www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2022; SP-11(4): 1737-1742

www.thepharmajournal.com
Received: 28-02-2022
Accepted: 30-03-2022

Dr. Nageena Nazir
Division of Agricultural Statistics, Sher-e-Kashmir University of Agricultural Sciences and Technology, Kashmir, Jammu and Kashmir, India

Dr. SA Mir
Division of Agricultural Statistics, Sher-e-Kashmir University of Agricultural Sciences and Technology, Kashmir, Jammu and Kashmir, India

Dr. BS Dhekale
Division of Agricultural Statistics, Sher-e-Kashmir University of Agricultural Sciences and Technology, Kashmir, Jammu and Kashmir, India

Corresponding Author
Dr. Nageena Nazir
Division of Agricultural Statistics, Sher-e-Kashmir University of Agricultural Sciences and Technology, Kashmir, Jammu and Kashmir, India

Parametric and non-parametric statistical techniques for modelling and monitoring area, production, productivity trends of dry fruits under temperate condition

Dr. Nageena Nazir, Dr. SA Mir and Dr. BS Dhekale

Abstract
The present investigation is carried out to study the trends in dry fruit area, production and productivity in Jammu and Kashmir for the period 2000-2001 to 2017-2018 based on the parametric and nonparametric regression models. In parametric models different linear models are employed. The statistically most suited parametric models are selected on the basis of highest adjusted R2, significant regression co-efficient and co-efficient of determination (R2). Appropriate model is selected based on the model performance measures such as, Root Mean Square Error, Mean Absolute Error, Mean Absolute Percentage Error, assumptions of normality and independence of residuals. Nonparametric estimates of underlying growth functions are computed at each and every time points. Relative growth rates of crop productions are estimated based on the best fitted trend functions. In this study it is found that non parametric/semi parametric regression comes out to be a good fit for trend in dry fruit in comparison to parametric regression. Even semi parametric spline is selected as the best fit model for trend analysis.

Keywords: Non-parametric statistical techniques, modelling, monitoring area, production, productivity trends, dry fruits

Introduction
Horticulture is the art and science of growing and handling fruits, nuts, vegetables, herbs, flowers, foliage plants, woody ornamentals and turf. The horticulture industry serves as a great advantage to the state due to its monopoly on walnuts, almonds, pears, cherries, hazelnuts, pecan nuts, strawberries and kiwi fruits. Production of dry fruits has increased to 2.89 lakh metric tons in 2017-2018 from 0.96 lakh metric tons in 2001-02. Jammu and Kashmir exports its horticulture products within the country well as to the other countries and in turn earns a substantial foreign exchange for the state. During the year 2013- 14, the state exported 8561.03 metric ton of dry fruits both almond and walnut in shell and kernel and earned foreign exchange of Rs. 365.25 crore (Digest of Statistics 2015-16). The horticulture industry contributes nearly 60 percent of the states revenue and 22 percent of gross state domestic product (GSDP). Among dry fruits, walnut occupies maximum area and production followed by almond in both the UT's (Jammu & Kashmir and Ladakh) (Lone *et al*, 2014; Rather *et al*, 2013).
The growth rates of crops are mostly estimated by the linear regression models. However, it might be the case that these models may not fit the data well. Under such conditions it becomes essential to apply nonparametric and semi-parametric regression, which is based on fewer assumptions. In last few years, nonparametric regression and semi-parametric regression technique for functional estimation has become increasingly popular as a tool for data analysis. These techniques impose only few assumptions about shape of function and therefore it is more flexible than usual parametric regression approaches. Smoothing techniques are commonly used to estimate the function non-parametrically (Härdle, 1990) [10]. Nonparametric regression models avoid restrictive assumptions of the functional form of the regression function. Semi-parametric regression model combine the components of parametric and nonparametric regression models, by keeping the easy interpretability of the former and retains some of the flexibility of the latter. Various scientists *viz*., (Chandran, 2004) [3] has applied nonparametric regression to study the growth rates of total foodgrain production of India during the period 1987 to 2001. Teczan (2010) [22] has studied the nonparametric regression technique to find out the growth rate trends of various crops.

