International Journal of Statistics and Applied Mathematics 2025; 10(7): 14-21

International Journal of Statistics and Applied Mathematics

**ISSN:** 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(7): 14-21

https://www.mathsjournal.com
Received: 05-07-2025
Accepted: 06-08-2025

**KS Tailor**
Assistant Professor, Department of Statistics, Maharaja Krishnakumarsinhji Bhavnagar University, Bhavnagar, Gujarat, India

**Corresponding Author:**
**KS Tailor**
Assistant Professor, Department of Statistics, Maharaja Krishnakumarsinhji Bhavnagar University, Bhavnagar, Gujarat, India

# An analysis of area, production and productivity of sugarcane in India using multivariate adaptive regression spline

**KS Tailor**

**Abstract**
Agriculture is the primary source of livelihood which forms the backbone of our country. Current challenges of water shortages, uncontrolled cost due to demand-supply, and weather uncertainty necessitate farmers to be equipped with smart farming. In particular, low yield of crops due to uncertain climatic changes, poor irrigation facilities, reduction in soil fertility and traditional farming techniques need to be addressed. Machine learning is one such technique employed to predict crop yield in agriculture. Various machine learning techniques such as prediction, classification, regression and clustering are utilized to forecast crop yield. In this paper, a non-parametric model called multivariate adaptive regression spline (MARS) is used to predict the sugarcane production in India. MARS (Friedman, 1991) is a non-parametric model that divides data into various partitions and formulates the relationship between independent and dependent spatial drivers.

**Keywords:** MARS, sugarcane, productivity, area

## 1. Introduction
Sugarcane is one of the major non-food grain crops in India. Sugar production has emerged as the biggest agro industries in the rural area of India during the last few decades. It has made a considerable impact on the economy of farmers particularly in irrigated areas. India is the second-highest producer of sugarcane in the world after Brazil. The largest producer of sugarcane in India is Maharashtra**,** which produced over 138 lakh tonnes of sugarcane in 2022-23. Uttar Pradesh, Karnataka, and Maharashtra together contribute to 80% of the total sugarcane production in India. Maharashtra produces 61.32 million tonnes of sugarcane on average per year. Sugarcane is a multipurpose crop, used in making sugar, jaggery, khansari, molasses, and even paper. In India, approximately 60% of the population is involved in agriculture and among the many crops cultivated in the nation, sugarcane is one of the most important Kharif crops. The climate of the country supports the plantation of sugarcane throughout the year.
Sugarcane is a tropical and subtropical crop that requires a hot and humid climate to grow. A tall, perennial grass species known as sugarcane or sugar cane is utilized in the production of sugar. The 2-6 m tall plants have thick, jointed, fibrous stalks that are rich in sucrose and accumulate in the internodes of the stalks. Several other states grow sugarcane in addition to Uttar Pradesh, which is the largest producer of sugarcane in India.
In this paper, an attempt is made to analyze the pattern of sugar cane crop production in India using a non-parametric model called multivariate adaptive regression spline (MARS). Secondary data is collected from the database of Directorate of Economics and Statistics, Ministry of Agriculture & Farmers' Welfare, New Delhi.

## Materials and Methods
**Materials:** This paper is based on sugarcane production in India. The data related to area, production and productivity of sugarcane in India is collected from the database of Directorate of Economics and Statistics, Ministry of Agriculture & Farmers' Welfare, New Delhi. The data for the years 2000-2001 through 2022-2023 has been collected from the website of

ministry of agriculture and farmers. The data is classified into seven different variables, namely V1= area in '000 hectares, V2 = production in '000 tones, V3 = productivity in tones per hectares, V4 = cane crushed in '000 tones, V5 = sugar recovery in %, V6 = sugar in '000 tones, V7 = number of sugar factories. Descriptive statistics for these variables are given in table 1.

**Table 1:** Descriptive Statistics

| | n | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| V1 | 23 | 3662 | 5883 | 4719.70 | 488.914 |
| V2 | 23 | 233862 | 494228 | 339698.8 | 61620.428 |
| V3 | 23 | 59.40 | 84.91 | 71.5217 | 6.87230 |
| V4 | 23 | 0 | 356400 | 220193.0 | 76315.792 |
| V5 | 23 | .00 | 11.01 | 9.9035 | 2.17449 |
| V6 | 23 | 0 | 35760 | 22809.96 | 8016.117 |
| V7 | 23 | 400 | 538 | 489.48 | 40.669 |

V1= area, V2 = production, V3 = productivity, V4 = cane crushed,
V5 = sugar recovery, V6 = sugar, V7 = number of sugar factories

**Five different MARS models have been evaluated. These models are as follows.**

- **Model 1:** Dependent variable: V2 = production, Independent variable: V1= area
- **Model 2:** Dependent variable: V3 = productivity, Independent variable: V1= area, V2 = production
- **Modal 3:** Dependent variable: V5 = sugar recovery, Independent variable: V3 = productivity, V4 = cane crushed and V7 = number of sugar factories
- **Modal 4:** Dependent variable: V4 = cane crushed, Independent variable: V1= area, V2 = production and V7 = number of sugar factories
- **Model 5:** Dependent variable: V6 = sugar, Independent variable: V3 = productivity, V4 = cane crushed, V5 = sugar recovery and V7 = number of sugar factories

**Methods**

Multivariate adaptive regression spline (MARS) is a data mining technique that can be used for solving regression type problems. (Hastie *et al.*, 2001) [52].

MARS is an effective machine learning algorithm that define the relation between a dependent variable and a set of independent variables. (Celik *et al.*, 2019) [25].

