ORIGINAL ARTICLE

Analysis of macro nutrient related growth responses using multivariate adaptive regression splines

Meleksen Akin[1] · Sadiye Peral Eyduran[2] · Ecevit Eyduran[3] · Barbara M. Reed[4]

Abstract

Strawberry micropropagation is generally based on Murashige and Skoog mineral salts, and many cultivars grow well on this medium. However, the diverse species found in germplasm collections often do not thrive, which indicates a need to optimize the mineral nutrients. In this study, Multivariate Adaptive Regression Splines (MARS), was employed to predict shoot quality, multiplication, and leaf color responses of three strawberry species in response to the major tissue culture nutrients by generating functional associations. MARS is a non-parametric approach that can be used to deal with continuous and categorical data without requiring the strict distributional assumptions of the basic linear models. The MARS algorithm is capable of capturing non-linear patterns between the input and target variables. NH_4NO_3, $CaCl_2 \cdot 2H_2O$, $MgSO_4 \cdot 7H_2O$, KNO_3 and KH_2PO_4 were tested in a range of $0.5 \times$ to $3 \times MS$ medium, within a computer-generated optimal design that consisted of 32 treatment combinations. The plant responses were affected by all of the major salts tested and the genotype factor. Multivariate Adaptive Regression Splines captured the significant factors and their interactions to predict optimal major salts suitable for all three strawberry species: 3300 mg L^{-1} NH_4NO_3, 862.4 mg L^{-1} $CaCl_2$, 1110 mg L^{-1} $MgSO_4$, 3439 mg L^{-1} KNO_3, and 329.8 mg L^{-1} KH_2PO_4. This study identified the major nutrient needs of the three strawberry species and provides an alternative statistical technique for tissue culture data analyses.

Key Message

The MARS statistical approach was used to predict macro nutrient related growth responses of three strawberry species. The objective of the study was to make a gentle introduction to the MARS algorithm and show its potential application to tissue culture research.

Keywords *Fragaria* · In vitro culture · MARS · Mineral nutrition · Statistical analysis

---

✉ Meleksen Akin
akinmeleksen@gmail.com

1 Department of Landscape Architecture, Agricultural Faculty, Igdir University, Igdir, Turkey

2 Department of Horticulture, Agricultural Faculty, Igdir University, Igdir, Turkey

3 Department of Business Administration, Economics and Administrative Sciences Faculty, Igdir University, Igdir, Turkey

4 USDA-ARS-Retired, National Clonal Germplasm Repository, 33447 Peoria Rd, Corvallis, OR 97333, USA

Introduction

While strawberries were first grown on a variety of media in the early days of micropropagation, they are currently most commonly grown on MS medium (Murashige and Skoog 1962), or variations such as half-strength MS mineral nutrients. Simpson and Bell (1989) noted that most *Fragaria × ananassa* cultivars grew well on MS medium, but proliferation varied greatly by genotype. They concluded that strawberries do not have a uniform response to any single growth medium recipe due to the differences in response to N^6 benzyladenine (BA) concentrations. A study of *Fragaria chiloensis* cultivars from Chile found that MS was suitable for the two cultivars studied (Quiroz et al. 2017). Most diverse germplasm or breeding collections can use MS medium for the majority of accessions, but often there

are genotypes or species that do not thrive despite changes in plant growth regulators. The cultivated strawberry, *Fragaria × ananassa* Duchesne ex Rozier has a very narrow genetic background, but the 20 or more other species are more genetically diverse and include a wide range of ploidy levels (Darrow 1966).

The lack of study on the mineral nutrition of strawberry combined with the genetic diversity present in germplasm collections and breeding programs, provides incentives to improve the mineral nutrition in the medium for micropropagated strawberries. Developing a mineral nutrition model can provide the basis for an improved medium for diverse groups of plants.

The performance of a good statistical model depends on the experimental design, clear understanding of the variable structure, and application of the right statistical method. Understanding the data type is of primary importance to determine the appropriate statistical technique (Mertler and Vannatta 2002; Niedz and Evens 2016). Variables are classified as qualitative (subgrouped as nominal and ordinal), and quantitative (labeled as discrete and continuous). Nominal variables are 'names' with two or more categories without hierarchical order. Ordinal data are similar to nominal except that the categories have distinct order (level A is more severe than level B). Discrete variables are counts that include integers, and continuous variables are measurements along a continuum, which could be divided into smaller fractions (Dettori and Norvell 2018). In vitro data structure could be classified as nominal (genotype), ordinal (shoot quality rated as poor, moderate, and good), discrete (counts of shoot number: 1, 2, 3…), and continuous (measurements of shoot length). Conventional linear models such as ANOVA and regression, should only be used with continuous data that shows a linear relationship between the dependent and explanatory variables (Quinn and Keough 2002). However, plant tissue culture is a complicated biological process that involves multiple design inputs that typically interact in non-linear ways to affect the tested plant responses. Therefore, the basic linear methods are generally not suitable for analyzing in vitro data.

A number of advanced statistical techniques are promoted in tissue culture research as robust alternatives to basic modeling methods. These include Response Surface Methodology (RSM), Decision trees, and Artificial Neural Networks (ANNs) (Akin et al. 2017; Nezami-Alanagh et al. 2018; Poothong and Reed 2014). The RSM is a parametric approach that is able to model non-linear data by showing polynomial relationships between the tested variables. However, RSM can analyze only continuous variables, and therefore it generates genotype (nominal variable) specific models (which means a separate optimal growth area for each genotype) (Akin et al. 2016). Classification and Regression Tree (CART), Chi-square Automatic Interaction Detector

(CHAID), and Exhaustive Chi-square Automatic Interaction Detector (Exhaustive CHAID), are non-parametric data mining methods that are able to evaluate both continuous and categorical data, without requiring the restrictive normality and heteroscedasticity assumptions of the parametric methods (Olden et al. 2008). These techniques generate visual, easy to interpret decision trees and can include nominal variables as factors, and thus provide common optimal factor ranges for the tested genotypes (Akin et al. 2017). Artificial Neural Networks are non-parametric tools that are powerful in detecting non-linear associations, and capable of dealing with continuous and categorical data. However, ANNs do not provide a neat mathematical equation that shows the relative contribution of each independent variable in the neural model. Therefore, ANNs are considered as a "black box" (Olden et al. 2008).