Sahu and Pal (2004) [18] used nonparametric regression (Lowess) and semi- parametric (spline) for modeling of pest incidences. Dhekale *et al.*, 2017 [6], employed the nonparametric regression model to study the trends of tea in India. Yasmeen *et al*, (2018, 2019) [24, 23] employed non parametric and parametric regression models to study the trend of Area, Production and productivity of Apple and cherry in Kashmir. The current study is aimed to develop appropriate parametric and nonparametric regression models to fit the trends in area, production and productivity of dry fruits in Kashmir.

Material & Method

For present study, to study the trends and growth rates, long term data for last 18 years pertaining to the area, production and productivity of Dry fruit is collected from Directorate of Horticulture.

The descriptive measures of central tendency and dispersions along with the simple and compound growth rates are used to explain the features of the data (Mishra *et al.*, 2012) [12].

Trend Models

Parametric Regression Models

To find out the path of the production process different parametric trend models are fitted. Among the fitted models, the best model is selected on the basis of their goodness of fit (R^2) value and significance of the coefficients. The dependent variable Y is area, production and productivity and independent variable X is the time points (years).

Non-parametric and semi-parametric regression models

The model considered here is of the form

$$y_i = m(x_i) + e_i \, , x_i = i/n, i = 1,2,\ldots,n$$

Where, y_i is observation of i^{th} time point, $m(.)$ is trend function which is assumed to be smooth and are e_i random errors with mean zero and finite variance. Since there is no assumption of parametric form of function, this approach is flexible and robust to deviations from an assumed model form. To obtain an estimate of the mean response value at a point X, most of the smoothers are averaging the Y – values of observations having predictor values closer to the target value X. The averaging is done in neighborhoods around target value. The main decision to be made in any of the smoothing techniques is to fix the size of neighborhood which is typically expressed in terms of an adjustable smoothing parameter or bandwidth. Intuitively, large neighborhoods will provide an estimate with low variance but potentially high bias, and conversely for small neighborhoods. Lowess regression, introduced i Y th i) (·mi ε) (·m by Cleveland (1979), is obtained on the basis of the data points around it within a band of certain width. The point xi is the midpoint of the band. The data points within the band are assigned weights in a way so that xi has the highest weight. The weights for the other data points decline with their distance from xi according to a weight function. The weighted least squares method is used to find the fitted value corresponding to xi, which is taken as the smoothed value. The procedure is repeated for all the data points. The spline method of estimation make use of the penalized least squares method (Simonoff, 2012) [21], which balances the fitting of the data closely. The objective is to estimate *m* by means of a function that fits the data well and is as smooth as possible. A measure of smoothness of mis the integral of the square of its second derivative as

$$\sum_{i=1}^{n}(Y_i - m(x_i))^2 + \lambda\int_a^b (m^{''}(x))^2 \, dx$$

Where $\lambda > 0$ is a fixed constant and $x_i \in [a,b]$ the first term is the sum of squares of the residuals; it provides a measure of how well the function *m* fits the data. The integral of the above equation is a measure for the roughness / smoothness of the function. The functions which are highly curved will result in a large value of the integral; straight lines result in the integral being zero. The roughness penalty, controls the emphasis which one wishes to place on smoothness. By increasing the value of, one places more emphasis on smoothness; as becomes large the function approaches a straight line. On the other hand, a small value of λ emphasizes the fit of *m* to the data points: as λ approaches a function that interpolates the data points.

Result and Discussion

The maximum growth rate is observed in production of dry fruits over the years, whereas the minimum growth rate is exhibited by area of the dry fruits.