It is a non-parametric procedure, for invention adaptive regressions that uses piecewise basis functions to define relationships between a dependent variable and a set of estimations. MARS allows for the capture of linear and additive relationships and for the separation in excess of all nodes at each step, rather than just the terminal ones. Hence, MARS compose a bended regression line to fit the data from subgroup to subgroup and from spline to spline. (Friedman, 1991) [47].

In every spline, MARS splits the data anymore inside many subgroups. Several knots are constituted by MARS. These knots can be established between distinct input variables or distinct intervals in the same input variable, to separate the subgroups. The data of each subgroup are called basis function (BF). The model takes the form of an expansion in product spline basis functions, where the number of basic functions as well as the parameters associated with each one (product degree and knot locations) are automatically determined by the data (Friedman, 1991; Sephton 2001) [47, 63].

The MARS algorithm constructs models from two sided functions of the predictors (x) of the form:

$$(x-t)_+ = \begin{cases} & x > t \\ 0, otherwise & \end{cases}$$

These serve as basis functions for linear or nonlinear expansion that approximates some true underlying function *f(x).*

The MARS model for a response variable *y*, and M terms, can be given in the sequent equation:

$$y = f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m K_{km}(X_{v(k,m)})$$

Where the aggregate is over the M terms in the model and $\beta_0$ is an intercept, $\beta_m$ is a coefficient of basis functions, $K_{km}(X_{v(k,m)})$ is a basis function, here $v(k,m)$ is an index of a predictor for an m[th] component of k[th] product (Hastie *et al.*, 2001) [52]. Function H is defined as,

$$H_{km}(X_{v(k,m)}) = \prod_{k=1}^{K} h_{km}$$

Where $X_{v(k,m)}$ is the predictor in the k[th] of the m[th] product. Here, k is a parameter interaction order. For order of interactions *K*=1, the model is additive and for *K=2* the model pairwise interactive (Friedman, 1991) [47].

During forward step, a number of basis functions are added to the model according to a predetermined maximum which should be considerably larger (twice as much at least) than the optimal (best least-squares fit) (Hastie *et al.*, 2001) [52].

A backward procedure is applied in which the model is pruned by removing those basis functions that are associated with the smallest increase in the goodness of-fit. Generalized Cross Validation error is a measure of the goodness of fit that takes into account both the residual error and the model complexity as well. It is formulated by (Koronacki and Ćwik 2005) [57].

$$GCV = \frac{\sum_{i=1}^{N}(y_i - f(x_i))^2}{\left[1-\frac{C}{n}\right]^2}$$

With,

$$C = 1 + cd$$

Where n is the number of cases in the data set, d is the effective degrees of freedom, which is equal to the number of independent basis functions. The quantity C is the penalty for adding a basis function (Hastie *et al.*, 2001) [52].

To comparatively test the estimate criteria of MARS, the following goodness of fit criteria were used (Willmott and Matsuura, 2005; Liddle, 2007; Takma *et al.*, 2012; Eyduran *et al.*, 2019) [66, 64, 33]:

1. Pearson correlation coefficient (r) between the actual and predicted dependent variable values,
2. Coefficient of determination,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2}$$

3. Adjusted Coefficient of determination,

$$Adj.R^2 = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2}$$

4. Root mean square error (RMSE) given by following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

5. Standard deviation ratio $SD_{ratio}$

$$SD_{ratio} = \sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \overline{\epsilon})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

6. Akaike Information Criterion (AIC)

$$AIC = nlog\sum_{i=1}^{n}\left(\frac{(Y_i - \hat{Y}_i)^2}{n}\right) + 2k$$

7. Corrected Akaike Information Criterion (AICc)

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

Where k is the number of selected terms and n is the sample size. (Hu, 2007) [53].

Here, $Y_i$ is the observed dependent variable value of i$^{th}$ variable, $\hat{Y}_i$ is the predicted dependent values of i$^{th}$ variable, $\overline{Y}$ is the average of the dependent variable of the variable, $\varepsilon_i$ is the residual value of i$^{th}$ variable, $\overline{\epsilon}$ is the average of the residual values, k: number of the selected terms in the model, and n: total number of variable. The residual value of each observation is expressed as $\varepsilon_i = Y_i - \hat{Y}_i$.

The MARS analysis was performed using the earth package of R software (Milborrow, 2011; Milborrow, 2018; R Core Team, 2014; Eyduran *et al.*, 2019) [58, 59, 33].

**Results:** In this study, five different MARS models were developed to predictive five different dependent variables such as production, productivity, sugar recovery, crane crushed and sugar. The goodness-of-fit statistics (r, $R^2$, Adj. $R^2$, SDratio, AIC, AICc and GCV) were calculated to measure predictive performances of the developed MARS models. Results of predictive performances of the MARS models are reported in Table 2. It was understood that the fitted MARS models had high predictive accuracy (Table 2). Grzesiak and Zaborski (2012) [50] reported that the model having SD ratio less than 0.40 had a good fit, as also reported by Eyduran *et al.* (2019) [33]. For instance, it was determined that 99.9% of total variability in production was explained.

