STM JOURNALS

Research & Reviews: Journal of
Agricultural Science and Technology

ISSN: 2278-2206 (Online)
ISSN: 2349-3682 (Print)
Volume 15, Issue 1, 2026
January–April
DOI (Journal): 10.37591/RRJoAST

https://journals.stmjournals.com/rrjoast

Research RRJoAST

A Multivariate Adaptive Regression Splines Based Study of Soil Parameters and Their Impact on Onion Yield in Bhavnagar District

Arjun Nayak[1,*], K.S. Tailor[2]

Abstract

Bhavnagar district is one of the prominent onion-growing areas in the Saurashtra region of Gujarat, encompassing key talukas such as Mahuva, Talaja, Ghogha, Jesar, and Palitana. Onion cultivation in the district is carried out across three distinct seasons: rabi, kharif, and late kharif with harvesting periods extending from April to May for the rabi crop and from October to March for the kharif and late kharif crops. The productivity of onion crops is strongly influenced by soil characteristics, particularly pH balance and nutrient availability. This study examines the effects of soil pH and NPK (Nitrogen, Phosphorus, and Potassium) levels on onion yield in the Bhavnagar district of Gujarat, a major onion-producing region. A multivariate adaptive regression splines (MARS) modeling approach was employed to capture nonlinear relationships and complex interactions between soil parameters and crop yield. The MARS model effectively identified key spline functions representing critical thresholds in pH and nutrient levels that significantly impact yield. The findings provide valuable insights for site-specific soil management and precision agriculture practices, enabling farmers to optimize fertilizer application and soil conditioning strategies. This data-driven approach supports sustainable agricultural planning, improved resource-use efficiency, and enhanced crop productivity, offering a practical decision-support framework for policymakers, agronomists, and extension services involved in onion cultivation in semi-arid regions.

Keyword: MARS model, onion crop, regression, statistical analysis, soil parameters

INTRODUCTION

Onion is an important horticultural crop in Gujarat, valued for its culinary, economic and export significance. In Bhavnagar district, where agriculture is a dominant livelihood, onion cultivation contributes substantially to farmers' income and to regional vegetable supply chains. Soil factors play a central role in determining onion growth, bulb quality, and final yield: soil texture, structure, nutrient status (especially N, P, K), pH, salinity (electrical conductivity), organic carbon, water-holding capacity and bulk density all interact with climatic variables to shape crop performance. Soil texture and structure influence root development and moisture availability for onion, a crop with a relatively shallow rooting system. Sandy loam or loamy soils with good drainage generally favour bulb formation by preventing waterlogging and enabling aeration. Conversely, compacted or heavy clay soils can restrict root penetration, reduce oxygen in the root zone, and increase disease pressure, negatively affecting bulb size and marketability. Soil water retention and infiltration behaviour interact with irrigation scheduling; therefore, measuring field capacity, permanent wilting point and available water-holding capacity is important when linking

***Author for Correspondence**
Arjun Nayak
E-mail: nayakarjun720@gmail.com

[1]Research scholar, Department of Statistics, M. K. Bhavnagar University, Bhavnagar, Gujarat, India.
[2]Assistant professor, Department of Statistics, M. K. Bhavnagar University, Bhavnagar, Gujarat, India.

Received Date: January 06, 2026
Accepted Date: January 24, 2026
Published Date: February 08, 2026

Citation: Arjun Nayak, K.S. Tailor. A Multivariate Adaptive Regression Splines Based Study of Soil Parameters and Their Impact on Onion Yield in Bhavnagar District. Research & Reviews Journal of Agricultural Science and Technology. 2026; 15(1): 53–64p.

soil to yield outcomes in Bhavnagar's cropping systems. Chemical soil properties particularly pH, electrical conductivity (EC), organic carbon (OC), and available N, P and K are direct determinants of nutrient availability to the onion crop. Onion prefers near-neutral to slightly acidic soils for optimal micronutrient uptake, and extremes of pH can induce deficiencies or toxicities. High soil salinity (elevated EC) can be especially harmful to onions, reducing bulb size and marketable yield; thus, assessing salinity and sodium hazard (e.g., exchangeable sodium percentage or sodium adsorption ratio where relevant) is recommended for coastal or irrigated areas. Organic carbon influences nutrient cycling and soil physical quality; low OC frequently correlates with reduced yield stability and poorer response to fertilizer inputs.

Several earlier studies have examined issues closely aligned with the objectives of the present research.

Singh *et al.* (2021) [1] have conducted a field experiment to study the effect of different levels of nitrogen, phosphorus, and potassium on the growth and bulb yield of onion. Pusa Madhavi under Lucknow conditions. He was using randomized block design with twelve treatment combinations. The study concluded that balanced and combined application of nitrogen, phosphorus, and potassium was more effective than individual nutrient application, and NPK at 100:50:80 kg ha^{-1} was the most suitable dose for achieving higher growth and bulb yield of onion.