The positive compound growth of production (4.31 per annum) reveals that there is no decrease in the production of dry fruitover the years with a maximum of 0.289 million kilogram and minimum of 0.096 million kilogram. Similarly, the simple growth rate (7.59 per annum) is observed in production indicates an increase in the production of dry fruits in Kashmir over the years (Table 1).

This is due to the fact that a large area of land is being brought under agriculture we have noticed a compound growth rate of area (0.75 per annum) under dry fruit cultivation indicating that a large portion of the land is being utilized for the latter.

Trend analysis of area, production and productivity

Parametric techniques

The linear models used here are the cubic model or the third degree polynomial model and the quadratic model or second degree polynomial model. The value of *b3* for area is negative which indicates that area under dry fruit cultivation decreased in the last part of the cultivation period and the value of *b1* and *b2* being positive clearly indicates that there was an increase in the cultivation area. Further, the negative value of *b3* for production is an indication of the decrease in the production during the last period of the study and the positive values of *b1* and *b2* indicates an increase in the production (Table-2).

Non-parametric and semi-parametric regression

Trend analysis of area, production and productivity using nonparametric (Loess) and semi-parametric (spline) regression are presented in the tables 3, 4 and 5. In Table 3 the value of R2 is 0.90 for Loess and 0.95for Spline regression. The AICc, RMSE, MAPE, MAE, MaxAPE and MaxAE values comes out be small for Spline regression for the area under dry fruit cultivation. The area under the dry fruit cultivation has increased over the years of study and is shown in figure 1.

On comparing the values of AICc, RMSE, MAPE, MAE, MaxAPE and MaxAE for production and productivity the spline regression has the smallest values. The increasing trend in the production and productivity over the years of study is shown in the figure 3 and 5. It can be observed that upto 2013-14 there is sharp increase in production and productivity. However, a decline in production and productivity can also be observed during the year 2014-15 is observed which is due to the floods that occurred during the said year (Islam and Shrivastava, 2017) [11].

The values of area are initially fitted at the smoothing parameters in order to obtain the best fit of the data points we obtain the graph of the data points in the neighborhood of the smoothing parameters and look for the curve which covers all the points of the data. The one which covers maximum points is the best fit of the data points. In figure 2 the smooth curve fits are obtained for area in the neighborhood of smoothing parameters i.e., at 0.23, 0.40, 0.528 and 0.835. It is observed that the best fit is obtained at smooth=0.528. In figure 4 smooth fits for production are plotted in the neighborhood of the smoothing parameter at 0.34, 0.50, 0.98 and 1.50 and it is observed that the best fit obtained for smooth=0.98. Figure 6 provides the fits for productivity in the neighborhood of the smoothing parameters i.e., at smooths equal to 0.405, 0.806, 1.40 and 1.61. The best fit is observed to be at the smooth=0.806.

Even values of RMSE, MAE, MAPE, MaxAE and MaxAPE for area production and productivity of Kashmir for non-parametric regression has observed lower values than the parametric regression (Tables 3, 4, 5). This is clear indication of the superiority of these techniques over the parametric models. These models perform very well in visualizing the past trends where the parametric models fails to.

Among the nonparametric and semi-parametric regression, the spline regression has shown the lowest values of AICc, RMSE, MAPE, MAE, MaxAPE and MaxAE for area, production and productivity of dry fruitin Kashmir hence spline regression is the best fitted model for dry fruit production in Kashmir (Fig. 5). Various scientist viz. Aydin (2007) and Pal (2011) observed similar results where the spline gave the better results than the Loess smoothing.