**Table 2:** Goodness of fit criteria for MARS algorithm.

| Variables | r | $R^2$ | Adj $R^2$ | RMSE | SDratio | AIC | AICc | GCV |
|---|---|---|---|---|---|---|---|---|
| V2 | 0.999 | 0.999 | 0.999 | 1555.779 | 0.026 | 348 | -46 | 2420450 |
| V3 | 0.9023 | 0.9983 | 0.997 | 0.277 | 0.041 | -49 | -46 | 0.07658 |
| V4 | 0.9994 | 0.999 | 0.999 | 2386.728 | 0.032 | 368 | 371 | 5696469 |
| V5 | 0.9980 | 1 | 1 | 0.133 | 0.063 | -83 | -79 | 0.0177 |
| V6 | 0.9995 | 0.999 | 0.999 | 242.468 | 0.031 | 263 | 266 | 58791 |

**Modal 1**

In model 1, variable V2 is considered as dependent variable. All the models in this paper are created by R software using earth package.

Summary of mars modal 1 (V2~.,) is presented below, coefficients (Intercept) 337725.56

- h(4857-V1)-66.30
- h(V1-4857) 82.18
- h(70-V3) -4186.17
- h(V3-70) 5029.91

Selected 5 of 5 terms, and 2 of 28 predictors (nprune=100)
Termination condition: Reached maximum RSq 0.9990 at 5 terms

**Importance:** V3, V1, year2001-02-unused, year2002-03-unused, year2003-04-unused, year2004-05-unused, Number of terms at each degree of interaction: 1 4 (additive model) GCV 2420450 RSS 55670346 GRSq 0.9993336 RSq 0.9993336 CVRSq 0.9975392

Hence the prediction equation in terms of four basic functions for the mars model can be written as,

$$V2 = 337725.56 - 66.30 * h(4857 - V1) + 82.18 * h(V1 - 4857) - 4186.17 * h(70 - V3) + 5029.91 * h(V3 - 70)$$

It shows that 5 out of 5 terms were used from the original1 predictor. It can be seen that a variable V1 and V3 are included with a knot at 4857 and 70 respectively, the coefficient for h(4857-V1) is -66.30 and the coefficient for h(70-V3) is -4186.17. It is also clear that total 99.75% variability is explained by the model.

Various plots of mars model 1 is presented below.

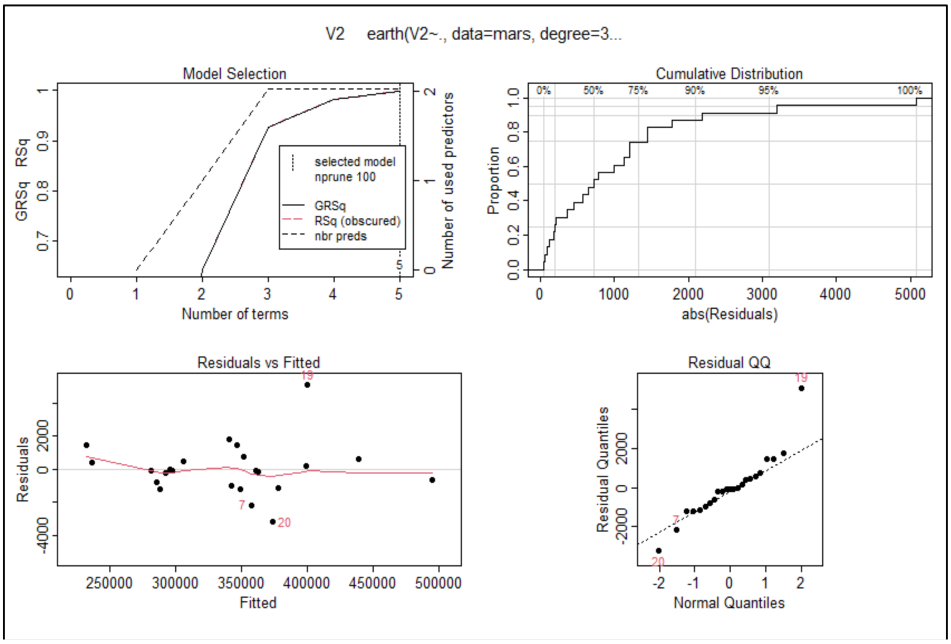


**Fig 1:** model 1 plot

Figure 1 illustrates the model selection plot that graphs the GCV (left-hand y-axis and solid black line) based on the number of terms retained in the model (x-axis) which are constructed from a certain number of original predictors (right-hand y-axis). It also includes cumulative distribution curve, residuals Vs. fitted curve and residual QQ plots.

**Modal 2**

In this variable V3 is considered as dependent variable. Summary of mars modal 2 (V3~.,) is presented below,

Coefficients (Intercept) 72.065731

- h(4857-V1) 0.015688
- h(V1-4857) -0.015807
- h(348188-V2) -0.000232
- h(V2-348188) 0.000195

Selected 5 of 5 terms, and 2 of 28 predictors (nprune=100)
Termination condition: RSq changed by less than 0.001 at 5 terms Importance: V2, V1, year2001-02-unused, year2002-03-unused, year2003-04-unused, year2004-05-unused,
Number of terms at each degree of interaction: 1 4 (additive model) GCV 0.07657788 RSS 1.761291 GRSq 0.9983049 RSq 0.9983049 CVRSq 0.9434533Hence the prediction equation in terms of four basic functions for the mars model can be written as,

$$V3 = 72.065731 + 0.015688 * h(4857 - V1) - 0.015507 * h(V1 - 4857) - 0.000232 * h(348188 - V2) + 0.000195 * h(V2 - 348188)$$

It shows that 5 out of 5 terms were used from the original 2 of 28 predictors. It can be seen that a variable V1 is included with a knot at 4857, the coefficient for h(4857-V1) is 0.015688, and variable V2 is included with a knot 348188, the coefficient for h(348188-V2) is -0.000232. It is also clear that total 99.83% variability is explained by the model.
Various plots of mars model 2 is presented below.