Kabir (2020) [2] have conducted a field experiment at the SAU Farm of Sher-e-Bangla Agricultural University, Dhaka, during October 2019 to March 2020 to evaluate the effect of different sources of nitrogen on the growth and yield of tomato (Solanum lycopersicum L.) under field conditions. The study concluded that application of 100% nitrogen through USG at 100 kg ha^{-1} was the most effective in improving growth and yield attributes of tomato under the agro-climatic conditions of Sher-e-Bangla Agricultural University.

Garai *et al*. (2024) [3] the authors developed a hybrid model that integrates Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and machine learning algorithms. The authors concluded that optimization-based hybrid models, which combine decomposition, machine learning, and stochastic techniques, present a promising avenue for handling complex agricultural and economic time series forecasting.

Mirani *et al.* (2021) [4] reviewed the role of machine learning in advancing agricultural production within the framework of Agriculture 4.0. The authors emphasized that while monitoring and managing key agricultural factors was previously challenging, computational intelligence and machine learning techniques now enable accurate analysis, quantification, monitoring, and prediction of crop performance. They highlighted the robustness of machine learning in handling large datasets and making reliable predictions, particularly in crop yield forecasting.

DATA BASE

The present study was carried out in Bhavnagar district of Gujarat, focusing on the talukas of Mahuva, Talaja, Ghogha, and Jesar, which represent the major onion-growing zones of the district. Among these, Mahuva is recognized as a leading centre of onion production, largely attributed to its favourable soil characteristics, including appropriate soil reaction (pH), and relatively balanced availability of essential nutrients such as nitrogen, phosphorus, and potassium, which support bulb development and yield formation. The study was based on secondary data collected from Statistical Branch of the District Panchayat Office, Bhavnagar, Krishi Vigyan Kendra, Sanosara and Soil and Science Department, Junagadh Agricultural University. The period for this study is 2018 to 2024. This seven-year time span was considered adequate to reflect recent variations and trends in soil parameters influencing onion productivity.

METHODOLOGY

Multivariate adaptive regression splines (MARS) is a data mining technique that can be used for solving regression-type problems (Hastie *et al*., 2001) [5].

As a widening of classification and regression tree (CART) algorithm, MARS is an effective machine learning algorithm that defines the relation between a dependent variable and a set of independent variables (Celik and Boydak, 2020) [6].

It is a non-parametric procedure, for invention adaptive regressions that uses piece wise 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) [7].

In every spline, MARS splits the data an anymore inside many sub groups. 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 sub group 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) [7, 8].

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

$$(x - t)+= \begin{cases} x > t \\ 0, \text{otherwis} \end{cases}$$

These serve as basic 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:

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

Where the aggregate is over the M terms in the model and β_0 is an intercept, β_m is a coefficient of basic 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) [5]. Function H is defined as,

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

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 pair wise interactive (Friedman, 1991).

During forward step, a number of basic 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) [5].

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) [9].

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

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).

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) [10, 11, 12, 13]:

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

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

2. Adjusted Coefficient of determination,

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

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

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

4. Standard deviation ratio $SDratio$

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

5. Akaike Information Criterion (AIC)

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

6. Corrected Akaike Information Criterion (AICc)

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

Where k is the number of selected terms and n is the sample size. (Hu,2007). 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, $\bar{Y}$ is the average of the dependent variable of the variable, ε_i is the residual value of i^{th} variable, $\bar{\epsilon}$ 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 (Zabihi, 2016; Milborrow, 2018; R Core Team, 2014; Eyduran et al., 2019) [14, 15, 16, 13].

RESULTS AND DISCUSSION

The dataset considered in this study consists of seven major variables related to onion cultivation. Variable V1 represents onion yield, expressed in tonnes per hectare, while V2 denotes total production measured in thousand tonnes. The cultivated area under onion crop, recorded in hectares, is captured through variable V3. Soil characteristics are represented by V4 and V5, where V4 corresponds to soil pH and V5, V6 and V7 represents the combined status of major soil nutrients, namely nitrogen, phosphorus, and potassium (NPK). The Descriptive Statistics of all selected variables are summarized in Table 1.

Table 1 presents the descriptive statistics of yield, production, area, and key soil parameters influencing onion cultivation in the study area. Onion yield (V1) showed limited variability, indicating relatively stable productivity across observations, whereas production (V2) and cultivated area (V3) exhibited wide ranges, reflecting substantial spatial and temporal variation. Soil pH (V4) remained within a narrow range, suggesting generally favourable soil reaction for onion growth. In contrast, nitrogen (V5), phosphorus (V6), and potassium (V7) levels displayed moderate to high variability, highlighting differences in soil fertility status across fields. The statistics indicate that while yield stability is maintained, variations in nutrient availability may play a significant role in influencing onion productivity in the region.