Table 1: Performance of Dry fruit production in Kashmir during 2001-2018

Area ('000 hectare)		Production ('000 MT)		Productivity (Quintals per hectare)	
Minimum	79.36	Minimum	96.32	Minimum	12.11
Maximum	113.01	Maximum	289.83	Maximum	31.82
Mean	97.70	Mean	185.37	Mean	18.88
Standard Deviation	9.94	Standard Deviation	55.14	Standard Deviation	6.78
CV(%)	10.38	CV(%)	29.08	CV(%)	35.07
Skewness	-0.15	Skewness	0.18	Skewness	0.738
SGRA	1.61	SGRA	7.59	SGRA	5.23
CGRA	0.75	CGRA	4.31	CGRA	3.51

CV= coefficient of variation, SD= standard deviation, SGAR= simple growth rate per annum, CGAR= compound growth rate per annum

Table 2: Trends in area, production and productivity of Dry fruits in Kashmir Division

	R-Square	Constant b0	b1	b2	b3	RMSE	MAPE	MAE	MaxAPE	MaxAE
Area	0.87	35.38	2.90	0.19	-0.018	2.40	1.80	1.82	4.96	5.04
Production	0.92	59.93	3.97	0.75	-0.19	7.31	3.87	6.46	11.99	12.35
Productivity	0.90	11.39	0.50	0.02	0.004	1.18	4.99	0.96	10.74	4.59

Area in '000 hectares, Production in '000 metric tons, Productivity in quintals per hectare

Table 3: Trends in area of Dry fruit in Kashmir using non-parametric and semi- parametric regression

Loess	Splines	
Bandwidth	0.41	0.27
R-Square	0.90	0.95
AICc	4.11	2.12
RMSE	0.82	0.64
MAPE	0.18	0.10
MAE	0.77	0.72
MaxAPE	3.09	2.39
MaxAE	1.65	1.39

Table 4: Trends in production of Total Dry Fruits in Jammu and Kashmir using non-parametric and semi- parametric regression

Loess	Splines	
Bandwidth	0.62	0.43
R2	0.93	0.98
AICc	6.17	3.15
RMSE	9.58	2.05
MAPE	0.246	0.11
MAE	2.55	1.18
MaxAPE	17.28	4.24
MaxAE	26.23	5.46

Table 5: Trends in productivity of Dry fruit in Kashmir using non-parametric and semi- parametric regression

Loess	Splines	
Bandwidth	0.59	0.41
R2	0.94	0.97
AICc	6.68	3.15
RMSE	5.17	2.01
MAPE	0.27	0.15
MAE	2.42	1.16
MaxAPE	9.15	4.23
MaxAE	10.20	5.59

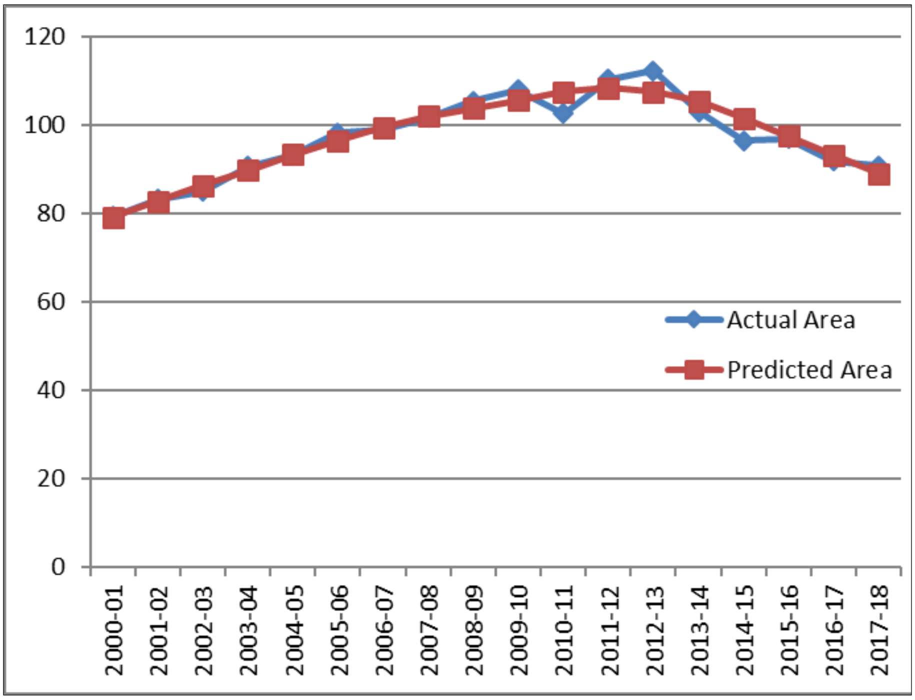

Fig 1: Observed and expected trends of area under Dry fruit cultivation using spline in Kashmir