Another promising statistical approach not previously utilized in tissue culture data analyses is Multivariate Adaptive Regression Splines (MARS). The MARS approach is a non-parametric regression technique that reveals complex non-linear relationships and interactions by a sequence of spline functions of the independent variables, without requiring the restrictive distributional assumptions of the general linear models. The MARS algorithm can handle both categorical and numerical variables. It shows the contribution of each predictor to the response using subsets of piecewise linear regression splines (Everingham and Sexton 2011; Friedman 1991). A spline is a flexible curve (polynomial) that is fixed at various points or knots and shows the association between the target and explanatory variable. The knots are randomly distributed within the range of each explanatory variable and detect the regions of relationship change between the input and output variables. The linear functions that form the MARS model combine additively or interactively, and the relative importance of the inputs on the target variable are illustrated clearly in a mathematical equation (Eyduran et al. 2019; Zhang and Goh 2014).

This study investigated the potential of the MARS algorithm to reveal the intrinsic non-linear and multidimensional relationships of some responses of tissue culture grown strawberries. This process aims to provide a basic introduction to the MARS algorithm and to encourage future application of this promising approach for tissue culture research, and to improve the in vitro culture of strawberries.

Materials and methods

Plant material and culture conditions

Shoots of *Fragaria bucharica* Losinsk. (Local no. CFRA 1906.001; PI 657844), *Fragaria chiloensis* (L.) Mill. subsp. *chiloensis patagonica* Staudt (CFRA 1084.002; PI

616535), and *Fragaria x ananassa* (Duchesne ex Weston) Duchesne ex Rozier 'Jucunda' (CFRA 256.002; PI 551623) strawberry species from the U.S. Department of Agriculture, National Clonal Germplasm Repository (NCGR), Corvallis, Oregon, were micropropagated on MS medium (Murashige and Skoog 1962), with LS vitamins (Linsmaier and Skoog 1965), and with the following per liter: 80 mg adenine sulfate, 170 mg NaH_2PO_4, 30 g sucrose, 4.44 ∝ M BA, 4.92 ∝ M indole-3-butyric acid (IBA), 0.03 ∝ M gibberellic acid (GA) and 6 g L^{-1} agar (A1111; PhytoTechnology Laboratories$^{\rightarrow}$, Lenexa, KS) at pH 5.7. The medium pH was adjusted using H_3PO_4 and KOH, and all medium ingredients were added before autoclaving. The plant growth regulators and vitamins were obtained from PhytoTechnology Laboratories®, and all stock solutions were prepared in house. Each Magenta™ box (Magenta™ GA7, Magenta,

Chicago, IL) contained 40 mL medium, and was autoclaved for 20 min at 121 °C (118 kPa). Growth room conditions were 80 μmol m^2s^{-1} light intensity with a 16-h photoperiod of half warm-white and half cool-white fluorescent lamps (Ecolux Starcoat, 32 watts; GE, Fairfield, CT) at 25 ± 2 °C.

Experimental design

An I-optimal experimental design (DOE) was built by Design-Expert software (Design-Expert 2010) using NH_4NO_3, $CaCl_2·2H_2O$, $MgSO_4·7H_2O$, KNO_3, and KH_2PO_4 as inputs within 32 treatment combinations. The factors were varied from 0.5× to 3×MS medium salts. Treatment 32 was the control NCGR *Fragaria* medium consisting of MS salts (1×) with adenine sulfate and NaH_2PO_4 (Table 1). All treatment

Table 1 I-optimal design consisting of MS (Murashige and Skoog 1962) major salts

Treatments	NH_4NO_3	$CaCl_2·2H_2O$	$MgSO_4·7H_2O$	KNO_3	KH_2PO_4
1	1.75	1.66	0.50	1.75	0.50
2	3.00	3.00	0.50	0.50	3.00
3	2.23	2.36	1.17	0.50	1.50
4	3.00	0.50	0.50	0.50	0.50
5	0.50	3.00	3.00	0.50	3.00
6	0.50	0.50	3.00	3.00	3.00
7	0.50	3.00	0.50	1.93	1.94
8	0.50	0.50	0.50	0.50	3.00
9	0.50	0.50	3.00	0.50	0.50
10	0.50	3.00	0.50	0.50	0.50
11	1.27	1.43	2.25	2.50	1.42
12	1.27	1.43	2.25	2.50	1.42
13	1.58	3.00	1.59	3.00	3.00
14	0.50	3.00	3.00	3.00	0.50
15	3.00	3.00	3.00	3.00	3.00
16	2.00	1.45	0.84	1.50	3.00
17	1.58	3.00	1.59	3.00	3.00
18	3.00	0.50	1.69	1.81	1.65
19	3.00	3.00	0.50	3.00	0.50
20	0.85	0.50	3.00	0.85	3.00
21	3.00	0.50	3.00	3.00	0.50
22	0.50	1.96	0.50	3.00	3.00
23	2.00	1.45	0.84	1.50	3.00
24	0.50	3.00	0.50	1.93	1.94
25	1.27	1.43	2.25	2.50	1.42
26	3.00	0.50	0.50	3.00	3.00
27	3.00	0.50	3.00	0.50	3.00
28	2.13	3.00	3.00	0.50	1.38
29	0.50	0.50	0.50	3.00	0.50
30	3.00	3.00	2.14	1.38	0.50
31	0.50	1.65	1.75	0.50	1.75
32	1.00	1.00	1.00	1.00	1.00

The concentration ranges of the treatment combinations are expressed as ×MS levels

combinations contained the same type and concentration of plant growth regulators, vitamins and sucrose as noted above. For each treatment, two boxes with five plantlets for each genotype were used (n = 10). Boxes were randomized on the growth room shelf. Cultures were grown on each treatment for 9 week, with transfers to fresh medium every 3 week.