In the present study, two separate Multivariate Adaptive Regression Splines (MARS) models were formulated to estimate the major response variables, namely onion yield, total production. Model adequacy was evaluated using multiple goodness-of-fit indicators, including the correlation coefficient (r), coefficient of determination (R^2), adjusted R^2, root mean square error (RMSE), SD ratio, Akaike Information Criterion (AIC and AICc), and generalized cross-validation (GCV). The results summarized in Table 2 indicate that all developed models achieved excellent predictive performance. Consistent with the guidelines suggested by Grzesiak and Zaborski (2012) and Eyduran et al. (2019), SD ratio values below 0.40 reflect a well-fitted model. Accordingly, the low SD ratios obtained confirm the robustness of the MARS models, which collectively accounted for nearly 99.93% of the variability in onion production.

MARS MODELS

In this investigation, two separate Multivariate Adaptive Regression Splines (MARS) models were formulated and systematically evaluated to examine the associations among the principal variables influencing onion crop yield. Each model was designed to capture nonlinear relationships and interaction effects between the dependent and explanatory variables, thereby enabling a comprehensive assessment of their individual impacts. The structural framework of these models is described as follows:

Table 1. Descriptive statistics.

	N	Minimum	Maximum	Mean	Std. Deviation
V1	70	23.53	26.05	25.1248	0.53954
V2	70	9289.22	248159.34	117733.8021	72604.31783
V3	70	515.00	114110.00	8016.2429	13350.43030
V4	70	6.08	7.70	6.8651	0.41591
V5	70	75.80	158.70	119.5657	24.17759
V6	70	28.00	77.60	50.5571	11.97565
V7	70	127.70	291.50	202.1614	45.00676

Table 2. Goodness of fit criteria for MARS algorithm.

Variables	r	R^2	Adj. R^2	RMSE	SD_{ratio}	AIC	AIC_C	GCV
V1	0.9986305	0.997	0.988	0.028	0.052	-396	-72	0.0007854
V2	0.9993893	0.999	0.992	2518.896	0.035	1216	2030	6344838

Model 1. Dependent variable.
V1 = yield, independent variable:
V2= production
V3 = area
V4=pH levels
V5=Nitrogen levels
V6=Phosphorus levels,
V7=Potassium levels

Model 2. Dependent variable.
V2 = production, independent variable:
V3= area,
V4=pH levels,
V5=Nitrogen levels,
V6=Phosphorus levels,
V7=Potassium levels.

Model 1

In Model 1, onion yield (V1) was specified as the response variable. This model, similar to the other MARS models developed in the present investigation, was implemented using the R statistical environment through the *earth* package, which is widely employed for Multivariate Adaptive Regression Splines analysis. The *earth* package facilitates the identification of nonlinear relationships and interaction effects among predictor variables. A comprehensive description and detailed results of MARS Model 1 (V1 ~.) are presented in the subsequent section.

Coefficients (Intercept)	27.8841
h(93032.4-V2)	0.0001
h(V2-93032.4)	0.0003
h(1707-V3)	0.0007
h(V3-1707)	0.0006
h(V3-2942)	-0.0005
h(7.31-V4)	-1.0382
h(V4-7.31)	-2.7370
h(V5-100.2)	-0.0859
h(V5-105.3)	-0.0187
h(V5-116.2)	0.7245
h(V5-124.3)	-0.2014
h(127-V5)	-0.0665
h(V5-127)	0.1220
h(61.7-V6)	-0.0458
h(V6-61.7)	0.1715
h(V7-167.9)	-0.0173
h(216.7-V7)	-0.0103
h(V7-216.7)	-0.0539
h(V7-234.8)	-0.0871
h(V7-250.6)	0.0314
V2 * h(V5-127)	0.0000
V2 * h(V7-216.7)	0.0000
V4 * h(V5-116.2)	-0.0766
V5 * h(V6-61.7)	-0.0021
V5 * h(V7-216.7)	0.0007
h(V5-116.2) * V7	0.0002
h(107928-V2) * h(V7-167.9)	0.0002

h(V2-107928) * h(V7-167.9)	0.0004
h(V3-1707) * h(V4-7.08)	0.0002
h(V3-1707) * h(7.08-V4)	-0.0001
h(V3-1707) * h(V5-113.1)	0.0000
h(V3-1707) * h(113.1-V5)	0.0000
h(4924-V3) * h(V5-105.3)	0.0000
h(V3-4924) * h(V5-105.3)	0.0000
h(5255-V3) * h(61.7-V6)	0.0000
h(V3-5255) * h(61.7-V6)	0.0000
h(V3-1707) * h(V7-188.5)	0.0000
h(V3-1707) * h(188.5-V7)	0.0000
h(V3-1707) * h(V7-207.1)	0.0000
h(7.31-V4) * h(V5-113.1)	0.0771
h(7.31-V4) * h(113.1-V5)	0.0560
h(7.31-V4) * h(V5-120.4)	-0.1973
h(6.73-V4) * h(61.7-V6)	-0.0390
h(V4-6.73) * h(61.7-V6)	-0.0252
h(122.6-V5) * h(61.7-V6)	0.0020
h(V5-122.6) * h(61.7-V6)	-0.0039
h(118.1-V5) * h(V7-167.9)	-0.0022
h(V5-118.1) * h(V7-167.9)	-0.0006
V3 * h(7.31-V4) * h(V5-113.1)	0.0000
V4 * h(V5-122.6) * h(61.7-V6)	0.0008
V4 * h(118.1-V5) * h(V7-167.9)	0.0004

Selected 52 of 52 terms, and 6 of 6 predictors. Termination condition: RSq changed by less than 0.001 at 52 terms.