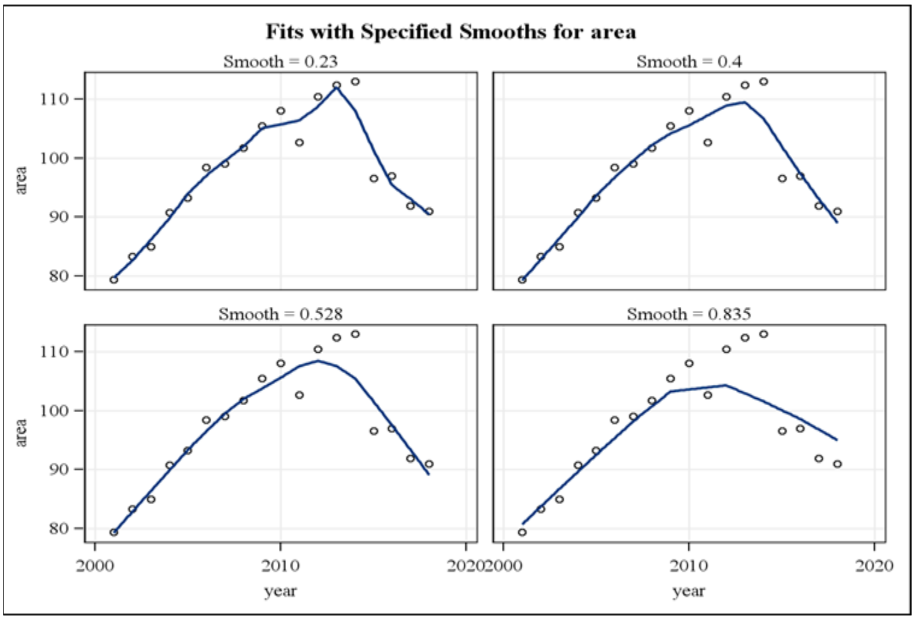

Fig 2: Fits with specified smooths for area

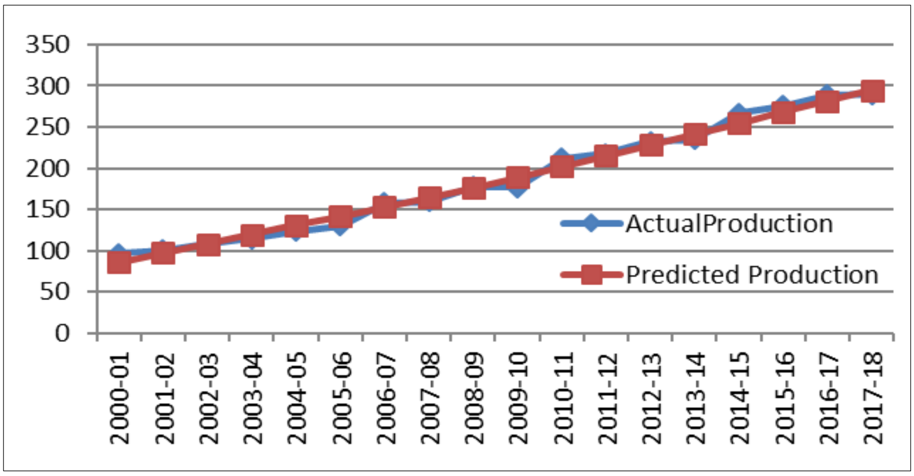

Fig 3: Observed and expected trends of Production under Dry fruit cultivation using spline in Kashmir