Data structure

Three plantlets selected on a diagonal, from each Magenta™ box (n = 6) were evaluated, and the remaining four shoots were photographed for additional documentation. Shoot quality, which is a subjective visual assessment of shoot growth was evaluated as 1 = poor (no growth or bad appearance), 2 = moderate (some growth and moderate appearance), and 3 = good (well-developed shoot culture with expanded leaves and good color). Multiplication was also quantified with the same scale of quality depending on the proliferation rate of the plant, with 1 = no shoot production, 2 = 2 to 4 shoots, and 3 = > 4 shoots. Leaf color was estimated as > 50% of leaves on the shoot as follows: 1 = yellow or red, 2 = light green, 3 = green (Niedz et al. 2007; Reed et al. 2013).

Statistical analyses

Although shoot quality, multiplication, and leaf color responses are ordinal-type data, these inputs were treated as continuous by taking the mean response of six shoots per treatment for each strawberry species. The outputs were converted to continuous variables to balance the within group variation. The genotype as a nominal variable was also included within the analyses. A multiple-response model was built by the MARS approach (Table 2), using STATISTICA 7.1 software (StatSoft 2005).

The MARS algorithm is constructed by piecewise linear basis functions (*BFs*) of the following form:

$$BF1 = \max(0, x - t)\begin{cases} x - t, x > t \\ 0, x \leq t \end{cases} \tag{1a}$$

$$BF2 = \max(0, t - x)\begin{cases} t - x, t > x \\ 0, t \leq x \end{cases} \tag{1b}$$

where x is the variable range and t is the knot.

The MARS model is a linear combination of basis functions:

$$Yi = a_0 + a_1 BF1_i + a_2 BF2_i + \dots \tag{2}$$

where Y_i is the dependent variable, a_0 is the intercept, and a_1 and a_2 are the coefficients of the related basis functions (Emamgolizadeh et al. 2015; Everingham and Sexton 2011). For better understanding of the model, a simple MARS example consisting of three piecewise linear functions or splines (*BF*1, *BF*2, and *BF*3), connected at two points or knots generated by RStudio software (R Core Team 2017) is illustrated in Fig. 1. The MARS mathematical equation is expressed as:

$$Yi = a_0 + a_1 BF1_i + a_2 BF2_i + a_3 BF3_i \tag{3}$$

in which $BF1 = \max(0, x - 3)$, and $BF2 = \max(0, 3 - x)$, and $BF3 = \max(0, x - 6)$. In this case the knot points are $t = 3$ and 6. These two knots delimit the x input range in three regions, in which different linear relationships between the response and the x explanatory variable are detected (Fig. 1).

The MARS model building process consists of forward and backward stepwise selections to define the knots and splines. First, the model is overfitted by forward selection of more basis functions than required to express the response variable, and subsequently backward pruning is performed by deleting the least significant splines one at a time until an optimal model is obtained. The model is refitted after each basis function removal, and each reduced sub-optimal model is tested with the Generalized Cross-Validation (GCV) method to prevent overfitting. (Everingham and Sexton 2011). The model with the lowest GCV score is considered to be the best (Zakeri et al. 2010). An optimal MARS model is also expected to have a Pearson correlation coefficient value close to one, as the coefficient approaches zero the predictive ability of the model decreases. The MARS algorithm is sensitive to multicollinearity, which refers to correlation between inputs. When polynomial terms are included to model curvature multicollinearity problem could occur in regression type methods (Friedman 1991). Pearson correlation coefficients between pairs of the salt inputs were estimated to test for multicollinearity.

MARS is an adaptive procedure that allows users to specify many options within the model, including a maximum number of basis functions, the level of interactions between variables, and a minimum number of terms to leave between each knot depending on the data. A four-way interaction was specified, and the algorithm was assigned a maximum of 74 basis functions and a minimum of 32 terms. These values were selected to generate the best model with the highest Pearson correlation and lowest GCV score.

Results

The design of the experiment provided a diversity of responses for the tested strawberries which varied with the mineral nutrients (Fig. 2). The MARS model was able to reveal the complex relationships between the tested salt inputs and the responses by showing the relevance of each predictor on the target variable in a mathematical equation

Table 2 Basis functions and corresponding coefficients of the Multivariate Adaptive Regression Splines (MARS) model for plant quality, multiplication, and leaf color prediction in Fragaria bucharica, Fragaria chiloensis and *Fragaria x ananassa* 'Jucunda'

