Ekoloji 27(106): 395-404 (2018)

Study on the Relationship between Blueberry Soil and Leaf in Southern China–Based on MLR-PCA Prediction Model

Jian Li [1], Ming-Yue Wang [1], Guang Chen [2*], Lin Wu [3], Min Zhang [1]

[1] Jilin Agricultural University, College of Information Technology, Changchun, Jilin 130118, CHINA

[2] Jilin Agricultural University, College of Life Sciences, Changchun, Jilin 130118, CHINA

[3] Jilin Agricultural University, College of Horticulture, Changchun, Jilin 130118, CHINA

[*] Corresponding author: chg61@163.com

Abstract

This paper selects six blueberry varieties from Miluo City as the sample, constructs MLR-Principal Component Prediction Model for Soil and Leaf Nutrients, studies deeply the correlation and predictive function between different blueberry varieties' soil and leaf nutrient contents. Compared with several prediction methods, the principal component prediction function combined with multiple regression is more systematic and has a better effect on the prediction of blueberry nutrition status. The results show that the nutrient content of the leaves is highly correlated with the elements of the soil, the nutritional status of different blueberry varieties' leaves can be fully expressed with predictive functions, and uses the MLR regression equation and regression prediction model to guide the technicians for matching the fertilization programs, positively guides the blueberry planting industry in practical applications.

Keywords: blueberry, soil, leaf blade, prediction, correlation

Li J, Wang M-Y, Chen G, Wu L, Zhang M (2018) Study on the Relationship between Blueberry Soil and Leaf in Southern China–Based on MLR-PCA Prediction Model. Ekoloji 27(106): 395-404.

INTRODUCTION

Blueberry is Vaccinium spp. plant of the family Ericaeae, a kind of wild berry, it has been listed as one of the five health foods of mankind by the FAO, because of its rich nutritional value and high economic value (Li 2017). International blueberries are mainly grown in 27 countries and most varieties are produced in temperate regions (Sakhanokho et al. 2018), in the selection of blueberry research directions, foreign scholars prefer the blueberries's biological characteristics (Mannozzi et al. 2018), interspecific hybridization (Sakhanokho et al. 2018), fruit utility (Istek et al. 2017), simulation (Qu and Drummond 2018), etc. But research on the law of elements' increase and decrease in blueberry leaves or soil remains hadn't been undeveloped. At present, there are more than 90 varieties of blueberries in China, and the blueberry cultivation area has achieved about 12,000 hm^2, is mainly distributed in more than 10 provinces such as Shandong, Liaoning, Jilin and Guizhou (Lei et al. 2017). However, domestic scholars mainly study on the aspects of picking, torage, antioxidation, drought resistance and extraction techniques, studies on the nutrient elements' dynamics and correlations of blueberries' soil and leaves are not yet substantial (Shi et al. 2017). Therefore, the study focused on the soil and leaves of six blueberries in Miluo City, Hunan Province, focused on exploring the degree of interaction between different nutrient components and the nutritional prediction of the leaves.

Gray prediction, state space models, exponential smoothing methods and principal component regression analysis methods are the means to achieve forecasting function, in practical applications, only a single algorithm can't fully express the law of crop development, so they are usually used two or more blending algorithms for fusion operation. Mir et al. (2017) used clustering and principal component analysis to evaluate the diversity of apple germplasm, Maleki et al. (2014) combined genetic algorithm, stepwise multiple linear regression and artificial neural network methods to predict aromatic sulfonamide derivatives, Riccardi et al. (2014) performed a nondestructive evaluation of buckwheat and leeks leaves' chlorophyll content using simple multiple regression analysis. So in the research of agriculture and fruit prediction methods, there is still a lack of a hybrid

 Received: 2 Oct 2017 / Accepted: 24 Sep 2018

method that combines principal component regression analysis.

This paper used factor analysis, principal component analysis and multiple linear regression analysis for 6 blueberry (Misty, O'Neal, Duke, Legacy, Bluecrop and Northland) varieties' multivariate correlations analysis between soil and leaf nutrients, established separately the multivariate principal component regression equation and principal component load diagram (PCA) between soil and leaf, sorted out the comprehensive principal component regression prediction function of each blueberry variety, and then screened out the most influential element in different blueberry varieties, coefficient of influence between nutritional components and the leaf nutrient prediction function for soil. We hope to provide reference for the cultivation and production of blueberries in Miluo City, Hunan Province (Shen 2010, Wu 2016), and provide theoretical reference value and positive guidance for blueberry planting in South China.

MATERIALS AND METHODS

The experiment was conducted from May 2016 to March 2017, the sampling site was located at the blueberry planting base in Miluo City, Hunan Province. A total of 6 varieties for testing, and good growth status. Took 0-25cm deep soil under every point canopy of each species in the area, evenly mixed the 1.5-2.0kg soil collected at 5 points, and used a marker to store the number inside the polyethylene plastic bag. Leaf samples of each species in each orchard were selected diagonally by 25 strains of robust and disease-free plants, four strains were harvested in per plant. Each repeat 100 leaves, 3 repeats group make up 1 sample. At the same time, determined final elemental of the selected soil and leaf samples.

RESULTS AND ANALYSIS

Linear Correlation Judgment and Data Significant Difference Analysis

Linear correlation judgment

Six blueberry varieties (Misty, O'Neal, Duke, Legacy, Bluecrop, and Northland) are respectively labeled as a, b, c, d, e, f, and according to the standardized data trend of six blueberry varieties' soil and leaves conducted correlation analysis by SPSS software (**Table 1**).

