0.043 | 17.45 |
| S2 | Linear | 0.49 | 0.944 | 22.79 |
| Exponential | 0.52 | 0.927 | 22.11 |
| Quadratic | 0.55 | 0.882 | 22.98 |
| Cubic | 0.62 | 0.819 | 20.55 |
| S3 | Linear | 0.0024 | 0.899 | 13.56 |
| Exponential | 0.0022 | 0.903 | 13.40 |
| Quadratic | 0.28 | 0.766 | 11.39 |
| Cubic | 0.45 | 0.673 | 9.06 |

From Table 6, the following results are obtained:

1. In each of the three states S1, S2 and S3, the values of $R^{2}$ are more for the cubic model as compared to the linear, quadratic and exponential models. Moreover, the values of $R^{2}$ are nearly the same for both linear and exponential models in each state.
2. In each state S1, S2 and S3, we observe that $R^{2}>0.5$ for the cubic model, whereas $R^{2} value$ for the linear, exponential and quadratic model is less then cubic model. Hence, among all the four models, cubic model is the best fitted model.
3. In each state, the values of RMSE are least for cubic model as compared to the linear and exponential models. Furthermore, the values of RMSE are nearly the same for both linear and exponential models.
4. In each state, the values of RMAPE are least for cubic model as compared to the linear, exponential and quadratic models. Also, the values of RMAPE are least for cubic model as compared to the linear, exponential and quadratic models except S1.

**3.1 Formulation of Hypotheses**

We test the following null hypotheses:

$H\_{0L}: $ Linear model fits the given data on soybean production.

$H\_{0E}: $ Exponential model fits the given data on soybean production.

$H\_{0Q}: $ Quadratic model fits the given data on soybean production.

$H\_{0C}: $ Cubic model fits the given data on soybean production.

against the following respective alternative hypotheses:

$H\_{1L}: $ Linear model does not fit the given data on soybean production.

$H\_{1E}: $ Exponential model does not fit the given data on soybean production.

$H\_{1Q}: $ Quadratic model does not fit the given data on soybean production.

$H\_{1C}: $ Cubic model does not fit the given data on soybean production.

The above-mentioned hypotheses for model fitting on soybean production are tested using chi-square test statistic, in the concerned states S1, S2 and S3 of India.

**3.2 Hypotheses Testing and Validation**

The chi-square values have been computed for the linear, exponential, quadratic and cubic models (i.e., $χ\_{L}^{2}$, $χ\_{E}^{2}$, $χ\_{Q}^{2}$ and $χ\_{C}^{2}$) in the concerned states of India, and the findings are depicted in Table 7. The chi-square values, on fitting the concerned models, have been obtained using the following formulae:

$$χ\_{L}^{2}=\sum\_{t=1}^{n}\frac{(y\_{t}-L\_{t})^{2}}{L\_{t}}=\sum\_{t=1}^{18}\frac{(y\_{t}-L\_{t})^{2}}{L\_{t}} ,$$

$$χ\_{E}^{2}=\sum\_{t=1}^{n}\frac{(y\_{t}-E\_{t})^{2}}{E\_{t}}=\sum\_{t=1}^{18}\frac{(y\_{t}-E\_{t})^{2}}{E\_{t}} ,$$

$$χ\_{Q}^{2}=\sum\_{t=1}^{n}\frac{(y\_{t}-Q\_{t})^{2}}{Q\_{t}}=\sum\_{t=1}^{18}\frac{(y\_{t}-Q\_{t})^{2}}{Q\_{t}} ,$$

$$χ\_{C}^{2}=\sum\_{t=1}^{n}\frac{(y\_{t}-C\_{t})^{2}}{C\_{t}}=\sum\_{t=1}^{18}\frac{(y\_{t}-C\_{t})^{2}}{C\_{t}} ,$$

where the terms ‘$y\_{t}$’, ‘$L\_{t}$’, ‘$E\_{t}$’, ‘$Q\_{t}$’ and ‘$C\_{t}$’ have been utilized from the Tables 2, 3 and 4, for the concerned states S1, S2 and S3 of India.