Terms	Coefficients (a_i)			Basis functions
	Quality	Multiplication	Leaf color	
1	2.68	3.06	2.74	Intercept
2	− 0.75	− 1.13	− 0.73	$\max(0, NH_4NO_3\text{-}2)$
3	− 0.89	− 1.37	− 0.87	$\max(0, 2\text{-}NH_4NO_3)$
4	0.28	0.46	0.26	$\max(0, KNO_3\text{-}1.81)$
5	− 0.35	− 0.4	− 0.34	$\max(0, 2\text{-}NH_4NO_3) \times \max(0, Jucunda)$
6	0.52	0.74	0.55	$\max(0, 2\text{-}NH_4NO_3) \times \max(0, MgSO_4\text{-}1.75)$
7	0.42	0.73	0.49	$\max(0, 2\text{-}NH_4NO_3) \times \max(0, 1.75\text{-}MgSO_4)$
8	0.6	0.78	0.54	$\max(0, NH_4NO_3\text{-}2) \times \max(0, 1.81\text{-}KNO_3)$
9	0.41	0.45	0.33	$\max(0, 2\text{-}NH_4NO_3) \times \max(0, 1.81\text{-}KNO_3)$
10	− 0.61	− 0.87	− 0.56	$\max(0, 2\text{-}NH_4NO_3) \times \max(0, MgSO_4\text{-}1.75) \times \max(0, F.\ bucharica)$
11	− 0.2	− 0.25	− 0.27	$\max(0, CaCl_2\text{-}1.96)$
12	0.15	0.22	0.2	$\max(0, 1.81\text{-}KNO_3) \times \max(0, KH_2PO_4\text{-}0.5) \times \max(0, Jucunda)$
13	− 1.11	− 1.35	− 1.1	$\max(0, NH_4NO_3\text{-}2) \times \max(0, 1.81\text{-}KNO_3) \times \max(0, F.\ bucharica)$
14	− 0.26	− 0.18	− 0.35	$\max(0, F.\ bucharica)$
15	− 0.46	− 0.71	− 0.46	$\max(0, KNO_3\text{-}1.81) \times \max(0, F.\ bucharica)$
16	− 0.47	− 0.64	− 0.44	$\max(0, 1.81\text{-}KNO_3) \times \max(0, F.chiloensis)$
17	0.21	0.25	0.19	$\max(0, MgSO_4\text{-}0.5) \times \max(0, KH_2PO_4\text{-}0.5) \times \max(0, F.chiloensis)$
18	− 0.23	− 0.31	− 0.27	$\max(0, 1.96\text{-}CaCl_2) \times \max(0, KNO_3\text{-}1.81)$
19	0.24	0.41	0.26	$\max(0, KNO_3\text{-}1.81) \times \max(0, KH_2PO_4\text{-}1.94)$
20	− 0.98	− 1.39	− 1	$\max(0, KH_2PO_4\text{-}1.94)$
21	− 1.34	− 1.67	− 1.28	$\max(0, CaCl_2\text{-}1.96) \times \max(0, F.\ bucharica)$
22	− 0.39	− 0.4	− 0.36	$\max(0, 1.94\text{-}KH_2PO_4) \times \max(0, Jucunda)$
23	1.08	1.49	1.04	$\max(0, CaCl_2\text{-}1.96) \times \max(0, KNO_3\text{-}0.5) \times \max(0, F.\ bucharica)$
24	− 0.61	− 0.78	− 0.56	$\max(0, NH_4NO_3\text{-}0.5) \times \max(0, CaCl_2\text{-}1.96) \times \max(0, KNO_3\text{-}0.5) \times \max(0, F.\ bucharica)$
25	0.87	1.01	0.83	$\max(0, NH_4NO_3\text{-}0.5) \times \max(0, CaCl_2\text{-}1.96) \times \max(0, F.\ bucharica)$
26	0.53	0.76	0.51	$\max(0, 2\text{-}NH_4NO_3) \times \max(0, MgSO_4\text{-}1.75) \times \max(0, KH_2PO_4\text{-}0.5) \times \max(0, F.\ bucharica)$
27	0.42	0.56	0.47	$\max(0, NH_4NO_3\text{-}0.5) \times \max(0, KH_2PO_4\text{-}1.94) \times \max(0, F.\ bucharica)$
28	0.15	0.28	0.2	$\max(0, 1.81\text{-}KNO_3) \times \max(0, KH_2PO_4\text{-}0.5) \times \max(0, F.\ chiloensis)$
29	− 0.03	− 0.05	− 0.03	$\max(0, CaCl_2\text{-}0.5) \times \max(0, MgSO_4\text{-}0.5) \times \max(0, KH_2PO_4\text{-}0.5) \times \max(0, F.\ chiloensis)$
30	− 0.27	− 0.28	− 0.3	$\max(0, 2\text{-}NH_4NO_3) \times \max(0, MgSO_4\text{-}1.75) \times \max(0, F.\ chiloensis)$
31	− 0.09	− 0.09	− 0.07	$\max(0, NH_4NO_3\text{-}0.5) \times \max(0, MgSO_4\text{-}0.5) \times \max(0, KH_2PO_4\text{-}0.5) \times \max(0, F.\ chiloensis)$
32	0.12	0.11	0.1	$\max(0, NH_4NO_3\text{-}0.5) \times \max(0, KH_2PO_4\text{-}0.5) \times \max(0, F.\ chiloensis)$

The model consists of 32 terms leaved between each knot and 74 basis functions with four-way interactions. Four cut-offs, 2.0, 1.96, 1.81, and 1.94 are the most frequent in the multi-response MARS model equation, which constituted the basis functions with the corresponding MS salts. These cut-offs were selected to make the multi-response MARS model simpler by deleting the related basis functions. Consequently, the simplified MARS model consisted of the marked coefficients and basis functions

(Table 2). The algorithm successfully incorporated the genotype-nominal variable within the analysis, and therefore was able to detect the common nutrient requirements for all three strawberry species. The predictive ability of the model for the tested outcomes was also high, with a Pearson correlation coefficient about 0.90 (p < 0.01). The GCV value was approximately 0.49. The estimated Pearson correlation coefficients between the pairs of nutrient salt factors were found to be insignificant, which indicates that no multicollinearity problem was present. Plant quality, multiplication,

and leaf color were affected by all of the tested salt inputs and the genotype factor (Table 2).

Shoot quality model

The MARS model prediction equation for shoot quality is presented in Supplement 1. To further optimize the target responses and provide common nutrient concentrations for the tested strawberry species, ranges for the multiple-response MARS model were selected as $2 \times NH_4NO_3$,