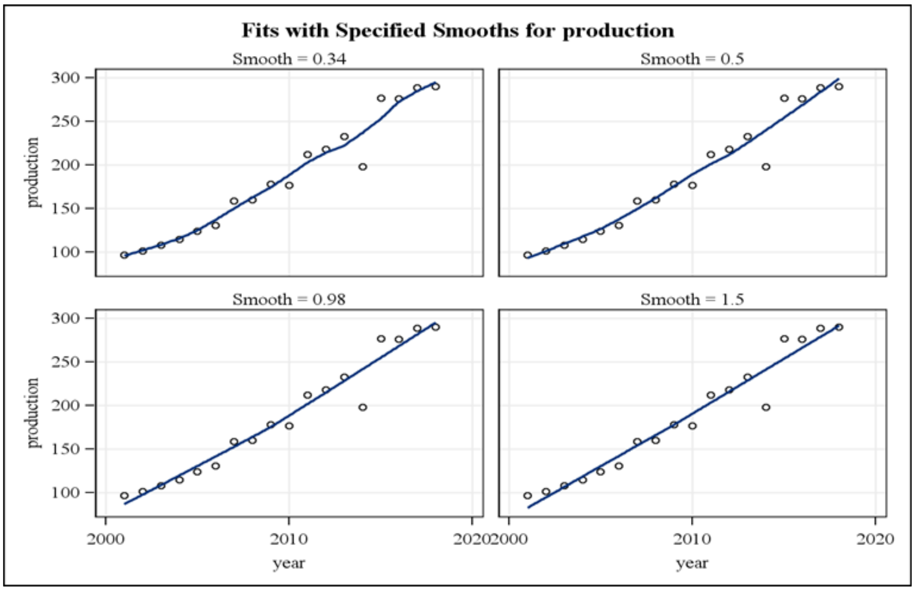

Fig 4: Fits with specified smooths for production

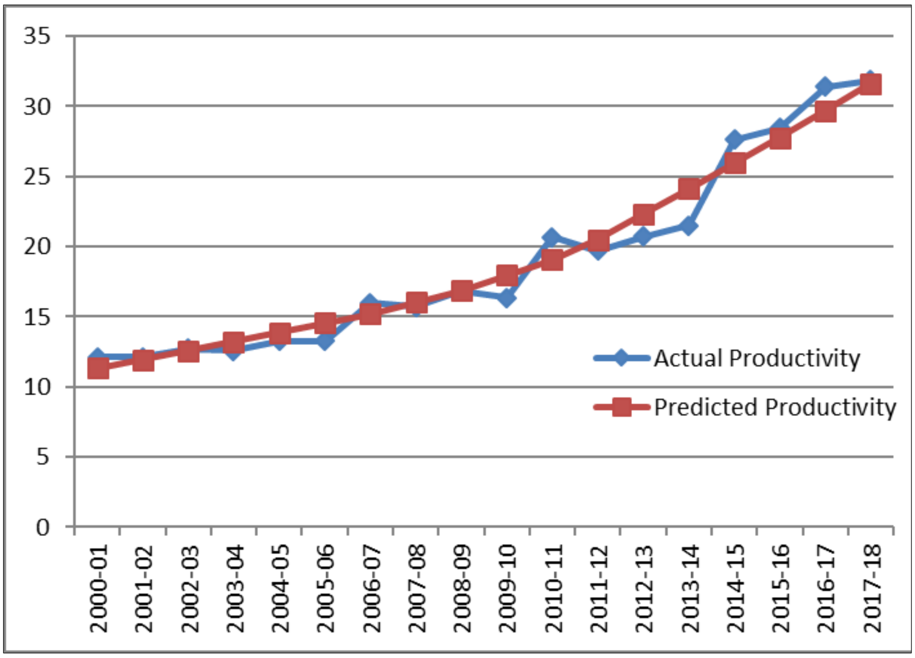

Fig 5: Observed and expected trends of Productivity under Dry fruit cultivation using spline in Kashmir