**Table 7:** Values of chi-square statistic on fitting linear, exponential, quadratic and cubic models

|  |  |
| --- | --- |
| **Chi-square values** | **States** |
| **S1** | **S2** | **S3** |
| Linear Model ($χ\_{L}^{2}$) | 0.2757 | 3.9625 | 2.5383 |
| Exponential Model ($χ\_{E}^{2}$) | 0.1413 | 3.9951 | 2.5866 |
| Quadratic Model ($χ\_{Q}^{2}$) | 0.1484 | 3.8120 | 1.8232 |
| Cubic Model ($χ\_{C}^{2}$) | 0.1491 | 3.2175 | 1.3982 |

The tabulated values of chi-square ($χ^{2}$) at 1% and 5% levels of significance with 17 degrees of freedom are given, respectively, by

$χ\_{0.01,17}^{2}=33.41$ and $χ\_{0.05,17}^{2}=27.59$

From Table 7, the following results are obtained:

1. $χ\_{L\left(S\_{i}\right)}^{2}<χ\_{0.01,17 }^{2}$and $χ\_{L\left(S\_{i}\right)}^{2}<χ\_{0.05,17}^{2} \left(i=1,2,3\right)$
2. $χ\_{E\left(S\_{i}\right)}^{2}<χ\_{0.01,17 }^{2}$and $χ\_{E\left(S\_{i}\right)}^{2}<χ\_{0.05,17}^{2} \left(i=1,2,3\right)$
3. $χ\_{Q\left(S\_{i}\right)}^{2}<χ\_{0.01,17 }^{2}$and $χ\_{Q\left(S\_{i}\right)}^{2}<χ\_{0.05,17}^{2} \left(i=1,2,3\right)$
4. $χ\_{C(S\_{i})}^{2}<χ\_{0.01,17 }^{2}$and $χ\_{C(S\_{i})}^{2}<χ\_{0.05,17}^{2} (i=1,2,3)$

Hence, on the basis of above results, the null hypotheses $H\_{0L}$, $H\_{0E}$, $H\_{0Q}$ and $H\_{0C }$ are accepted at 1% and 5% levels of significance. So, we conclude that the linear, exponential, Quadratic and cubic models fit the given time series data on soybean production for the concerned states S1, S2 and S3 of India.

**4. CONCLUSION**

The present paper deals with time series analysis of soybean production in major soybean producing states of India. The secondary time series data on soybean production pertaining to the period (2005-2022) have been utilized for the analysis. The trends and growth patterns of soybean production in various selected Indian states have been analyzed by applying established statistical models specifically, the linear, exponential, quadratic, and cubic models to the relevant time series data.

It has been observed from the empirical results of section 3 that the cubic model is more precise and suitable, as compared to the linear, exponential and quadratic models, to analyze the trends in soybean production across the selected Indian states- S1 (Karnataka), S2 (Maharashtra), and S3 (Madhya Pradesh). The growth patterns of soybean production in the states S1 (Karnataka) and S2 (Maharashtra) is increasing. Moreover, in the state S3 (Madhya Pradesh), we observe that the growth pattern of soybean production is slightly decreasing.

In order to test the “Goodness of Fit” of the linear, exponential, quadratic and cubic models for the states S1, S2 and S3, the chi-square test statistic values (i.e., $χ\_{L}^{2}$, $χ\_{E}^{2}$ ,$χ\_{Q}^{2} $and $χ\_{C}^{2}$) have been computed for the respective states. These values are then compared with the tabulated values of chi-square at 1% and 5% levels of significance. It has been observed that all the considered models fit the given time series data on soybean production for the concerned states.

The present study could be enhanced further by considering the scenario of soybean production in the other states of India. Moreover, on considering the benefits and usefulness of soybean, the potential farmers could be encouraged for its cultivation.