Fig. 1 Graphical representation of a simple Multivariate Adaptive Regression Splines (MARS) model represented by three piecewise linear basis functions (BF1, BF2, and BF3) connected to each other with knots (indicated in blue). The knots show the regions of association change between the explanatory and target variable. These two knots delimit the x input range in three regions, in which different linear relationships between the response and the x explanatory variable are detected
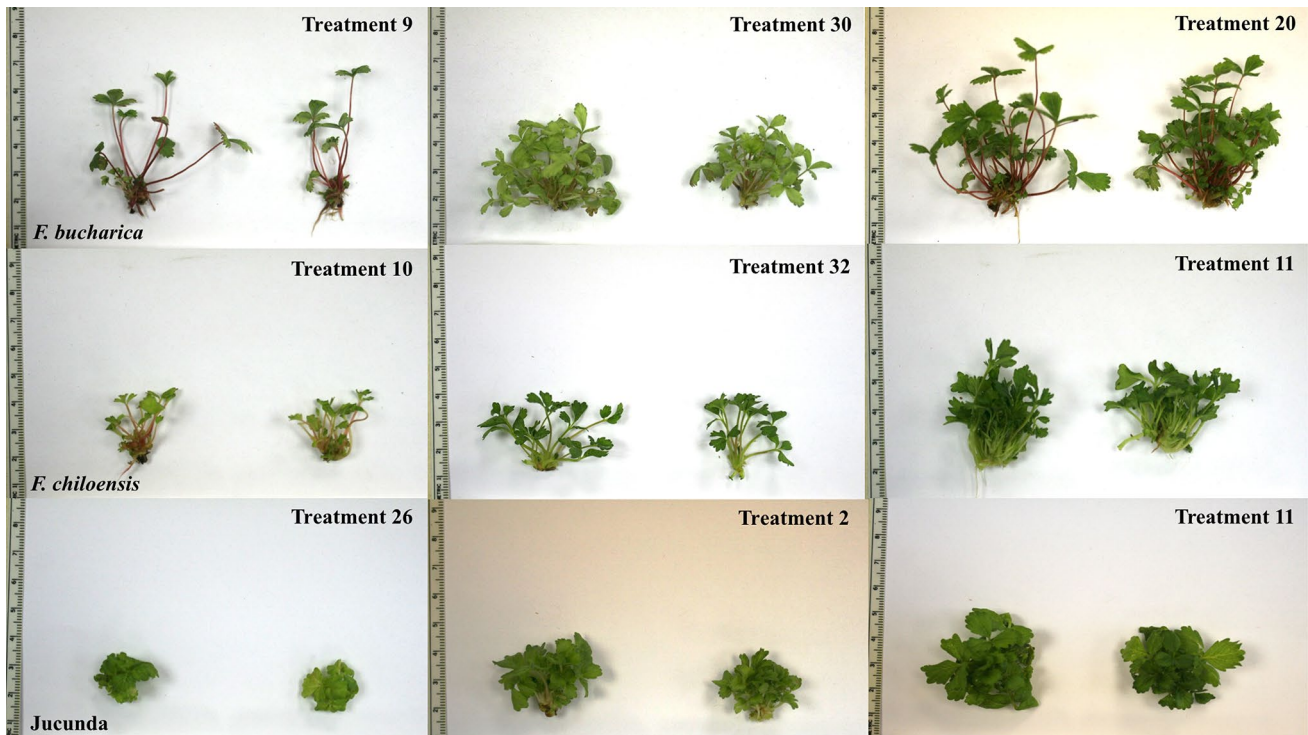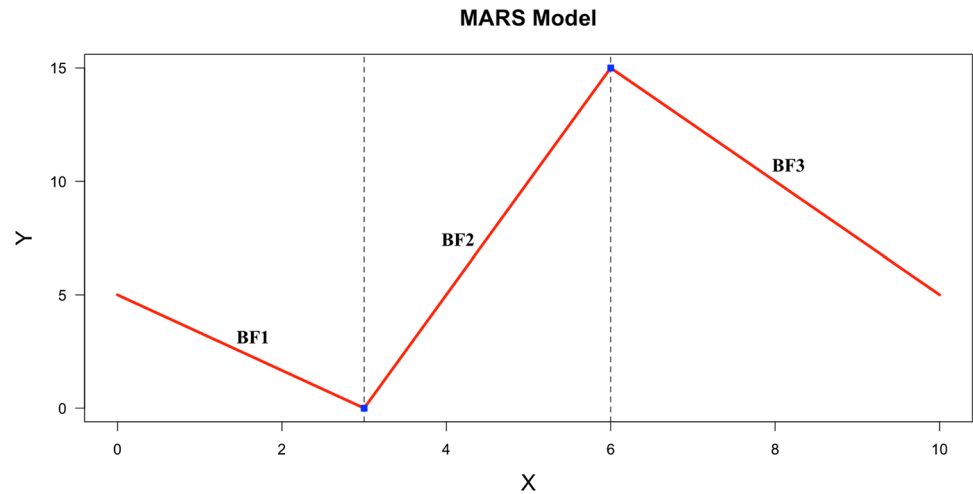
Fig. 2 Growth response of shoots of *Fragaria bucharica*, *Fragaria chiloensis* and *Fragaria x ananassa* 'Jucunda' grown on treatments with variations in MS major salt concentrations (Table 1). Treatments were chosen to show the range of plant quality, multiplication, and leaf color produced by changes in mineral nutrition.

$1.96 \times CaCl_2$, $3 \times MgSO_4$, $1.81 \times KNO_3$, and $1.94 \times KH_2PO_4$. These values were defined by looking at the MARS equation (Table 2). Four cut-offs, 2.0, 1.96, 1.81, and 1.94, were the most frequent in the equation, which constituted the basis functions with the corresponding salts [max (0, NH_4NO_3-2), max (0, $CaCl_2$-1.96), max (0, KNO_3-1.81), and max (0, KH_2PO_4-1.94)]. These cut-offs were selected to make the model simpler by deleting the related basis functions.

Finally, $MgSO_4$ was assigned a value of 3 to maximize the responses. After assigning these values, most of the basis functions were deleted according to the rules in Eq. (1a) and (1b) (max (0, NH_4NO_3-2) was equal to 0 when $NH_4NO_3 < 2$), and the model for plant quality was simplified to:

Quality $= 2.68 - 0.26 \times$ max (0, *F. bucharica*) $+ 0.21 \times$ max (0, $MgSO_4$-0.5) $\times$ max (0, KH_2PO_4-0.5) $\times$ max (0, *F.chiloensis*) $- 0.03 \times$ max (0, $CaCl_2$-0.5) $\times$ max (0,

$MgSO_4$-0.5) × max (0, KH_2PO_4-0.5) × max (0, *F. chiloensis*) − 0.09 × max (0, NH_4NO_3-0.5) × max (0, $MgSO_4$-0.5) × max (0, KH_2PO_4-0.5) × max (0, *F. chiloensis*) + 0.12 × max (0, NH_4NO_3-0.5) × max (0, KH_2PO_4-0.5) × max (0, *F. chiloensis*)

The model was further reduced based on the rules of Eq. (1a) and (1b), by subtracting the assigned salt values and the knot points (*e.g.*, max (0, $MgSO_4$-0.5) was 3-0.5 = 2.5 when $MgSO_4$ > 0.5).