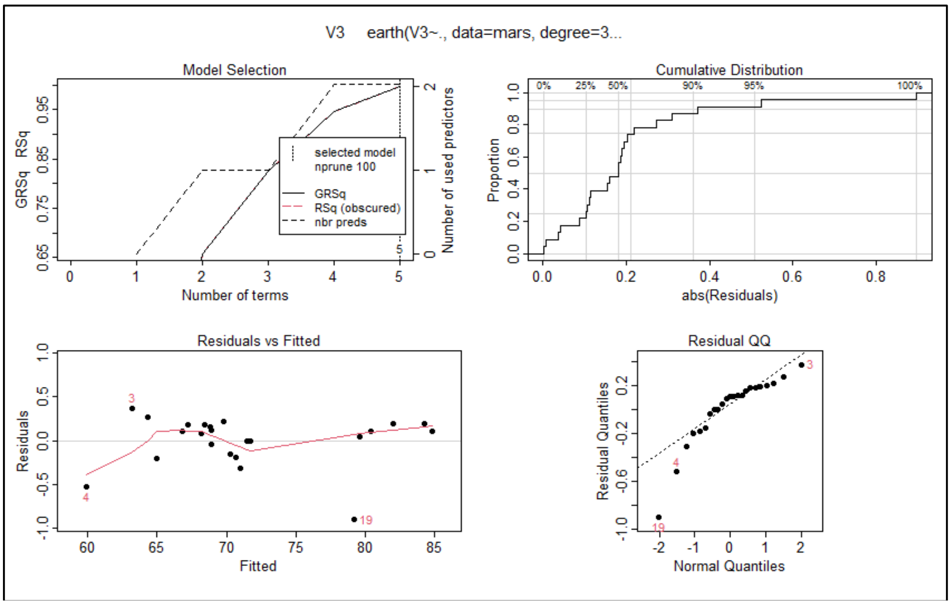


**Fig 2:** Model 2 plot

**Model 3**

This model is created by taking V5 variable as dependent variable and its summary is presented below,

Coefficients (Intercept) 10.2533333

- year2015-16 0.3666667
- year2017-18 0.4766667
- year2018-19 0.7566667
- year2019-20 0.6066667
- year2022-23-10.2533333

Selected 6 of 6 terms, and 5 of 28 predictors (nprune=100) Termination condition: RSq changed by less than 0.001 at 6 terms Importance: year2022-23, year2018-19, year2019-20, year2017-18, year2015-16, year2001-02-unused, year2002-03-unused, Number of terms at each degree of interaction: 1 5 (additive model) GCV 0.01770435 RSS 0.4072 GRSq 0.9960856 RSq 0.9960856 CVRSq -5.011345

Hence the prediction equation in terms of five basic functions for the mars model can be written as,

$$V5 = 10.2533333 + 0.366667 * (year\ 2015 - 16) + 0.4766667 * (year\ 2017 - 18) + 0.7566667 * (year\ 2018 - 19) + 0.6066667 * (year\ 2019 - 20) - 10.253333 * (year\ 2022 - 23)$$

It shows that 6 out of 6 terms were used from the original 5 out of 28 predictors. It is also clear that total 99.61% variability is explained by the model.

Various plots of mars model 3 is presented below.