From **Table 1**, there was a clear linear relationship between most of the soil elements and the nutrient elements in the leaves, such as the correlation coefficient between the N, P, K, Ca, Mg, Mn, Zn elements between the six soil varieties and the N element in their respective leaves was greater than 0.5, indicated that there was a significant linear correlation between the soil and leaves' N element of the six varieties. Therefore, a multiple linear regression method can be used to further explore the regression equation and influence degree between soil and leaves in different blueberry varieties.

Analysis of data significance differences

Under normal circumstances, when the experimental results of multiple sets of data reached 0.05 levels or 0.01 levels ($F_{0.05}$ and $F_{0.01}$), there was a significant or extremely significant difference between the marker data, also known as the selected data are statistically significant. This paper' data selected from May to September of 2016, so noted A as the significant difference between N, P, K, Ca, Mg, Fe, Mn, Cu, and Zn elements, and B as data significant differences from May to September, AB as significant differences in interactions between elements and months, the calculation results are shown in **Table 2**.

We can see that all significant differences F value in the analysis were greater than $F_{0.05}$, indicated they had a significant difference between the data; Some of the values appeared $F_{0.05} < F < F_{0.01}$, they had a very significant difference between the data. It can be seen that all the data selected in this study had significant differences, they were both statistically significant and could be used for the trial's study.

Multiple Linear Regression (MLR)

The subject of this study is the influence degree and the correlation analysis on the nutrient elements of soil and leaves in different blueberry varieties, the research type belongs to the multi-indicator and multi-indicator, if it applies simply multivariate analysis of variance and stepwise regression analysis, it is impossible to thoroughly study the correlation between all nutrient elements. This study applies the idea of multiple regression, firstly constructs multiple linear regression equations for different nutrients in each leaf of blueberry varieties, in order to preliminarily predict the trend and influence degree of different nutrients ingredient in the soil on each type of nutrient elements in the leaves.