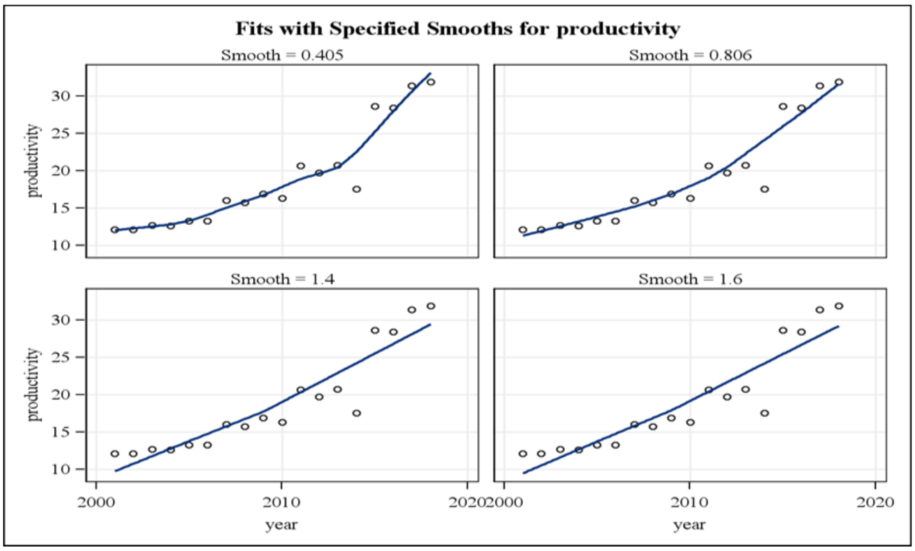

Fig 6: Fits with specified smooths for productivity

Conclusion

- From the results it is observed that values of R-Square AIC, RMSE, and other criterion for area production and productivity of total dry fruits using Semi and non-parametric regression has observed lower values than the parametric regression (Tables 3, 4, 5) indicating the superiority of these techniques over the parametric models.
- Among the nonparametric and semi-parametric regression, the spline regression has shown the lowest values of R-Square AICc, RMSE, MAPE, MAE, MaxAPE and MaxAE for area, production and productivity of total dry fruits in Jammu & Kashmir hence spline regression is the best fitted model.
- Thus it can be concluded that parametric regression usually utilized in studying the trend seems not to perform better than the nonparametric and semi-parametric regression. And out of the nonparametric and semi-parametric regression methods the semi-parametric regression (spline) is the best fit for the trend analysis of the of total dry fruit production in Jammu & of Kashmir.