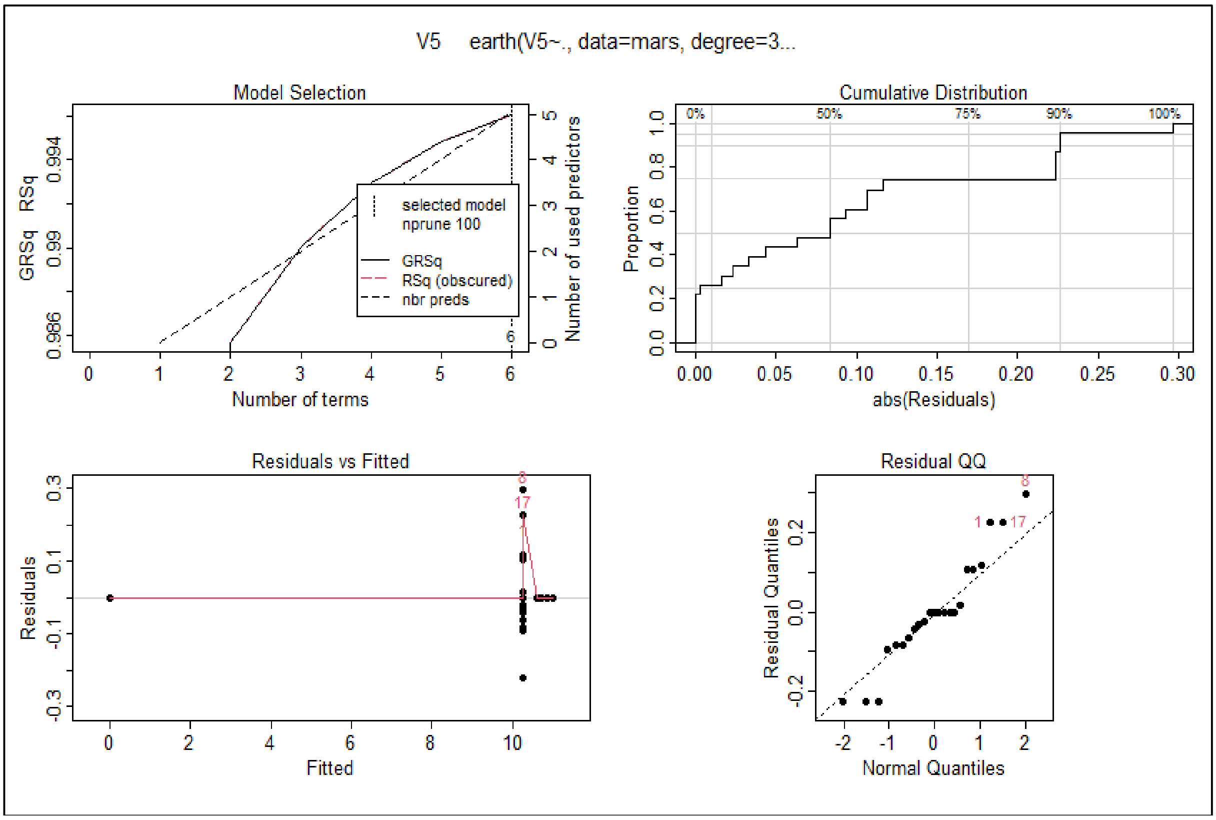


**Fig 3:** Model 3 plot

**Model 4:** This model is created by taking V4 variable as dependent variable and its summary is presented below,

Coefficients (Intercept) 240007.028

- h(10.25-V5) 45.344
- h(V5-10.25) -28431.212
- h(24394-V6) -9.849
- h(V6-24394) 9.801

Selected 5 of 5 terms, and 2 of 28 predictors (nprune=100) Termination condition: RSq changed by less than 0.001 at 5 terms Importance: V6, V5, year2001-02-unused, year2002-03-unused, year2003-04-unused, year2004-05-unused, Number of terms at each degree of interaction: 1 4 (additive model) GCV 5696469 RSS 131018782 GRSq 0.9989775 RSq 0.9989775 CVRSq 0.7343142 Hence the prediction equation in terms of four basic functions for the mars model can be written as,

$$V4 = 240007.028 + 45.344 * h(10.25 - V5) - 28431.212 * h(V5 - 10.25) - 9.849 * h(24394 - V6) + 9.801 * h(V6 - 24394)$$

It shows that 5 out of 5 terms were used from the original 2 out of 28 predictors. It can be seen that a variable V5 is included with a knot at 10.25, the coefficient for h(10.25-V5) is -28431.212, and variable V6 is included with a knot 24394, the coefficient for h(V6-24394) is 9.801. It is also clear that total 99.90% variability is explained by the model.

Various plots of mars model 4 is presented below.

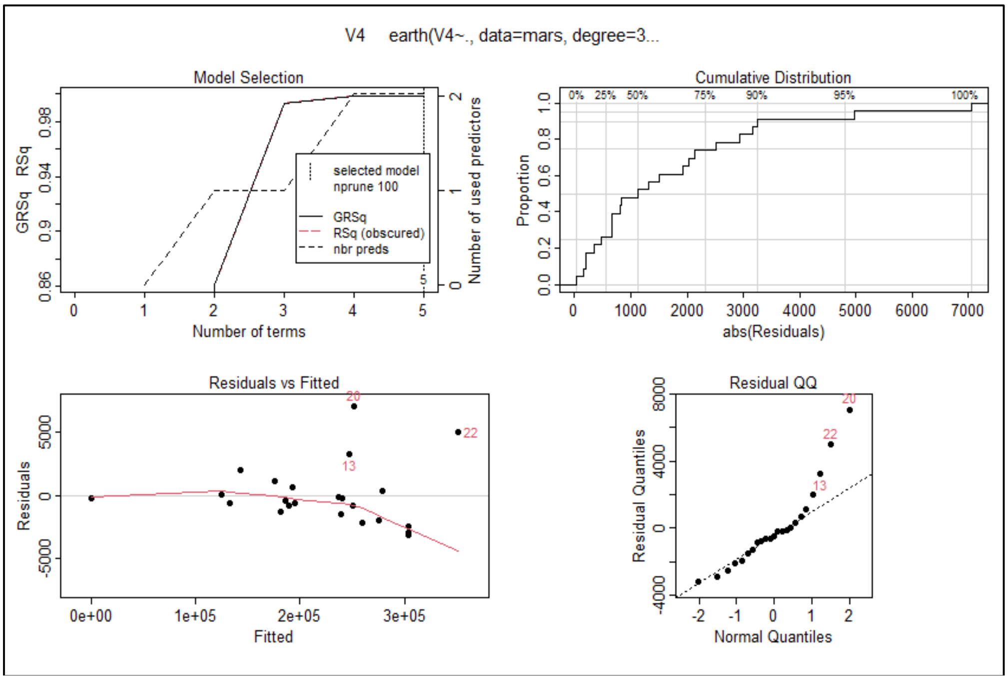


**Fig 4:** Model 4 plot

**Model 5**

This model is created by taking V6 variable as dependent variable and its summary is presented below,

Coefficients (Intercept) 24232.7485

- h(238176-V4) -0.1019
- h(V4-238176) 0.1011
- h(10.25-V5) 1.0327
- h(V5-10.25) 2911.2661

Selected 5 of 5 terms, and 2 of 28 predictors (nprune=100)
Termination condition: Reached maximum RSq 0.9990 at 5 terms Importance: V4, V5, year2001-02-unused, year2002-03-unused, year2003-04-unused, year2004-05-unused,
Number of terms at each degree of interaction: 1 4 (additive model) GCV 58790.53 RSS 1352182 GRSq 0.9990435 RSq 0.9990435 CVRSq 0.7673597

Hence the prediction equation in terms of two basic functions for the mars model can be written as,

$$V6 = 24232.7485 - 0.1019 * h(238176 - V4) + 0.1011 * h(V4 - 238176) + 1.0327 * h(10.25 - V5) + 2911.2661 * h(V5 - 10.25)$$

It shows that 5 out of 5 terms were used from the original 2 out of 28 predictors. It can be seen that a variable V4 is included with a knot at 238176, the coefficient for h(238176-V4) is -0.1019, and variable V5 is included with a knot 10.25, the coefficient for h(V5-10.25) is 2911.2661. It is also clear that total 99.90% variability is explained by the model. Various plots of mars model 5 is presented below.