Table 1. Data correlation between soil and leaves of 6 blueberries

		Leaf nutrient elements								
		N	**P**	**K**	**Ca**	**Mg**	**Fe**	**Mn**	**Cu**	**Zn**
Soil	N	a:0.961** b:0.608 c:0.911** d:0.719 e:0.611 f:0.87	a:0.748 b:0.959* c:-0.063 d:0.986 e:0.994** f:0.283	a:-0.68 b:0.604 c:0.827 d:-0.746 e:-0.506 f:0.786	a:-0.363 b:0.81 c:0.16 d:-0.485 e:0.92* f:088*	a:0.933* b:0.679 c:-0.753 d:-0.362 e:-0.815 f:0.949*	a:0.814 b:0.168 c:0.207 d:-0.411 e: 0.064 f:0.485	a:-0.222 b:-0.254 c:0.729 d:0.766 e:0.706 f:-0.238	a:0.719 b:-0.663 c:-0.917* d:-0.88* e:-0.486 f:0.594	a:-0.866 b:-0.563 c:0.956* d:-0.052 e:0.801 f:0.535
	P	a:0.955* b:0.870 c:0.347 d:0.906 e:-0.646 f:0.585	a:0.771 b:0.606 c:0.783 d:0.708 e:-0.987** f:0.376	a:0.643 b:0.350 c:0.670 d:-0.978** e:0.641 f:0.593	a:-0.383 b:0.905* c:-0.804 d:-0.397 e:-0.98** f:0.581	a:0.985** b:0.794 c:-0.782 d:-0.663 e:0.903* f:0.522	a:0.885* b:0.246 c:-0.445 d:-0.563 e:0.026 f:0.513	a:-0.359 b:-0.744 c:0.234 d:0.676 e:-0.525 f:-0.267	a:0.618 b:-0.547 c:-0.544 d:-0.855 e:0.317 f:0.270	a:-0.921* b:-0.872 c:0.372 d:-0.150 e:-0.719 f:0.257
	K	a:0.936* b:0.841 c:0.884* d:0.241 e:0.149 f:-0.443	a:0.556 b:0.060 c:0.884* d:-0.028 e:0.237 f: 0.172	a:-0.746 b:-0.211 c:0.037 d:-0.161 e:-0.760 f: -0.348	a:-0.094 b:0.565 c:0.597 d:-0.817 e:0.389 f: -0.542	a:0.760 b:0.745 c:-0.261 d:-0.653 e:-0.648 f: -0.647	a:0.568 b:0.495 c:-0.975** d:-0.184 e:-0.780 f: 0.270	a:0.108 b:-0.898* c:-0.277 d:0.251 e:-0.525 f: -0.750	a:0.891* b:-0.408 c:0.303 d:-0.300 e:0.489 f:-0.721	a:-0.643 b:-0.855 c:-0.712 d:0.808 e:-0.047 f:-0.900**
	Ca	a:0.749 b:0.309 c:-0.889* d:0.980** e:0.154 f:0.661	a:0.057 b:0.559 c:-0.064 d:0.668 e:0.119 f:0.284	a:-0.783 b:0.790 c:-0.795 d:-0.869 e:-0.742 f:0.642	a:0.509 b:0.467 c:0.112 d:-0.654 e:0.271 f:0.728	a:0.373 b:0.537 c:0.925* d:-0.856 e:-0.574 f:0.675	a:0.070 b:0.869 c:-0.007 d:-0.857 e:-0.869 f:0.122	a:0.596 b:0.246 c:-0.592 d: 0.517 e:-0.589 f:0.743	a:0.956* b:-0.858 c:0.826 d:-0.868 e:0.476 f:0.920*	a:-0.163 b:0.040 c:-0.821 d:-0.025 e:-0.075 f:0.986**
	Mg	a:-0.630 b:0.522 c:-0.313 d:-0.300 e:0.505 f:0.903*	a:-0.197 b:0.935* c:-0.017 d:-0.540 e:0.326 f:0.681	a:0.819 b:0.799 c:-0.765 d:0.153 e:-0.813 f:0.892*	a:0.107 b:0.733 c:-0.454 d:-0.165 e:0.369 f:0.861	a:-0.567 b:0.712 c:0.116 d:-0.144 e:-0.578 f:0.786	a:-0.437 b:0.585 c:-0.809 d:0.464 e:-0.954* f:0.761	a:-0.119 b:0.026 c:-0.975** d:0.006 e:-0.194 f:-0.190	a:-0.465 b:-0.900** c:0.576 d:0.245 e:-0.037 f:0.749	a:0.520 b:-0.302 c:-0.717 d:0.649 e:0.387 f:0.501
	Fe	a:0.056 b:0.094 c:0.460 d:0.127 e:-0.070 f:-0.175	a:-0.663 b:0.491 c:0.420 d:-0.413 e:0.375 f:-0.833	a:-0.480 b:0.850 c:0.909* d:0.128 e:0.466 f:-0.325	a:0.911* b:0.322 c:0.015 d:0.459 e:0.080 f:-0.058	a:-0.373 b:0.332 c:-0.376 d:-0.074 e:0.063 f:0.219	a:-0.638 b:0.753 c:0.302 d:-0.632 e:0.548 f:-0.805	a:0.988** b:0.456 c:0.819 d:-0.745 e:0.857 f:0.472	a:0.508 b:-0.733 c:-0.855 d:0.341 e:-0.754 f:-0.006	a:0.553 b:0.265 c:0.856 d:-0.749 e:0.504 f:0.336
	Mn	a:0.569 b:0.672 c:0.747 d:0.978** e:0.633 f:0.338	a:0.630 b:0.774 c:0.273 d:0.632 e:0.623 f:-0.410	a:-0.454 b:0.834 c:0.949* d:-0.908* e:-0.034 f:0.197	a:-0.692 b:0.822 c:-0.142 d:-0.737 e:0.464 f:0.441	a:0.820 b:0.834 c:-0.831 d:-0.926* e:-0.152 f:0.643	a:0.876 b:0.854 c:0.168 d:-0.783 e:0.367 f:-0.405	a:-0.610 b:-0.135 c:0.776 d:0.594 e:0.943* f:0.550	a:0.058 b:-0.994** c:-0.858 d:-0.894* e:-0.873 f:0.511	a:0.553 b:-0.377 c:0.856 d:0.124 e:0.832 f:0.732
	Cu	a:0.524 b:-0.095 c:0.185 d:-0.374 e:-0.487 f:-0.743	a:-0.131 b:0.661 c:0.327 d:0.378 e:-0.472 f:-0.623	a:-0.587 b:0.857 c:0.839 d:0.386 e:0.043 f:-0.783	a:0.675 b:0.232 c:0.034 d:0.126 e:-0.213 f:-0.764	a:0.088 b:0.159 c:-0.338 d:0.641 e:0.088 f:-0.600	a:-0.212 b:0.404 c:0.672 d:0.343 e:0.083 f:-0.442	a:0.763 b:0.644 c:0.932* d:0.076 e:-0.811 f:-0.578	a:0.896* b:-0.591 c:-0.436 d:0.072 e:0.969** f:-0.952*	a:0.127 b:0.369 c:0.530 d:-0.057 e:-0.863 f:-0.875
	Zn	a:-0.687 b:-0.727 c:0.741 d:-0.079 e:0.670 f:-0.798	a:-0.531 b:-0.827 c:0.192 d:0.255 e:0.918* f:-0.563	a:0.543 b:-0.758 c:0.954* d:0.055 e:-0.388 f:-0.816	a:0.315 b:-0.851 c:-0.011 d:-0.750 e:0.760 f:-0.829	a:-0.602 b:-0.881* c:-0.758 d:-0.171 e:-0.619 f:-0.700	a:-0.507 b:-0.801 c:0.294 d:0.284 e:0.020 f:-0.418	a:-0.024 b:0.209 c:0.847 d:0.477 e:0.862 f:-0.559	a:-0.646 b:0.994** c:-0.858 d:-0.284 e:-0.770 f:-0.970**	a:0.570 b:0.473 c:0.888* d:0.922* e:0.952* f:-0.924*