References

1. Aydin D. A comparison of the nonparametric regression models using smoothing spline and kernel regression. World Academy of Science, Engineering and Technology. 2007;36:253-57.
2. Bryon Parman J, Allen Featherstone M. A comparison of parametric and nonparametric estimation methods for cost frontiers and economic measures Journal of Applied Economics, 2019, 22(1).
3. Chandran K. A Study on Some Aspects of Nonlinear Statistical Models in Agriculture, Indian Agricultural Statistics Research Institute Indian Agricultural Research Institute: New Delhi.
4. Cleveland WS. The elements of graphing data (revised edition). Hobart Press, Summit, NJ, USA, 1994.
5. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis, John Wiley & sons, Inc, 2003.
6. Dhekale B, Sahu P, Vishwajith K, Mishra P, Narsimhaiah L. Application of Parametric and Nonparametric Regression Models for Area, Production and Productivity Trends of Tea (*Camellia sinensis*) in India, Indian Journal of Ecology. 2017;44:192-200.
7. Pandey G, Shukla SK. The Walnut Industry in India-Current Status and Future Prospects. International Journal of Fruit Science. 2007;6(4):67-75.
8. Mir GM, Uzma Iqbal, Owais Nisa. Portable Walnut Grading Machine Best Replacement for perforated Cylindrical Grader. International Journal of Research & Innovation. 2016;3(8):82-86.
9. Hamdy Mahmoud FF. Parametric versus Semi and Nonparametric Regression Models, 2019. arXiv:1906.10221v1 [stat.ME].
10. Härdle W. Applied Nonparametric Regression, Cambridge university press, 1990.
11. Islam RT, Shrivastava S, A Study on Area, Production and Productivity of Apples in J&K from2006-07 to 2015-16, International Journal of Scientific Research and Management. 2017;5:6513-6519.
12. Mishra P, Sahu P, Bajpai P, Nirnjan H. Past Trends and Future Prospects in Production, and Export Scenario of Tea in India, International Review of Business and Finance. 2012;4:25-33.
13. Mohammad Taufique, Vajahat Khursheed. Walnut Industry in Jammu and Kashmir: A Geographical Analysis. Res. J Humanities and Social Sciences. 2018;9(4):793-798.
14. Nageena Nazir, SA Mir, Dhekale BS. Modeling and Monitoring Walnut (*Juglans regia*) Area & Production Based on Parametric and Non-Parametric Regression Models. Int. J Curr. Microbiol. App. Sci. 2020;9(10):75-84.
15. Pal S. Modeling Pest Incidences in Some Agricultural Crops, Thesis Submitted to the Bidhan Chandra Krishi Viswavidyalaya under Department of Agricultural Statistics, 2011.
16. Paola Pollegioni, Keith Woeste, Francesca Chiocchini, Stefano Del Lungo, Marco Ciolfi, Irene Olimpieri, Virginia Tortolano, Jo Clark, Gabriel E. Hemery, Sergio Mapelli, Maria Emilia Malvolti. Rethinking the history of common walnut (*Juglans regia* L.) in Europe: Its origins and human interactions, 2017. https://doi.org/10.1371/journal.pone.0172541.
17. Suba SS, Rajarathinam A. Modeling and Monitoring Cereals Crop Production Based On Parametric and Nonparametric Regression Models. International Journal of Scientific Research Mathematical and Statistical Sciences. 2019;6(4):16-29.
18. Sahu PK, Pal S. Modelling Pest Incidence - Nonparametric Approach. Paper presented in 24th International Symposium on Forecasting, Sydeny, 2004.
19. Shah UN, Mir JI, Ahmed N, Fazili KM. Assessment of Germplasm Diversity and Genetic Relationships among walnut Genotypes through Microsatellite Markers. J Saudi Soc. Agri. Sci. 2016, 1-12.
20. Sheikh SH, Tripathi A, Socio- Economic Conditions of Apple Growers of Kashmir Valley: A Case Study of District Anantnag, International Journal of Educational Research and Technology. 2013;4:30-39.
21. Simonoff JS. Smoothing Methods in Statistics, Springer Science & Business Media, 2012.
22. Teczan N. Analysis of the Factors Affecting Productivity Using Non Parametric Regression Method. 10th Global Conference on Business & Economics, Rome, Itlay, 2010.
23. Yasmeena Ismail, Mir SA, Nageena Nazir, Wani MH, Wani SA, Pukhta MS. Trend Analysis of Area, Production and Productivity of Cherry in Jammu and Kashmir. Int. J Curr. Microbiol. App. Sci. 2019;8(2):2135-2144.
24. Yasmeena Ismail, Mir SA, Nageena Nazir SA. Utilization of Parametric and Nonparametric Regression Models for Production, Productivity and Area Trends of Apple (*Malus domestica*) in Jammu and Kashmir, India. Int. J Curr. Microbiol. App. Sci. 2018;7(04):267-276.