**References**

Abraham, E. R., Reis, J. G. M. D., Vendrametto, O., Neto, P. L. D. O. C., Toloi, R. C., Souza, A. E. D., and Morais, M. D. O. (2020). Time series prediction with artificial neural networks: An analysis using Brazilian soybean production. *Agriculture,* 10(10): 475.

Anonymous (2023). Agricultural Statistics at a Glance. *Directorate of Economics & Statistics, DAC&FW, Government of India.*

Filho, I. J. R., Corrêa, G. B., Freire, G. M., and Rezende, S. O. (2020). Forecasting future corn and soybean prices: an analysis of the use of textual information to enrich time-series. In *Proceedings*.

Gowda KE, S., Kumar M., Prakash, G., Gautam S., and Rajnish (2025). Assessment of growth rate and instability analysis of rice cultivation in some selected states of India. *International Journal of Agriculture Extension and Social Development,* 8(5), 797-802.

Gusso, A., Arvor, D., and Ducati, J. R. (2017). Model for soybean production forecast based on prevailing physical conditions. *Pesquisa Agropecuária Brasileira*, 52: 95-103.

Han, J., and Ng’ombe, J. N. (2023). The relation between wheat, soybean, and hemp acreage: a Bayesian time series analysis. *Agricultural and Food Economics*, 11(1): 1-12.

Kumar, M., Gowda KE, S., Prakash, G., and Menon, S. V. (2025). Assessment of Trend Pattern of Rice Production in Some Rice Growing States of India. *Journal of Experimental Agriculture International*, 47(1), 582-591.

Kumar, M., Prakash, G., and Rana, S. K. (2024). Statistical Modeling for Analysis of Growth and Trend Pattern of Wheat Production in Selected States of India. *Asian Journal of Research in Crop Science*, *9*(1), 66-75.

Larbi, P. A., and Green, S. (2018). Time series analysis of soybean response to varying atmospheric conditions for precision agriculture. *Precision agriculture*, 19(6): 1113-1126.

Masuda, T., and Goldsmith, P. D. (2009). World soybean production: area harvested, yield, and long-term projections. *International food and agribusiness management review*, 12(4): 1-20.

Parmar, P., and Devi, G. (2021). Growth and decomposition analysis of area, production and yield of soybean. *Gujarat Journal of Extension Education*, 32(1): 116-119.

Prashnani M, Dupare B, Vadrevu KP, and Justice, C. (2024) Towards food security: Exploring the spatio-temporal dynamics of soybean in India. *PLOS ONE,* 19(5): e0292005. <https://doi.org/10.1371/journal.pone.0292005>.

Ramteke, R., Gupta, G. K., and Singh, D. V. (2015). Growth and yield responses of soybean to climate change. *Agricultural research*, 4(3): 319-323.

Stepanov, A., Dubrovin, K., and Sorokin, A. (2022). Function fitting for modeling seasonal normalized difference vegetation index time series and early forecasting of soybean yield. *The Crop Journal*, 10(5): 1452-1459.

Thimmegowda, M. N., Manjunatha, M. H., Lingaraj, H., Soumya, D. V., Jayaramaiah, R., Sathisha, G. S., and Nagesha, L. (2025). Comparative analysis of machine learning and statistical models for cotton yield prediction in major growing districts of Karnataka, India. *Journal of Cotton Research*, 8(1), 6.

Thomasz, E. O., Corfield, K., Vilker, A. S., and Osman, M. (2023). Forecasting soybean production to enhance climate services for Agriculture in Argentina.*Climate Services*, 30: 100341.

Zeng, L., Wardlow, B. D., Wang, R., Shan, J., Tadesse, T., Hayes, M. J., and Li, D. (2016). A hybrid approach for detecting corn and soybean phenology with time-series MODIS data.*Remote Sensing of Environment*, 181: 237-250.