Table 2. Significant Differences Analysis of Six Blueberry Data

classification / difference	Soil				Leaf			
	F	$F_{0.05}$	$F_{0.01}$	**Significant**	**F**	$F_{0.05}$	$F_{0.01}$	**Significant**
A (element)	a: 5410283	1.917827	2.486232	★★	a: 521389	2.042986	2.715364	★★
	b:19039211			★★	b: 3425.85			★★
	c:12557420			★★	c: 411604.7			★★
	d:34452510			★★	d: 347231.6			★★
	e: 8270232			★★	e: 113755.9			★★
	f: 7134309			★★	f: 377023			★★
B (month)	a: 3.69935	2.454213	3.494555	★	a: 2.62852	2.472927	3.534992	★
	b: 11.43681			★★	b: 2.55111			★
	c: 4.24319			★★	c: 5.474478			★★
	d: 2.596829			★	d: 2.976621			★
	e: 3.909368			★★	e: 2.644635			★
	f: 3.122511			★	f: 2.541261			★
AB	a: 1.507393	1.504268	1.77844	★	a: 3.036828	1.572028	1.892373	★★
	b: 6.545495			★★	b: 2.542884			★★
	c: 1.625361			★	c: 2.772447			★★
	d: 3.696426			★★	d: 2.915943			★★
	e: 2.604963			★★	e: 1.753773			★
	f: 1.740352			★	f: 2.113083			★★

Note: ★ indicates significant difference;
★★ indicates a very significant difference

Table 3. Multivariate Linear Regression Equations of Mist Leaves - Soil

R^2	F	Sig	Equation
0.963	25.687	0.000	$y_N = -1.988x_N - 0.008x_P + 0.356x_{Ca} + 2.361x_{Mg} - 0.003x_{Mn} - 0.019x_{Cu} + 0.085x_{PH}$
0.993	59.227	0.001	$y_P = 0.39x_N + 0.002x_P + 0.637x_k + 0.2x_{Ca} + 1.303x_{Mg} - 0.002x_{Fe} + 0.007x_{Mn}$ $+0.012x_{Cu} - 0.016x_{\text{Organic}} + 2.499$
0.905	2.592	0.035	$y_K = -1.713x_N + 0.006x_P + 0.008x_k - 0.015x_{Ca} + 3.52x_{Mg} + 0.005x_{Fe} - 0.004x_{Mn}$ $-0.001x_{Cu} + 0.006x_{Zn} + 0.101x_{PH} - 0.081x_{\text{Organic}} - 7.565$
0.982	14.944	0.024	$y_{Ca} = -1.053x_N + 0.001x_P - 1.505x_k + 0.012x_{Ca} + 3.043x_{Mg} - 0.003x_{Fe} + 0.02x_{Mn} - 0.012x_{Cu}$ $+0.02x_{Zn} + 0.067x_{PH} - 0.071x_{\text{Organic}} + 1.772$
0.943	6.599	0.042	$y_{Mg} = -0.787x_N - 0.015x_P - 3.514x_k + 0.138x_{Ca} - 2.475x_{Mg} + 0.001x_{Fe} - 0.033x_{Cu}$ $-0.019x_{Zn} + 0.032x_{PH} + 0.063x_{\text{Organic}} - 1.834$
0.986	40.129	0.000	$y_{Fe} = -191.832x_N + 2.057x_P + 913.943x_k - 67.517x_{Ca} + 727.021x_{Mg} + 0.974x_{Fe} - 3.048x_{Mn}$ $-4.668x_{Zn} - 3.589x_{\text{Organic}} - 1013.297$
0.967	22.140	0.001	$y_{Mn} = -3.722x_P - 1468.955x_k - 912.715x_{Mg} + 0.908x_{Fe} - 0.411x_{Mn} - 3.922x_{Cu} + 8.048x_{PH}$ $+7.384x_{\text{Organic}} - 1280.681$
0.989	34.617	0.002	$y_{Cu} = 3.914x_N + 0.145x_P + 47.765x_k - 1.752x_{Ca} + 17.439x_{Mg} - 0.151x_{Mn} + 0.351x_{Cu}$ $+0.256x_{Zn} - 0.477x_{PH} + 0.208x_{\text{Organic}} - 29.617$
0.983	31.801	0.001	$y_{Zn} = -0.212x_P - 79.428x_k - 1.515x_{Ca} - 68.405x_{Mg} + 0.077x_{Fe} - 0.07x_{Mn} - 0.245x_{Cu}$ $+0.259x_{PH} + 0.576x_{\text{Organic}} - 92.364$

Let there be a linear relationship between the variable Y and the variable

$$X_1, X_2, \cdots, X_p:$$
$$Y = b_0 + b_0X_1 + \cdots + b_pX_p + \varepsilon \tag{1}$$

Among them, $b_0, b_1, \cdots, b_p (p \geq 2)$ is an unknown parameter, if $(x_{i1}, x_{i2}, \cdots, x_{ip}, y_i), (i = 1,2, \cdots, n)$ is a set of $n, (n > p + 1)$ independent observations of $(X_1, X_2, \cdots, X_p, Y)$, the multiple linear regression model can be expressed as:

$$y_i = b_0 + b_0x_{i1} + \cdots + b_px_{ip} + \varepsilon_i \quad (i = 1,2, \cdots, n) \tag{2}$$

Among them each ε_i is independent.

The data of the Misty breeds' soil and leaves was initially standardized, and then constructed separately multiple linear regression equations between the nutrients N, P, K, Ca, Mg, Fe, Mn, Cu, Zn in the leaves and the nutrients N, P, K, Ca, Mg, Fe, Mn, Cu, Zn, PH and organic matter in the soils, as shown in **Table 3**.

From **Table 3**, we can conclude the multiple linear equation expressions between the leaf nutrient elements and the soil nutrient content of the Misty Blueberry variety. The larger the coefficient before the different elements is, the greater the influence degree of this type's nutrient elements is, the multiple linear regression equation obtained in this study is used for the subsequent construction of the principal component regression analysis function.