Importance
V2, V3, V4, V5, V6, V7. Number of terms at each degree of interaction: 1 20 28 3
GCV 0.0007854 RSS 0.05498 GRSq 0.9973 RSq 0.9973 CVRSq -1.014.

Therefore, the predictive relationship for the MARS model, represented through fifty-two fundamental basis functions, is given as follows:

V1=27.8841 + 0.0001h(93032.4-V2) +0.0003h(V2-93032.4) + 0.0007h(1707-V3) + 0.0006 h(V3-1707) - 0.0005h(V3-2942) -1.0382 h(7.31-V4) - 2.7370 h(V4-7.31) - 0.0859h(V5-100.2) -0.0187h(V5-105.3) + 0.7245h(V5-116.2) - 0.2014h(V5-124.3) - 0.0665h(127-V5) + 0.1220

h(V5-127) - 0.0458h(61.7-V6) + 0.1715h(V6-61.7) - 0.0173h(V7-167.9) - 0.0103h(216.7-V7) - 0.0539h(V7-216.7) - 0.0871h(V7-234.8) + 0.0314h(V7-250.6) + 0.0000V2 * h(V5-127) + 0.0000V2 * h(V7-216.7) - 0.0766V4 * h(V5-116.2) - 0.0021V5 * h(V6-61.7) + 0.0007V5 * h(V7-216.7) + 0.0002h(V5-116.2) * V7 + 0.0002h(107928-V2) * h(V7-167.9) + 0.0004h(V2-107928) * h(V7-167.9) + 0.0002h(V3-1707) * h(V4-7.08) - 0.0001h(V3-1707) * h(7.08-V4) + 0.0000h(V3-1707) * h(V5-113.1) + 0.0000h(V3-1707) * h(113.1-V5) + 0.0000h(4924-V3) * h(V5-105.3) + 0.0000h(V3-4924) * h(V5-105.3) + 0.0000h(5255-V3) * h(61.7-V6) + 0.0000

h(V3-5255) * h(61.7-V6) + 0.0000h(V3-1707) * h(V7-188.5) + 0.0000h(V3-1707) * h(188.5-V7) + 0.0000h(V3-1707) * h(V7-207.1) + 0.0771h(7.31-V4) * h(V5-113.1) + 0.0560h(7.31-V4) * h(113.1-V5) - 0.1973h(7.31-V4) * h(V5-120.4) - 0.0390h(6.73-V4) * h(61.7-V6) -0.0252h(V4-6.73) * h(61.7-V6) + 0.0020h(122.6-V5) * h(61.7-V6) - 0.0039h(V5-122.6) * h(61.7-V6) -0.0022h(118.1-V5) * h(V7-167.9) - 0.0006 h(V5-118.1) * h(V7-167.9) + 0.0000 V3 * h(7.31-V4) * h(V5-113.1) + 0.0008V4 * h(V5-122.6) * h(61.7-V6) + 0.0004 V4 * h(118.1-V5) * h(V7-167.9).

It shows that 52 out of 52 terms were used from the original predictor. It can be seen that a variable V2, V3, V4, V5, V6 and V7 are included with a knot at 93032.4, 1707, 7.31, 127, 61.7 and 216.7 respectively, the coefficient for h(93032.4-V2) is 0.0001, the coefficient for h(1707-V3) is 0.0007, the coefficient for h(7.31-V4) is -1.0382, the coefficient for h(127-V5) is -0.0665, the coefficient for h(61.7-V6) is -0.0458 and the coefficient for h(216.7-V7) is -0.0103. It is also clear that total 99.86 % variability is explained by the model 1.

The various graphical outputs of MARS model 1 are shown below.

The diagnostic plots show that the Generalized Cross-Validation (GCV/GRSq) increases rapidly and then stabilizes, indicating that the optimal MARS model is achieved with about eight terms, where the mean out-of-fold is maximized. The cumulative distribution of absolute residuals indicates good prediction accuracy, with around 75–90% of residuals falling below 0.03–0.05. The residuals versus fitted plot shows residuals cantered around zero with no strong pattern, suggesting that homoscedasticity is largely satisfied, although slight curvature and a few influential observations are present. The residual QQ plot further indicates that residuals are approximately normally distributed, with only mild deviations in the tails due to a few outliers. The selected MARS model demonstrates good generalization ability, small prediction errors, and reliable performance, with minor attention needed for a few influential observations**.**