The final model for shoot quality was:

Quality = 2.68 − 0.26 × max(0, *F. bucharica*) + 0.76 × max (0, *F. chiloensis*) − 0.16 × max(0, *F. chiloensis*) − 0.49 × max (0, *F. chiloensis*) + 0.26 × max (0, *F. chiloensis*)

The mean quality of *F. bucharica* was predicted to be:

2.68 − 0.26 × 1 + 0.76 × 0 − 0.16 × 0 − 0.49 × 0 + 0.26 × 0 = 2.68 − 0.26 = 2.42

Genotype was treated as a nominal categorical variable, thus 1 was used for max (0, *F. bucharica*) and 0 otherwise [in this case for max (0, *F. chiloensis*)].

Using the same approach, the average shoot quality of *F. chiloensis* was:

2.68 − 0.26 × 0 + 0.76 × 1 − 0.16 × 1 − 0.49 × 1 + 0.26 × 1 = 3.05

The predicted mean quality of 'Jucunda' was

2.68 − 0.26×0 + 0.76×0 − 0.16×0 − 0.49×0 + 0.26×0 = 2.68

Shoot multiplication model

The MARS model for predicting multiplication (Table 2), after inserting 2 = NH_4NO_3, 1.96 = $CaCl_2$, 3 = $MgSO_4$, 1.81 = KNO_3 and 1.94 = KH_2PO_4 values, was reduced to:

Multiplication = 3.06 − 0.18 × max (0, *F. bucharica*) + 0.9 × max (0, *F. chiloensis*) − 0.26 × max (0, *F. chiloensis*) − 0.49 × max (0, *F. chiloensis*) + 0.24 × max (0, *F. chiloensis*)

The predicted mean multiplication of *F. bucharica* was:

3.06 − 0.18 × 1 + 0.9 × 0 − 0.26 × 0 - 0.49 × 0 + 0.24 × 0 = 3.06 − 0.18 = 2.88

The predicted average multiplication of *F. chiloensis* was:

3.06 − 0.18 × 0 + 0.9 × 1 − 0.26 × 1 − 0.49 × 1 + 0.24 × 1 = 3.06 + 0.9 − 0.26 − 0.49 + 0.24 = 3.45

The predicted mean multiplication of 'Jucunda' was estimated as 3.06.

Leaf color model

The MARS model to predict leaf color (Table 2), after replacing 2 = NH_4NO_3, 1.96 = $CaCl_2$, 3 = $MgSO_4$, 1.81 = KNO_3 and 1.94 = KH_2PO_4 values, was shortened to:

Leaf color = 2.74 − 0.35 × max (0, *F. bucharica*) + 0.68 × max (0, *F. chiloensis*) − 0.16 × max (0, *F.*

chiloensis) − 0.38 × max (0, *F. chiloensis*) + 0.22 × max (0, *F. chiloensis*)

The mean leaf color of *F. bucharica* was:

2.74 − 0.35 = 2.39

The average leaf color of *F. chiloensis* was:

2.74 + 0.68 − 0.16 – 0.38 + 0.22 = 3.1

The predicted mean leaf color of "Jucunda" was estimated as 2.74. The mean quality, multiplication, and leaf color of 'Jucunda' depended only on the intercepts of the model under 2 × NH_4NO_3, 1.96 × $CaCl_2$, 1.81 × KNO_3 and 1.94 × KH_2PO_4. Therefore, it can be concluded that the responses of "Jucunda" won't change under any $MgSO_4$ concentration. However, the aim was to define a common medium for the tested strawberry species. Therefore, 3 × $MgSO_4$ was specified for the final medium.

The software user doesn't need to make all of those tedious calculations, because only stating the factor values will be enough to obtain the predicted responses. Those equations have been demonstrated to show how transparent and computationally practical the MARS algorithm is for analyzing in vitro culture experiments.

Shoot response to treatment combinations

As expected, none of the treatment combinations (Table 1) exactly matched the suggested optimal macronutrient ranges predicted by the multiple response MARS algorithm. However, growth of shoots of *F. bucharica*, *F. chiloensis* and 'Jucunda' all showed improvement on some treatments, with some of the nutrients in similar ranges compared to the control NCGR *Fragaria* medium treatment (Fig. 3).

Discussion

Developing a growth medium for diverse species of *Fragaria* provides a useful tool for germplasm collections and breeding stocks. Many of the commonly grown cultivars can be cultured on standard MS medium, however many of the wild species are more difficult. This is the first time that modeling has been used to determine improved medium components for strawberry germplasm.