Principal Component Regression Analysis

Based on constructing the multivariate linear regression equations between the leaves and soil elements of different blueberry varieties, this study performs principal component regression analysis of the different leaf inner elements. This method can ensure the extraction of different independent variables (extract information), use the contribution rate to eliminate the related influence between the respective variables.

Leaf internal correlation coefficient matrix of the Misty breed

The data is standardized by using the $Z - score$ algorithm, and then the normalized data is used to calculate the correlation coefficient matrix, the coefficient matrix can be used to detect whether there is a problem of multiple collinearity features between the leaves' nutrient elements.

$Z - score$ standardized formula is as follows:

$$Z = \frac{x - \mu}{\sigma} \tag{3}$$

x is the original data, μ is the mean of all data, and σ is the standard deviation.

The normalized correlation coefficient matrix of the Misty breed leaves' inner elements is shown in equation (4).

$$R = \begin{bmatrix} 1.000 & & & & & & & & \\ .382 & 1.000 & & & & & & & \\ -.698 & -.022 & 1.000 & & & & & & \\ -.086 & -.518 & -.242 & 1.000 & & & & & \\ .767 & .650 & -.537 & -.385 & 1.000 & & & & \\ .633 & .729 & -.326 & -.623 & .946 & 1.000 & & & \\ -.025 & -.595 & -.353 & .725 & -.415 & -.564 & 1.000 & & \\ .550 & -.275 & -.377 & .376 & -.059 & -.182 & .630 & 1.000 & \\ -.681 & -.686 & .422 & .549 & -.969 & -.992 & .476 & .123 & 1.000 \end{bmatrix} \tag{4}$$

From the correlation coefficient matrix, only the correlation coefficients of Mg-Fe, Mg-Zn, and Fe-Zn were significant, and other internal nutrient elements did not reach a significant state, therefore, the situation of multicollinearity was not considered in this study,

Table 4. Characteristic Values and Cumulative Contribution of Leaves Nutrients in Misty Varieties

Element	Starting feature value			Extract square and load		
	total	mutated %	accumulate %	total	mutated %	accumulate %
1	4.801	53.341	53.341	**4.801**	53.341	**53.341**
2	2.657	29.520	82.861	**2.657**	29.520	**82.861**
3	.648	7.196	90.057			
4	.435	4.833	94.890			
5	.253	2.811	97.701			
6	.188	2.084	99.785			
7	.013	.146	99.931			
8	.005	.055	99.986			
9	.001	.014	100.000			

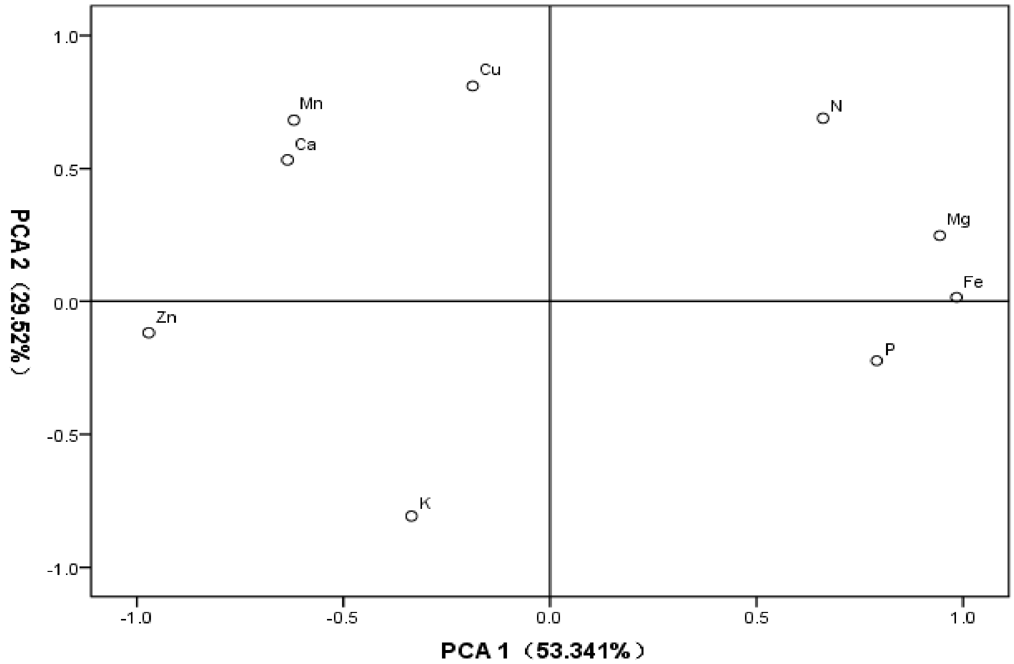

Fig. 1. The principal component (PCA) structure diagram of the Misty breed

and applied component analysis solving the multiple collinearity problem of individual elements. At the same time, it is known that in the Misty breed, N has obvious antagonistic effects with Ca, Mn and Zn, Ca and Mg are antagonistic, but Ca and Cu are synergistic.

Principal component analysis of the Misty breed

In this study, Principal Component Analysis (PCA) was used to solve the problem of multicollinearity between individual nutrient elements in the Misty breed's leaves, the cumulative contribution rate of the eigenvalues and the factor load matrix were used to determine the principal components' number of the target, and then the principal component regression equation leaves' elements were constructed in order to establish multiple regression equations between the soil and leaves of the Misty breed. The eigenvalues and cumulative contribution rates of the Misty breed's leaves are shown in **Table 4**.