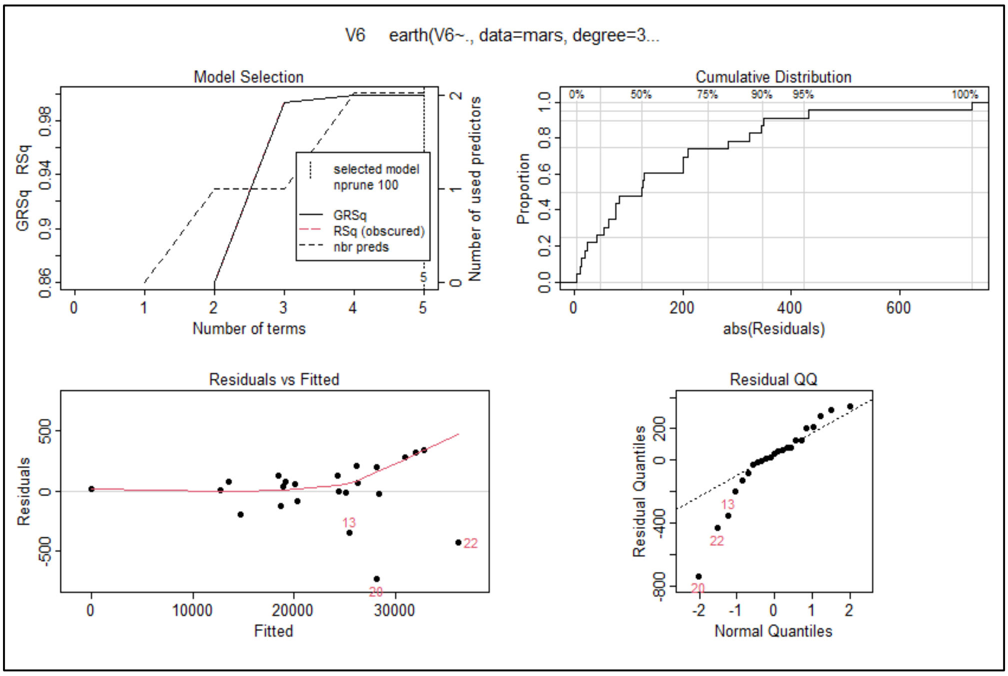


**Fig 5:** Model 5 plot

**Conclusion**

In this paper, MARS predictive models with the first and second degree interaction effects were developed using the MARS algorithm to estimate the productions of sugarcane, productivity, cane crushed, sugar and sugar recovery. Five different mars models were created using R-software. The established models demonstrated high explanatory power, with $R^2$ values of 0.999, 0.998, 0.996, 0.999, and 0.999, respectively, indicating that the models provide highly accurate predictions. The MARS algorithms were determined to be good predictors of the production and relationship between the other variables in agriculture. In agricultural sciences, student t test, one-way ANOVA, two-way ANOVA, multiple linear regression analysis has been widely used, but it can be suggested that the use of MARS models will be beneficial in future studies in agriculture.