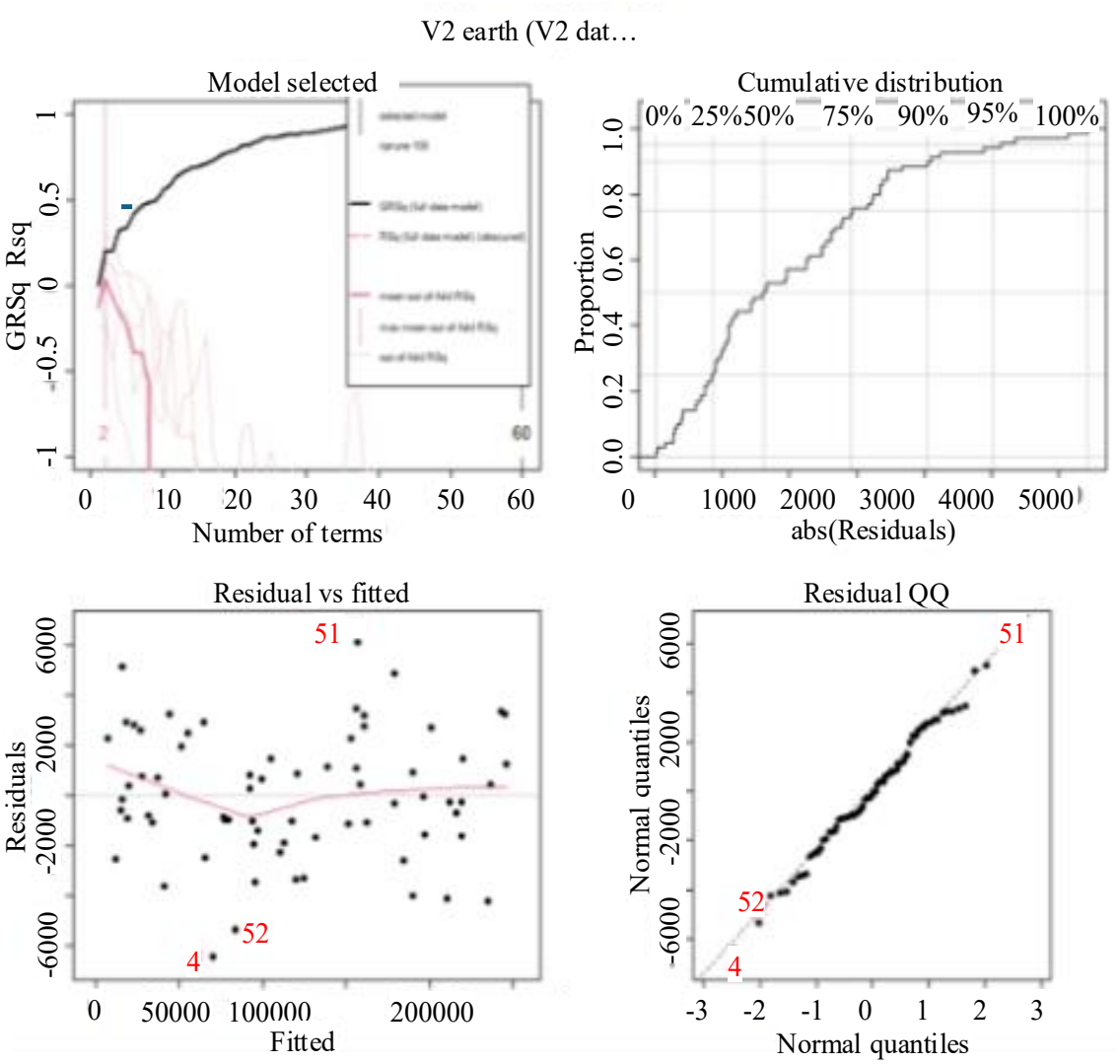

Figure 1. Model 1 plot.