Table 4 showed that the eigenvalues of the principal component analysis and cumulative contribution rates of all nutrient elements in the leaves, the total of the initial feature values in one column (second column) represented the eigenvalue corresponding to each component, and usually the principal component were used as the main components that the feature value was greater than 1, such as component 1 (feature value 4.801) and component 2 (feature value 2.657); while component 1 and component 2 cumulative contribution rate was 82.861% (>80%), indicated that the component total could explain the variance of 82.861% for all variables, the other seven components showed relatively little variable information, so they were discarded and principal component (PCA) diagram were shown in **Fig. 1**.

The principal component (PCA) diagram was actually a representation of the principal component element coefficient matrix in two or three-dimensional coordinates, thus, component 1 and component 2 were selected as the main components of the nutrient components in the Misty breed' leaves among the nine components, and calculated the coefficient matrix of the two principal component elements, as shown in equation (5).

$$R = \begin{bmatrix} .661 & .791 & -.335 & -.635 & .944 & .984 & -.620 & -.187 & -.971 \\ .689 & -.223 & -.808 & .532 & .248 & .015 & .682 & .810 & -.118 \end{bmatrix} \quad (5)$$

The two columns of the matrix represented the load vectors w_{11} and w_{12} of the two principal components respectively:

$$w_{11} = (0.661, 0.791, -0.335, -0.635, 0.944, 0.984, -0.620, -0.187, -0.971$$

and

$$w_{12} = (0.689, -0.223, -0.808, 0.532, 0.248, 0.015, 0.682, 0.810, -0.118)$$

Using the load vectors of the two principal components and the eigenvalues of each component obtained the coefficient vector of each variable in each principal component, the calculation method was shown in formula (6).

$$a_{ij} = \frac{w_{ij}}{\sqrt{\lambda}} \quad (6)$$

a_{ij} represented the coefficients of different independent variables in each principal component;

w_{ij} represented the load of different independent variables in each principal component; $\sqrt{\lambda}$ represented the eigenvalue square root of the arithmetic in each principal component. After calculation, the coefficient vectors a_{11} and a_{12} of each nutrient element can be obtained:

$$a_{11} = (0.302, 0.361, -0.153, -0.289, 0.431, 0.449, -0.283, -0.085, -0.443)$$

and

$$a_{12} = (0.423, -0.137, -0.496, 0.326, 0.152, 0.009, 0.418, 0.497, -0.724)$$

Coefficients' cumulative sum of squared on the vectors a_{11} and a_{12} were 0.99938 and 1.519084, respectively, all around the value 1. This indicated that the selected two principal components had practical significance. Therefore, the above coefficient vectors could be written as two principal component characteristic function equations, as shown in equations (a) and (b).

$$F_{11} = 0.302y_N + 0.361y_P - 0.153y_K - 0.289y_{Ca} + 0.431y_{Mg} + 0.449y_{Fe} - 0.283y_{Mn} - 0.085y_{Cu} - 0.443y_{Zn} \quad (a)$$

$$F_{12} = 0.423y_N - 0.137y_P - 0.496y_K + 0.326y_{Ca} + 0.152y_{Mg} + 0.009y_{Fe} + 0.418y_{Mn} + 0.497y_{Cu} - 0.724y_{Zn} \quad (b)$$

Among them, $y_N \sim y_{Zn}$ represented respectively the original variables of the nutrient elements in the leaves, we could know that the ratio of N, P, K, and Zn elements were relatively large from the principal component equations F_{11} and F_{12}, so the Misty Blueberry variety could be used for constructing the regression equation in the soil-leaf principal component.

Construction of Principal Component Regression Model in Misty Breed

Before constructing the principal component regression equation, we needed to consider the variance contribution rate of each principal component as the weight of each principal component, and combine multiple principal component function equations. The calculation method was as shown in formula (7):

$$F = \frac{P_1 F_{11} + P_2 F_{12}}{P_{Accumulation}} \quad (7)$$

Among them, P_1 and P_2 were the respective contribution rates of the two principal components, and $P_{\text{Accumulate}}$ was the cumulative contribution rate of the two principal components. The comprehensive statistical function of the Misty breed $F_{\text{Mistti}} = \frac{0.53341F_{11} + 0.2952F_{12}}{0.82861} = 0.644F_{11} + 0.356F_{12}$, as shown in equation (c).

$$F_{Mistti} = 1.351y_N + 0.568y_P - 1.144y_K - 0.102y_{Ca} + 1.192y_{Mg} + 0.998y_{Fe} + 0.061y_{Mn} + 0.624y_{Cu} - 2.151y_{Zn} \quad (c)$$

The multiple linear regression equation expressions between the leaf and soil of Misty Breed would be substituted in the comprehensive statistical function equation (c), the functional expression of the principal component regression equation between the leaf and the soil of Misty Breed was finally obtained, was showed in equation (d).

$$(F_{Mistti} = -190.342x_N + 2.3589x_P + 1019.482x_K - 64.441x_{Ca} + 824.554x_{Mg} + 0.8553x_{Fe} - 3.143x_{Mn} + 0.461x_{Cu} - 4.525x_{Zn} - 0.334x_{PH} - 4.075x_{Organic} - 864.53) \quad (d)$$