Modeling using non-parametric machine learning techniques has been utilized in only a few tissue culture studies (Akin et al. 2017; Kovalchuk et al. 2018; Nezami-Alanagh et al. 2018). This is mainly because the plant science researchers are not familiar with this type of statistical modelling. Therefore, this study aimed to introduce the MARS algorithm, which is a promising new tool to analyze in vitro data. The MARS algorithm is a powerful data mining technique that is able to reveal high dimensional, non-linear, and interactional effects among multiple inputs and output
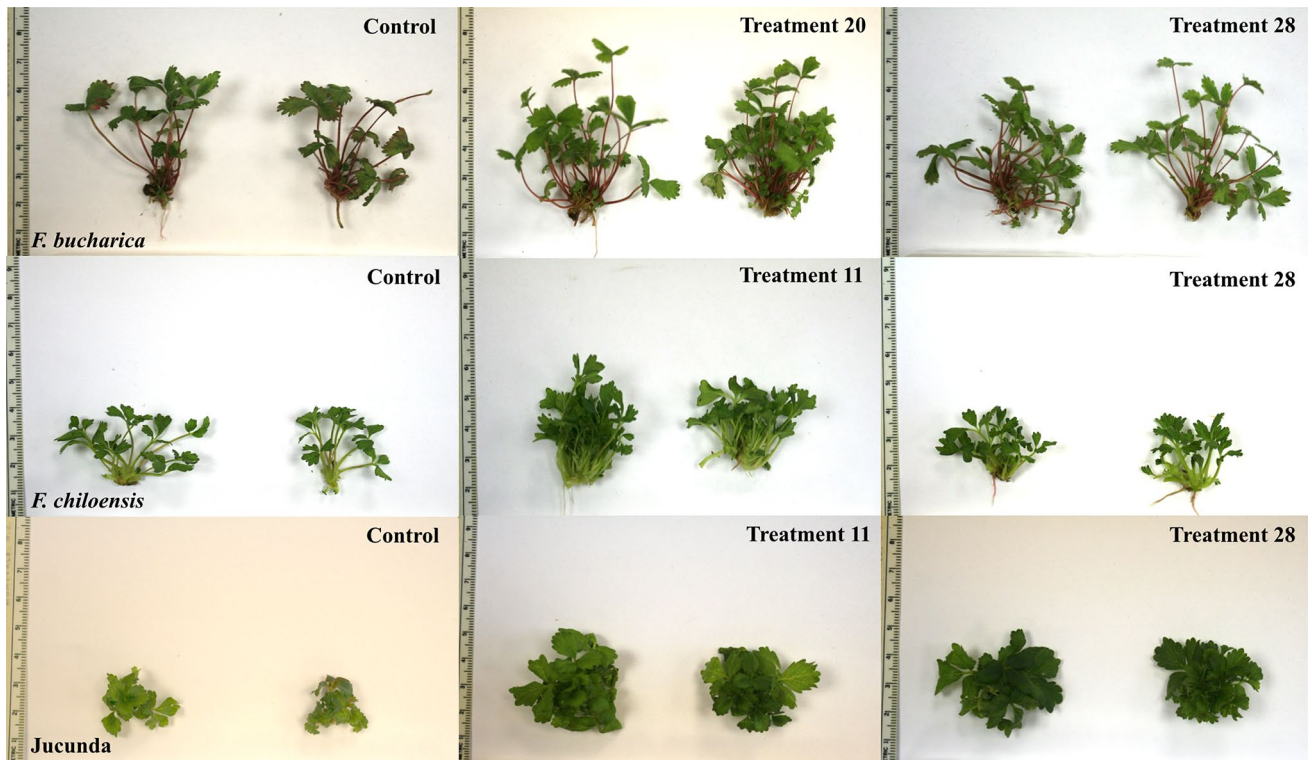
Fig. 3 A comn of shoots of *Fragaria bucharica*, *Fragaria chiloensis* and *Fragaria x ananassa* 'Jucunda' grown on the modified MS (NCGR medium) control (treatment 32) and two improved treatments (Table 1). These improved treatments included some, but not all, of the suggested optimal macronutrient ranges predicted by the Multivariate Adaptive Regression Splines (MARS) algorithm

without requiring the restrictive distributional assumptions of the traditional linear models (Friedman 1991).

Response Surface Methodology is a parametric method, which is able to describe non-linear associations and is used in tissue culture research. However, this method cannot include categorical variables such as genotype within the model. Therefore, RSM projects genotype-specific input requirements for the corresponding dependent variable (Akin et al. 2016). In a study of raspberry mineral nutrition, RSM generated individual models for each raspberry genotype. This resulted in five graphs for plant quality, one for each genotype, and each with a different result rather than a common model. The best treatments were with high NH_4NO_3 and KNO_3, as well as low minors and iron (compared to MS medium) for 'Canby'; high mesos, low minors and iron for 'Indian Summer'; high mesos and low iron for 'Nootka' and 'Trailblazer'; and high mesos and minors for 'Willamette'. Low KNO_3 and high mesos were required for higher multiplication, and depending on genotype NH_4NO_3 and iron were also significant (Poothong and Reed 2014). In a study of hazelnut shoot culture (Akin et al. 2016), analysis using the CHAID decision tree predicted higher plant quality for three genotypes at $NH_4NO_3 < 1.701 \times$ and $KH_2PO_4 < 2.012 \times$ DKW medium (Driver and Kuniyuki 1984). Multiplication was predicted

to be optimal at $NH_4NO_3 < 0.5 \times$ and $CaNO_3 < 1.725 \times$ (Akin et al. 2016). The CHAID algorithm can include nominal variables within the analyses and provides common nutrient requirements for the tested genotypes but is not capable of performing multiple responses simultaneously such as the MARS algorithm. The CHAID algorithm constructs separate visual trees for each target variable, and the user must determine the common ranges for the tested outcomes (Akin et al. 2016). Medium improvements for wild apricot cultures from Kazakhstan (Kovalchuk et al. 2017) using the Classification and Regression Tree (CART) algorithm, mapped $KH_2PO_4 < 2.41 \times$ and $MgSO_4 < 0.75 \times$ for plant quality, $0.77 \times < KH_2PO_4 < 0.91 \times$ for multiplication, $KH_2PO_4 < 2.41 \times$ and $K_2SO_4 < 1.22 \times$ WPM (Lloyd and McCown 1980) for leaf color. In addition, the CART regression tree is not able to generate a multi-response tree but forms separate models for each target variable. The neuro-fuzzy logic technique described the $SO_4^{2-} \times Cl^-$ interaction and the Fe^{2+}, K^+, NH_4^+ linear independent effects, as significant for quality of two pistachio rootstocks. $K^+ \times EDTA^- \times SO_4^{2-}$ and $Fe^{2+} \times BO_3^-$ interactions played crucial roles for multiplication (Nezami-Alanagh et al. 2018). Artificial neural networks can utilize both qualitative and quantitative variables and reveal complex non-linear effects. However, these methods do not provide a transparent mathematical

equation that shows the relative importance of the inputs on the output (Olden et al. 2008). None of these techniques have been used to improve strawberry growth medium.

Murashige and Skoog (1962) medium with some modifications was used as the standard for multiplication of a wide range of in vitro cultures at the NCGR, however a large variation in the response indicated that it is not ideal for many species and cultivars (Reed and Hummer 1995). As seen in the current study, a medium with increased amounts of five major MS nutrients ($2 \times NH_4NO_3$, $1.96 \times CaCl_2$, $3 \times MgSO_4$, $1.81 \times KNO_3$ and $1.94 \times KH_2PO_4$), is needed to provide improved growth for the three species tested without changing growth regulator concentrations. The next step in this study would be to test this medium on additional accessions to determine the applicability of this medium to large germplasm collections.