	coefficients		
(Intercept)	266209	h(V1-24.4908) * h(46.5-V6)	-53222
h(24.4908-V1)	-100659	h(V1-24.4908) * h(V6-54.6)	-32269
h(V1-24.4908)	-254137	h(V1-24.4908) * h(V6-49.8)	19581
h(V3-4026)	-94	h(V1-24.4908) * h(V7-209.1)	-9702
h(4604-V3)	-19	h(V1-24.4908) * h(209.1-V7)	13983
h(V3-4604)	-282	h(V1-24.4908) * h(V7-191.4)	4508
h(V3-5255)	-297	h(25.1752-V1) * h(234.8-V7)	5857
h(V3-8596)	117	h(V1-25.1752) * h(234.8-V7)	-8959
h(6.79-V4)	138813	h(25.2053-V1) * h(V7-147.1)	-2621
h(V4-6.79)	73530	h(V1-25.2053) * h(V7-147.1)	4277
h(142.7-V5)	3461	h(6034-V3) * h(142.7-V5)	0
h(V5-142.7)	5377	h(V3-6034) * h(142.7-V5)	0
h(43.7-V6)	-1926883	h(V3-4026) * h(V6-48.3)	-8
h(V6-43.7)	-18530	h(V3-4026) * h(48.3-V6)	9
h(V6-55.2)	44456	h(6248-V3) * h(V6-43.7)	-156
h(V7-147.1)	2741	h(V3-6248) * h(V6-43.7)	18
h(V7-175.1)	-4397	h(6.83-V4) * h(142.7-V5)	-10369
h(234.8-V7)	-6735	h(V4-6.83) * h(142.7-V5)	42205
h(V7-234.8)	-11181	h(6.8-V4) * h(V7-147.1)	2528
V1 * h(43.7-V6)	79006	h(V4-6.8) * h(V7-147.1)	2295
h(V3-4604) * V6	14	h(142.7-V5) * h(V7-180.6)	20
h(V3-8596) * V6	-1	h(142.7-V5) * h(180.6-V7)	50
V4 * h(V7-234.8)	3536	h(V6-43.7) * h(V7-178.7)	-224
h(V1-24.4908) * h(V3-7132)	-30	h(V6-43.7) * h(178.7-V7)	-66
h(V1-24.4908) * h(7132-V3)	-8	V3 * V4 * h(V7-234.8)	0
h(25.2053-V1) * h(V3-4026)	-32	V1 * h(6248-V3) * h(V6-43.7)	6
h(V1-25.2053) * h(V3-4026)	22	h(V1-24.4908) * V3 * h(V6-46.5)	3
h(V1-24.4908) * h(V4-6.74)	-146221	h(V3-6248) * V4 * h(V6-43.7)	-5
h(V1-24.4908) * h(6.74-V4)	-425519	h(V3-6248) * V5 * h(V6-43.7)	0
h(V1-24.4908) * h(V6-46.5)	-24862	h(V4-6.83) * h(142.7-V5) * V6	-949

Model 2

In Model 2, onion yield (V2) was specified as the response variable. This model, similar to the other MARS models developed in the present investigation, was implemented using the R statistical environment through the *earth* package, which is widely employed for Multivariate Adaptive Regression Splines analysis. The *earth* package facilitates the identification of nonlinear relationships and interaction effects among predictor variables. A comprehensive description and detailed results of MARS Model 2 (V2 ~.) are presented in the subsequent section.

Selected 60 of 60 terms, and 6 of 6 predictors. Termination condition: RSq changed by less than 0.001 at 60 terms.

Importance

V1, V3, V4, V5, V6, V7. Number of terms at each degree of interaction: 1 18 35 6
GCV 6344838 RSS 444138651 GRSq 0.9988 RSq 0.9988 CVRSq -54.5

Therefore, the predictive relationship for the MARS model, represented through sixty fundamental basis functions, is given as follows:

V2=266209 - 100659h(24.4908-V1) - 254137h(V1-24.4908) - 94h(V3-4026) - 19h(4604-V3) - 282h(V3-4604) - 297h(V3-5255) + 117h(V3-8596) + 138813 h(6.79-V4) + 73530 h(V4-6.79) + 3461h(142.7-V5) + 5377h(V5-142.7) - 1926883h(43.7-V6) – 18530h(V6-43.7) + 44456h(V6-55.2) + 2741h(V7-147.1) - 4397h(V7-175.1) - 6735h(234.8-V7) - 11181h(V7-234.8) + 79006 V1 * h(43.7-V6) + 14h(V3-4604) * V6 - 1h(V3-8596) * V6 + 3536 V4 * h(V7-234.8) - 30h(V1-24.4908) * h(V3-7132) - 8h(V1-24.4908) * h(7132-V3) - 32h(25.2053-V1) * h(V3-4026) + 22 h(V1-25.2053) * h(V3-4026) - 146221h(V1-24.4908) * h(V4-6.74) - 425519h(V1-24.4908) * h(6.74-V4) - 24862h(V1-24.4908) * h(V6-46.5) -53222h(V1-24.4908) * h(46.5-V6) - 32269 h(V1-24.4908) * h(V6-54.6) + 19581h(V1-24.4908) * h(V6-49.8) - 9702h(V1-24.4908) * h(V7-209.1) + 13983h(V1-24.4908) * h(209.1-V7) + 4508h(V1-24.4908) * h(V7-191.4) + 5857 h(25.1752-V1) * h(234.8-V7) - 8959h(V1-25.1752) * h(234.8-V7) - 2621h(25.2053-V1) * h(V7-147.1) + 4277h(V1-25.2053) * h(V7-147.1) + 0h(6034-V3)

* h(142.7-V5) + 0h(V3-6034) * h(142.7-V5) - 8h(V3-4026) * h(V6-48.3) + 9h(V3-4026) * h(48.3-V6) - 156h(6248-V3) * h(V6-43.7) + 18h(V3-6248) * h(V6-43.7) - 10369h(6.83-V4) * h(142.7-V5) + 42205h(V4-6.83) * h(142.7-V5) + 2528h(6.8-V4) * h(V7-147.1) + 2295h(V4-6.8) * h(V7-147.1) + 20h(142.7-V5) * h(V7-180.6) + 50h(142.7-V5) * h(180.6-V7) - 224h(V6-43.7) * h(V7-178.7) - 66h(V6-43.7) * h(178.7-V7) + 0 V3 * V4 * h(V7-234.8) + 6V1 * h(6248-V3) * h(V6-43.7) + 3h(V1-24.4908) * V3 * h(V6-46.5) - 5h(V3-6248) * V4 * h(V6-43.7) + 0h(V3-6248) * V5 * h(V6-43.7) - 949 h(V4-6.83) * h(142.7-V5) * V6.