Among them, $x_N \sim x_{Organic}$ are the starting variables for each nutrient element in the soil of the Missy cultivar. From the function expression (d), the larger the absolute value of the original variable's correlation coefficient in the soil was, the greater the relative influence degree between the integrated principal

Table 5. Characteristics and Cumulative Contribution of Leaves Nutrients in five blueberry varieties

Breeds	element	Starting feature value			Extract square and load		
		total	mutated %	accumulate %	total	mutated %	accumulate %
O'Neal	1	5.303	58.922	58.922	**5.303**	58.922	**58.922**
	2	1.880	20.886	79.808	**1.880**	20.886	**79.808**
Duke	1	3.999	44.431	44.431	**3.999**	44.431	**44.431**
	2	2.091	23.237	67.668	**2.091**	23.237	**67.668**
	3	1.565	17.393	85.060	**1.565**	17.393	**85.060**
Legacy	1	3.286	36.517	36.517	**3.286**	36.517	**36.517**
	2	2.890	32.111	68.628	**2.890**	32.111	**68.628**
	3	1.186	13.175	81.802	**1.186**	13.175	**81.802**
Bluecrop	1	4.122	45.804	45.804	**4.122**	45.804	**45.804**
	2	2.334	25.936	71.740	**2.334**	25.936	**71.740**
	3	1.399	15.550	87.290	**1.399**	15.550	**87.290**
Northland	1	4.877	54.194	54.194	**4.877**	54.194	**54.194**
	2	1.780	19.778	73.972	**1.780**	19.778	**73.972**
	3	1.272	14.136	88.108	**1.272**	14.136	**88.108**

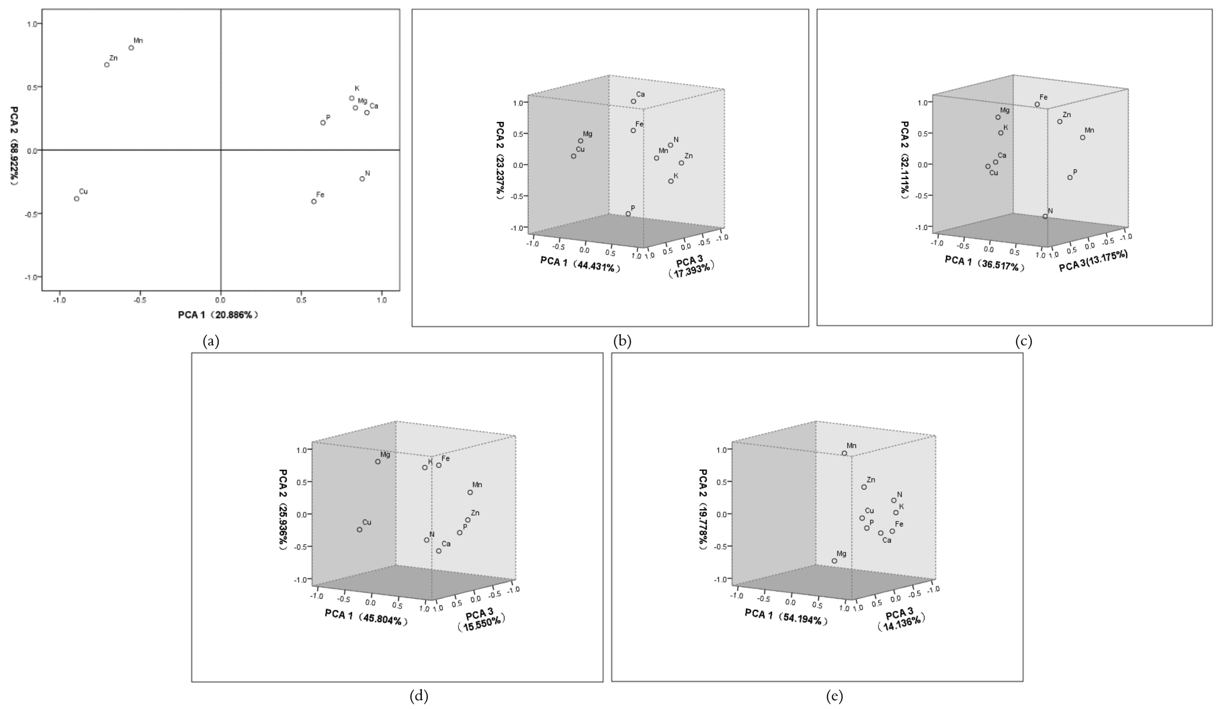

Fig. 2. Principal components (PCA) structure of five varieties

component function and the original variable was, the representation of the principal component of the variable was greater, such as N, K, Ca, Mg and etc.; also indicated that the elements Ca, Mn, Zn, pH, and organic matter in the soil were negatively correlated, the other elements were positively correlated, conducted analysis and integration combining antagonistic and synergy to each other in the nutrient elements' leaves. Therefore, in actual planting and fertilization planning, it was necessary to pay attention to the elements intake such as Ca and Mn.