**References**


8. Aboutalebian MA, Malmir M. Soybean yield and yield components affected by the mycorrhiza and bradyrhizobium at different rates of starter nitrogen fertilizer. Semina: Ciênc Agrár. 2017;38(4):2409-2418.
9. Agaoglu YS, Eyduran SP, Eyduran E. Comparison of some pomological characteristics of blackberry cultivars growth in Ayaş conditions. Ankara Univ Tarim Bilim Derg. 2007;13(1):69-74.
10. Akin M, Eyduran E, Reed BM. Using the CHAID data mining algorithm for tissue culture medium optimization. *In Vitro* Cell Dev Biol Anim. 2016;52:233.
11. Akin M, Eyduran SP, Ercisli S, Kapchina-Toteva V, Eyduran E. Phytochemical profile of wild blackberries, black and white mulberries from Southern Bulgaria. Biotechnol Biotechnol Equip. 2016;30(5):899-906.
12. Akin M, Eyduran E, Reed BM. Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut. Plant Cell Tissue Organ Cult. 2017;128:303-316.
13. Akin M, Eyduran E, Reed BM. Developing of optimal tissue culture medium for hazelnuts. IX Int Congr Hazelnut; 2017 Aug 15-19; Atakum, Samsun, Turkey.
14. Akin M, Eyduran E, Niedz RP, Reed BM. Developing hazelnut tissue culture medium free of ion confounding. Plant Cell Tissue Organ Cult. 2017;130:483-494. doi:10.1007/s11240-017-1238-z
15. Akin M, Hand C, Eyduran E, Reed BM. Predicting minor nutrient requirements of hazelnut shoot cultures using regression trees. Plant Cell Tissue Organ Cult. 2018;132:545-59. doi:10.1007/s11240-017-1353-x
16. Aksoy A, Ertürk YE, Eyduran E, Tariq MM. Comparing predictive performances of MARS and CHAID algorithms for defining factors affecting final fattening live weight in cultural beef cattle enterprises. Pak J Zool. 2018;50(6):2279-2286.
17. Arslan M, Arıoğlu H. Determining of soybean (*Glycine max* (L.) Merr.) cultivars and ideal plant type as a double crop for Amik Plain. J Agric Fac Cukurova Univ. 2003;18(3):39-46.
18. Atila SP, Agaoglu YS, Celik M. A research on the adaptation of some raspberry cultivars in Ayas (Ankara) conditions. Pak J Biol Sci. 2006;9(8):1504-1508.
19. Atila SP, Agaoglu YS, Celik M. A research on the adaptation of some blackberry cultivars in Ayas (Ankara) conditions. Pak J Biol Sci. 2006;9(9):1791-1794.
20. Aytekin I, Eyduran E, Koksal K, Akşahan R, Keskin I. Prediction of fattening final live weight from some body measurements and fattening period in young bulls of crossbred and exotic breeds using MARS data mining algorithm. Pak J Zool. 2018;50(1):189-195.
21. Bingöl Municipality. Bingöl Belediyesi stratejik planı (2015-2019). Bingöl Belediyesi; 2015.
22. Board JE. A regression model to predict soybean cultivar yield performance at late planting dates. Agron J. 2002;94:483-492.
23. Celik S, Eyduran E, Tatliyer A, Karadas K, Kara MK, Waheed A. Comparing predictive performances of some nonlinear functions and multivariate adaptive regression splines (MARS) for describing the growth of Daera Din Panah (DDP) goat in Pakistan. Pak J Zool. 2018;50(3):1187-1190.
24. Celik S, Eyduran E, Karadas K, Tariq MM. Comparison of predictive performance of data mining algorithms in predicting body weight in Mengali rams of Pakistan. Braz J Anim Sci. 2017;46(11):863-872.
25. Celik S, Akın M, Aliyev P, Eyduran SP, Eyduran E. A hybrid approach of combining factor analysis scores with MARS predictive model for regression problems. 1st Int Erciyes Agric Anim Food Sci Conf (AGANFOS); 2019 Apr 24-27; Kayseri, Turkey.
26. Chavan BH, Dahat DV, Rajput HJ, Deshmukh MP, Diwane SL. Correlation and path analysis in soybean. Int Res J Multidiscip. 2016;2(9):1-5.
27. Çalışkan S, Arslan M, Üremiş I, Çalışkan ME. The effects of row spacing on yield and yield components of full season and double-cropped soybean. Turk J Agric For. 2007;31:147-154.
28. Çevre ve Şehircilik İl Müdürlüğü. Bingöl İl çevre durum raporu. Bingöl Valiliği Çevre ve Şehircilik İl Müdürlüğü; 2011.
29. Çopur O, Gur MA, Demirel U, Karakus M. Performance of some soybean [*Glycine max* (L.) Merr.] genotypes double cropped in semi-arid conditions. Not Bot Horti Agrobo. 2009;37(2):85-91.
30. Daraz G, Hameed M, Ahmad F, Waheedullah. The response of different soybean varieties yield and yield components to different reduced irrigation levels in District Swat of Pakistan. J Biol Agric Healthcare. 2014;4(6):6-10.
31. El-Mohsen AAA, Mahmoud GO, Safina SA. Agronomical evaluation of six soybean cultivars using correlation and regression analysis under different irrigation regime conditions. J Plant Breed Crop Sci. 2013;5(5):91-102.
32. Eyduran E. The possibility of using data mining algorithms in prediction of live body weights of small ruminants. Can J Appl Sci. 2016;1:18-21.
33. Eyduran E, Akin M, Eyduran SP. Application of multivariate adaptive regression splines in agricultural sciences through R software. Ankara: Nobel Bilimsel Eserler; 2019. ISBN:978-605-2149-81-82.
34. Eyduran E, Akkus O, Kara MK, Tirink C, Tariq MM. Use of multivariate adaptive regression splines (MARS) in predicting body weight from body measurements in Mengali rams. Int Conf Agric For Food Sci Technol (ICAFOF); 2017 May 15-17; Cappadocia, Turkey.
35. Eyduran SP, Agaoglu YS. Some pomological and plant characteristics of currant varieties cultivated in Ankara (Ayaş) condition. Ankara Univ Tarim Bilim Derg. 2007;13(3):293-298.
36. Eyduran SP, Agaoglu YS, Eyduran E, Özdemir T. Comparison of some raspberry cultivars' herbal features by repeated random complete design statistic technique. Pak J Biol Sci. 2007;10(8):1270-1275.