It shows that 60 out of 60 terms were used from the original predictor. It can be seen that a variable V1, V3, V4, V5, V6 and V7 are included with a knot at 24.4908, 4604, 6.79, 142.7, 43.7 and 234.8 respectively, the coefficient for h(24.4908-V1) is -100659h, the coefficient for h(4604-V3) is -19, the coefficient for h(6.79-V4) is 138813, the coefficient h(142.7-V5) is 3461, the coefficient for 1926883h(43.7-V6) is -1926883 and the coefficient for 6735h(234.8-V7) is -6735.
It is also clear that total 99.93% variability is explained by the model 2.

The diagnostic plots for the MARS model 2 show that the Generalized Cross-Validation (GCV/GRSq) increases quickly with the number of terms and then gradually levels off, indicating diminishing returns from added complexity.

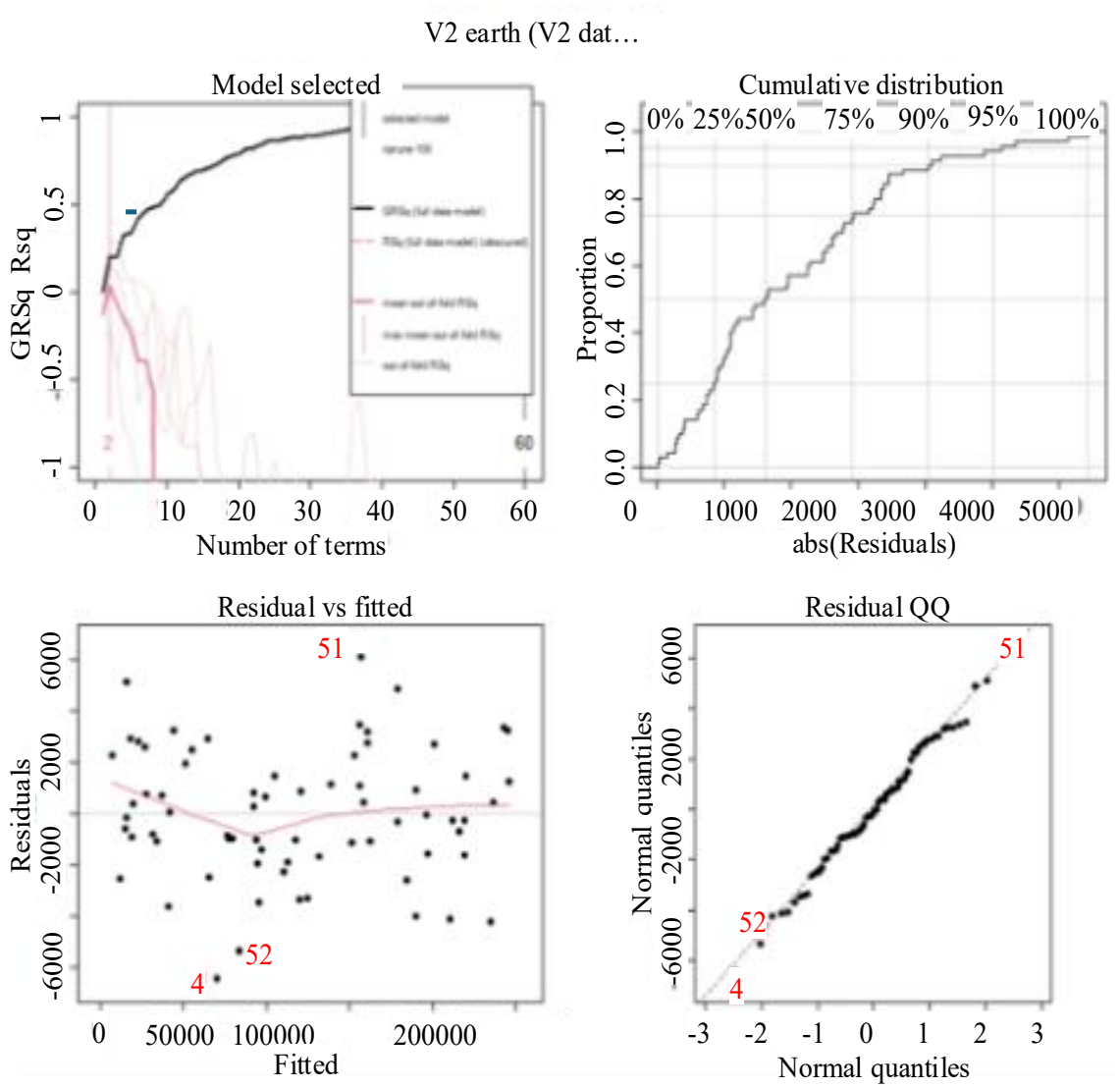

Figure 2. Model 2 plot.