Conclusions

The MARS algorithm is a promising tool to advance understanding and prediction of biological events. This method can reveal non-linear and multidimensional relationships, and typically outcompete basic linear models, which makes it ideal to use for in vitro data. MARS is a non-parametric approach that is able to handle continuous and categorical data, without the need to satisfy the restrictive distributional assumptions required by the traditional parametric techniques. The MARS model clearly showed the direct effect and the complex interactions of the parameters. The target variables were affected by all of the tested mineral nutrients and the genotype input. This study was able to define the common major nutrient requirements of the tested strawberry species, due to the inherent characteristic of the MARS algorithm to incorporate nominal variables as predictors within the model. Thanks to the ability of MARS to analyze multiple responses simultaneously, the optimal major salts for good shoot quality, multiplication, and greener leaves were predicted to be: 3300 mg L^{-1} NH_4NO_3, 862.4 mg L^{-1} $CaCl_2$, 1110 mg L^{-1} $MgSO_4$, 3439 mg L^{-1} NO_3 and 329.8 mg L^{-1} KH_2PO_4.

Author contributions MA planned and executed the experiment, collected the data, drafted the manuscript and assisted with analysis of the data. SPE helped with data analysis interpretation and writing the manuscript. EE performed the statistical analysis. BMR assisted with planning and analysis, supervised the study, and edited the manuscript.

Compliance with ethical standards

References

Akin M, Eyduran E, Reed BM (2016) Use of RSM and CHAID data mining algorithm for predicting mineral nutrition of hazelnut. PCTOC 128(2):303–316

Akin M, Hand C, Eyduran E, Reed BM (2017) Predicting minor nutrient requirements of hazelnut shoot cultures using regression trees. PCTOC 132(3):545–559

Darrow GM (1966) The strawberry: history, breeding and physiology. Holt, Rinehart and Winston, New York

Design-Expert (2010) Stat-Ease, Inc., Minneapolis, MN

Dettori JR, Norvell DC (2018) The anatomy of data. Global Spine J 8(3):311–313

Driver JA, Kuniyuki AH (1984) In vitro propagation of Paradox walnut rootstock. HortScience 19:507–509

Emamgolizadeh S, Bateni SM, Shahsavani D, Ashrafi T, Ghorbani H (2015) Estimation of soil cation exchange capacity using genetic expression programming (GEP) and multivariate adaptive regression splines (MARS). J Hydrol 529(3):1590–1600

Everingham YL, Sexton J (2011) An introduction to multivariate adaptive regression splines for the cane industry. Proceedings of the 2011 Conference of the Australian Society of Sugar Cane Technologists

Eyduran E, Akin M, Eyduran SP (2019) Application of multivariate adaptive regression splines through R Software. Nobel Academic Publishing, Ankara

Friedman J (1991) Multivariate adaptive regression splines. Ann Stat 19(1):1–67

Kovalchuk IY, Mukhitdinova Z, Turdiyev T, Madiyeva G, Akin M, Eyduran E, Reed BM (2017) Modeling some mineral nutrient requirements for micropropagated wild apricot shoot cultures. PCTOC 129(2):325–335

Kovalchuk IY, Mukhitdinova Z, Turdiyev T, Madiyeva G, Akin M, Eyduran E, Reed BM (2018) Nitrogen ions and nitrogen ion proportions impact the growth of apricot (*Prunus armeniaca*) shoot cultures. PCTOC 133(2):263–273

Linsmaier EM, Skoog F (1965) Organic growth factor requirements of tobacco tissue cultures. Physiol Plant 18:100–127

Lloyd G, McCown B (1980) Commercially feasible micropropagation of mountain laurel, Kalmia latifolia, by use of shoot-tip culture. Comb Proc Int Plant Prop Soc 30:421–427

Mertler C, Vannatta R (2002) Advanced and multivariate statistical methods: practical application and interpretation, 2nd edn. Pyrczak Publishing, Los Angeles

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant 15:473–497

Nezami-Alanagh E, Garoosi GA, Landin M, Gallego PP (2018) Combining DOE With neurofuzzy logic for healthy mineral nutrition of pistachio rootstocks in vitro culture. Front Plant Sci 9:1474. https://doi.org/10.3389/fpls.2018.01474

Niedz RP, Evens TJ (2016) Design of experiments (DOE)—history, concepts, and relevance to in vitro culture. In Vitro Cell Dev Biol Plant 52(6):547–562

Niedz RP, Hyndman SE, Evens TJ (2007) Using a gestalt to measure the quality of in vitro responses. Sci Hort 112(3):349–359

Olden JD, Lawler JJ, Poff NL (2008) Machine learning methods without tears: a primer for ecologists. Q Rev Biol 83:171–193

Poothong S, Reed BM (2014) Modeling the effects of mineral nutrition for improving growth and development of micropropagated red raspberries. Sci Hort 165(0):132–141

Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, Cambridge

Quiroz KA, Berríos M, Carrasco B, Retamales JB, Caligari PDS, García-Gonzáles R (2017) Meristem culture and subsequent

micropropagation of Chilean strawberry (*Fragaria chiloensis* (L.) Duch. Biol Res 50(1):20

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reed BM, Hummer KE (1995) Conservation of germplasm of strawberry (*Fragaria* Species). In: Bajaj YPS (ed) Cryopreservation of Plant Germplasm I, vol 32, Springer, Berlin

Reed BM, Wada S, DeNoma J, Niedz RP (2013) Improving in vitro mineral nutrition for diverse pear germplasm. In Vitro Cell Dev Biol Plant 49(3):343–355

Simpson DW, Bell JA (1989) The response of different genotypes of *Fragaria × ananassa* and their seedling progenies to in vitro micropropagation and the effects of varying the concentration of 6-benzylaminopurine in the proliferation medium. PCTOC 17(2–3):225–234

StatSoft (2005) STATISTICA, Inc., (Data Analysis Software System), Version 7.1. http://www.statsoft.com

Zakeri IF, Adolph AL, Puyau MR, Vohra FA, Butte NF (2010) Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents. J Appl Physiol 108:128–136

Zhang W, Goh ATC (2014) Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geosci Front 7:45–52