37. Eyduran SP, Özdemir T, Agaoglu YS. Ankara (Ayaş) koşullarında yetiştirilen böğürtlen çeşitlerinin bazı bitkisel özellikleri. Alatarım. 2007;6(1):18-25.
38. Eyduran SP, Eyduran E, Agaoglu YS. Estimation of fruit weight by cane traits for various raspberries (*Rubus idaeus* L.) cultivars. Afr J Biotechnol. 2008;7(17):3044-3052.
39. Eyduran SP, Eyduran E, Ağaoğlu YS. Estimation of fruit weight by cane traits for eight American blackberries (*Rubus fructicosus* L.) cultivars. Afr J Biotechnol. 2008;7(17):3031-8.
40. Eyduran SP, Eyduran E, Khawar KM, Agaoglu YS. Adaptation of eight American blackberry (*Rubus fructicosus* L.) cultivars for Central Anatolia. Afr J Biotechnol. 2008;7(15):2600-2604.
41. Eyduran SP, Akin M, Ercisli S, Eyduran E, Maghradze D. Sugars, organic acids, and phenolic compounds of ancient grape cultivars (*Vitis vinifera* L.) from Igdir Province of Eastern Turkey. Biol Res. 2015;48(2):1-8.
42. Eyduran SP, Akin M, Ercisli S, Eyduran E. Phytochemical profiles and antioxidant activity of some grape accessions (Vitis spp.) native to Eastern Anatolia of Turkey. J Appl Bot Food Qual. 2015;88:5-9.
43. Eyduran SP, Ercisli S, Akin M. Organic acids, sugars, vitamin C, antioxidant capacity, and phenolic compounds in fruits of white (*Morus alba* L.) and black (*Morus nigra* L.) mulberry genotypes. J Appl Bot Food Qual. 2015;88:134-138.
44. Eyduran E, Sevgenler H, Akın M, Eyduran SP. Usage of multivariate adaptive regression splines for predicting continuous responses. In: Int Agric Sci Congr; 2018 May 9-12; Van, Turkey. Anim Plant Sci. 2018.
45. Food and Agriculture Organization of the United Nations (FAO). Production statistics, crops. 2017. Available from: http://www.fao.org/faostat/en/#data/QC
46. Food and Agriculture Organization of the United Nations (FAO). The role of soybean in fighting world hunger. 2015. Available from: http://www.fao.org/3/abs958e.pdf
47. Friedman JH. Multivariate adaptive regression splines. Ann Stat. 1991;19(1):1-67.
48. Gecer MK, Akin M, Gundogdu M, Eyduran SP. Organic acids, sugars, phenolic compounds, and some horticultural characteristics of black and white mulberry accessions from Eastern Anatolia. Can J Plant Sci. 2016;96(1):27-33.
49. Gozuacik C, Eyduran E, Cam H, Kara MK. Detection of infection preferences of the alfalfa seed chalcid, Bruchophagus roddi Gussakovskiy, 1933 (Hymenoptera: Eurytomidae) in alfalfa (*Medicago sativa* L.) fields of Igdir, Turkey. Legume Res. 2018;41:150-154.
50. Grzesiak W, Zaborski D. Examples of the use of data mining methods in animal breeding. In: Karahoca A, editor. Data mining applications in engineering and medicine. Rijeka: InTech; 2012. doi:10.5772/2616
51. Gür MA, Çopur O, Karakuş M, Demirel U. Determination of yield and yield components for some soybean (*Glycine max* L. Merr.) genotypes under the Harran Plain conditions. GAP IV Tarım Kongresi; 2004 Sep 21-23; Şanlıurfa, Turkey.
52. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference and prediction. New York: Springer-Verlag; 2001.
53. Hu S. Akaike information criterion. Center for Research in Scientific Computation, North Carolina State University; 2007. Available from: http://www4.ncsu.edu/~shu3/Presentation/AIC.pdf
54. Karadas K, Tariq M, Tariq MM, Eyduran E. Measuring predictive performance of data mining and artificial neural network algorithms for predicting lactation milk yield in indigenous Akkaraman sheep. Pak J Zool. 2017;49(1):1-7.
55. Karasu A, Öz M, Göksoy AT. A study on the adaptation of some soybean (*Glycine max* (L.) Merr.) cultivars under Bursa conditions. Uludag Univ J Fac Agric. 2002;16(2):25-34.
56. Kleinschmidt A. Soybean seeding rate and row width effect on yield. 2009. Available from: http://agvanwert.wordpress.com
57. Kornacki J, Ćwik J. Statistical learning systems. Warsaw: WNT; 2005. Polish.
58. Milborrow S. Derived from mda:mars by T. Hastie and R. Tibshirani. earth: Multivariate adaptive regression splines. R package; 2011.
59. Milborrow S. Notes on the earth package. 2018. Available from: http://www.milbo.org/doc/earth-notes.pdf
60. Naidu CR, Reddy GK, Sumathi V, Reddy PVM. Response of soybean varieties to different sowing times. J Pharmacogn Phytochem. 2017;6(5):1092-1095.
61. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014. Available from: http://www.R-project.org
62. Sall K, Sinclair TR. Soybean genotypic differences in sensitivity of symbiotic nitrogen fixation to soil dehydration. Plant Soil. 1991;133:31-37.
63. Sephton P. Forecasting recessions: can we do better on MARS™? Fed Reserve Bank St Louis Rev. 2001;Mar/Apr:39-49.
64. Takma C, Atil H, Aksakal V. Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields. Kafkas Univ Vet Fak Derg. 2012;18:941-944.
65. Unal I, Onder M. Determination of some agricultural characteristics of the soybean (*Glycine max* L.) lines developed by hybridization method. Selçuk Univ J Agric Fac. 2008;22(45):52-57.
66. Willmott C, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res. 2005;30:79-82.
67. Worku M, Astatkie T. Effects of row spacing on productivity and nodulation of two soybean varieties under hot sub-moist tropical conditions in south-western Ethiopia. J Agric Rural Dev Trop Subtrop. 2015;116(2):99-106.
68. Yari V, Frnia A, Maleki A, Moradi M, Naseri R, Ghasemi M, *et al.* Yield and yield components of soybean cultivars as affected by planting date. Bull Environ Pharmacol Life Sci. 2013;2(7):85-90.
69. Yetkin SG, Arioglu H. Determination of yield and important plant characteristics of some soybean varieties and genotypes grown as a main crop in the Çukurova region. Çukurova Univ Fen Bilim Enst Derg. 2009;20(1):29-37.
70. Yılmaz HA, Efe L. Possibilities of growing of some soybean (*Glycine max* (L.) Merr.) cultivars as a double crop under Kahramanmaraş conditions. Turk J Agric For. 1998;22:135-142.