The optimal model is achieved with a small number of terms, where the mean out-of-fold is maximized, while adding more terms does not improve predictive performance and may lead to overfitting. The cumulative distribution of absolute residuals indicates moderate prediction accuracy, with residuals spread over a wider range compared to V1The residuals versus fitted plot shows residuals roughly centered around zero with no strong systematic pattern, although some curvature and heteroscedasticity are visible, along with a few influential outliers. The residual QQ plot shows that residuals generally follow the normal reference line, but with noticeable deviations in the tails caused by extreme observations. the MARS model 2 provides a reasonable but weaker fit than V1, with acceptable generalization performance but larger residual variability and several influential points that should be examined further.

CONCLUSION

The present analysis clearly demonstrates that onion yield in the Bhavnagar district is strongly governed by the combined influence of soil pH and the availability of major nutrients, namely nitrogen, phosphorus, and potassium. Onion crops performed best in soils that were slightly acidic to neutral, with an optimal pH range of 6.0 to 8.0 (which was suggested by MARS model), as this condition maximizes nutrient availability and uptake efficiency. Nitrogen emerged as a key driver of vegetative growth and bulb development, with higher and stable yields observed under medium to optimum available nitrogen levels (75 to 160) kg ha^{-1}, while both deficiency and excess were found to be detrimental to bulb formation. Phosphorus played a crucial role during early root establishment and bulb initiation, and soils maintaining available phosphorus in the range of (43 to 62) kg ha^{-1} supported improved yield stability. Potassium significantly influenced bulb size, quality, and stress tolerance, with superior yields associated with medium to adequate potassium levels of about (216 to 251) kg ha^{-1}. Overall, the findings emphasize that sustainable onion productivity in Bhavnagar depends not on any single soil factor, but on maintaining a balanced soil chemical environment, where pH remains near neutral and N, P, and K are managed within their optimal ranges through soil-test-based nutrient management practices.

REFERENCES

1. Singh GP, Meena ML, Pankaj TR. Effect of different levels of nitrogen, phosphorus and potassium on growth and bulb yield of onion. The Pharma Innovation Journal. 2021;10(10):1504-7.
2. KABIR MR. *EFFECT OF DIFFERENT SOURCES OF NITROGEN ON GROWTH AND YIELD OF TOMATO (Solanum lycopersicum L.)* (Doctoral dissertation, DEPARTMENT OF SOIL SCIENCE, SHER-E-BANGLA AGRICULTURAL UNIVERSITY, SHER-E-BANGLA NAGAR, DHAKA).
3. Garai S, Paul RK, Yeasin M, Paul AK. CEEMDAN-based hybrid machine learning models for time series forecasting using MARS algorithm and PSO-optimization. Neural Processing Letters. 2024 Mar 6;56(2):92.
4. Mirani A, Memon MS, Chohan R, Wagan AA, Qabulio M. Machine learning in agriculture: A review. LUME. 2021 Jan;10:5.
5. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference and prediction. New York: Springer-Verlag; 2001.
6. Celik S, Boydak E. Description of the relationships between different plant characteristics in soybean using multivariate adaptive regression splines (MARS) algorithm. JAPS: Journal of Animal & Plant Sciences. 2020 Apr 1;30(2).
7. Friedman JH. Multivariate adaptive regression splines. The annals of statistics. 1991 Mar;19(1):1-67.
8. Sephton P. Forecasting recessions: can we do better on MARS. Federal Reserve Bank of St. Louis Review. 2001 Mar 1;83(March/April 2001).
9. Kornacki J, Cwik J. Statistical Learning Systems (in Polish) WNT. Warsaw, Poland. 2005.
10. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research. 2005 Dec 19;30(1):79-82.

11. Liddle AD, Davies AH. Pelvic congestion syndrome: chronic pelvic pain caused by ovarian and internal iliac varices. Phlebology. 2007 Jun 1;22(3):100-4.
12. Takma Ç, Atıl H, Aksakal V. Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields.
13. Eyduran E, Akin M, Eyduran SP. Application of multivariate adaptive regression splines through R software. Ankara Turkey: Nobel Academic Publishing. 2019.
14. Zabihi M, Pourghasemi HR, Pourtaghi ZS, Behzadfar M. GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. Environmental Earth Sciences. 2016 Apr;75(8):665.
15. Milborrow S. Notes on the earth package. Retrieved October. 2014 Jan 28;31:2017.
16. 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