Principal Component Regression Analysis Model of the Other 5 Blueberry Varieties

Principal component analysis of five blueberry varieties

Based on the multiple regression and principal component regression analysis method of the Misty breed, established a principal component regression analysis model in five kinds of blueberry leaves and soils including O'Neill, Duke, Legacy, Bluecrop, and Northland, in order to accurately explain the influence and related degree of nutrient elements in the corresponding leaves on soil. After calculated normalized correlation coefficient matrix of the remaining 5 blueberry varieties, it was found that there was no multicollinearity between the soil and the leaf elements, and could perform the principal component regression analysis. Then the remaining 5 blueberry varieties' characteristic values and cumulative contribution rate of nutrient elements in the leaves was shown as **Table 5**. The cumulative contribution rate of each breed was more than 70%, it could explain the information variance of all variables completely, the PCA diagrams of O'Neill, Duke, Legacy, Bluecrop, and Northland were plotted, just as **Fig. 2**.

Table 6. Load Vectors for Each Principal Component

Breeds	element	Load Vector
O'Neal	1	$w_{21}=(0.878,0.635,0.813,0.907,0.836,0.578,-0.556,-0.896,-0.709)$
	2	$w_{22}=(-0.227,0.216,0.409,0.296,0.333,-0.406,0.807,-0.384,0.673)$
Duke	1	$w_{31}=(0.607,0.376,0.946, -0.024, -0.364,0.484,0.870, -0.930,0.777)$
	2	$w_{32}=(0.274, -0.727, -0.201,0.929,0.401,0.618,0.204,0.010, -0.002)$
	3	$w_{33}=(-0.350,0.486,0.124, -0.232,0.734,0.499,0.413,0.119, -0.406)$
Legacy	1	$w_{41}=(0.484,0.792,-0.023,-0.685,-0.351,-0.191,0.792,-0.931,0.548)$
	2	$w_{42}=(-0.798, -0.177, 0.564, -0.087, 0.726, 0.827, 0.415, -0.201, 0.687)$
	3	$w_{43}=(0.286,0.042,0.784, -0.017,0.394,-0.459,-0.309,-0.156, -0.026)$
Bluecrop	1	$w_{51}=(0.654,0.283,0.439,0.711,-0.662,0.374,0.898,-0.808,0.938)$
	2	$w_{52}=(-0.298,-0.425,0.762,-0.496,0.694,0.723,0.346,-0.331,-0.059)$
	3	$w_{53}=(0.594,-0.811,0.336,0.355,0.009,-0.127,-0.210,0.286,-0.091)$
Northland	1	$w_{61}=(0.796,0.678,0.919,0.917,0.399,0.708,0.290,0.853,0.792)$
	2	$w_{62}=(0.173,-0.188,0.014,-0.245,-0.657,-0.325,0.935,0.038,0.492)$
	3	$w_{63}=(-0.452,0.094,-0.333,0.067,0.555,-0.538,0.135,0.468,0.331)$

Table 7. Principal component characteristic function equations of different varieties

Feature Function Equation
$F_{21}=0.381y_N+0.276y_P+0.353y_K+0.394y_{Ca}+0.363y_{Mg}+0.251y_{Fe}-0.241y_{Mn}-0.389y_{Cu}-0.308y_{Zn}$
$F_{22}=-0.166y_N+0.158y_P+0.298y_K+0.216y_{Ca}+0.243y_{Mg}-0.296y_{Fe}+0.589y_{Mn}-0.280y_{Cu}+0.491y_{Zn}$
$F_{31}=0.304y_N+0.188y_P+0.473y_K-0.012y_{Ca}-0.182y_{Mg}+0.242y_{Fe}+0.435y_{Mn}-0.465y_{Cu}+0.389y_{Zn}$
$F_{32}=0.189y_N-0.503y_P-0.139y_K+0.642y_{Ca}+0.277y_{Mg}+0.427y_{Fe}+0.141y_{Mn}+0.007y_{Cu}-0.001y_{Zn}$
$F_{33}=-0.279y_N+0.388y_P+0.099y_K-0.185y_{Ca}+0.587y_{Mg}+0.399y_{Fe}+0.330y_{Mn}+0.095y_{Cu}-0.325y_{Zn}$
$F_{41}=0.267y_N+0.437y_P-0.013y_K-0.378y_{Ca}-0.194y_{Mg}-0.105y_{Fe}+0.437y_{Mn}-0.514y_{Cu}+0.302y_{Zn}$
$F_{42}=-0.469y_N-0.104y_P+0.332y_K-0.051y_{Ca}+0.427y_{Mg}+0.486y_{Fe}+0.244y_{Mn}-0.118y_{Cu}+0.404y_{Zn}$
$F_{43}=0.263y_N+0.039y_P+0.719y_K-0.016y_{Ca}+0.362y_{Mg}-0.421y_{Fe}-0.284y_{Mn}-0.143y_{Cu}-0.024y_{Zn}$
$F_{51}=0.322y_N+0.139y_P+0.216y_K+0.350y_{Ca}-0.326y_{Mg}+0.184y_{Fe}+0.442y_{Mn}-0.398y_{Cu}+0.462y_{Zn}$
$F_{52}=-0.195y_N-0.278y_P+0.499y_K-0.325y_{Ca}+0.454y_{Mg}+0.473y_{Fe}+0.226y_{Mn}-0.217y_{Cu}-0.039y_{Zn}$
$F_{53}=0.502y_N-0.686y_P+0.284y_K+0.3y_{Ca}+0.008y_{Mg}-0.107y_{Fe}-0.178y_{Mn}+0.242y_{Cu}-0.077y_{Zn}$
$F_{61}=0.361y_N+0.307y_P+0.416y_K+0.415y_{Ca}+0.181y_{Mg}+0.321y_{Fe}+0.131y_{Mn}+0.386y_{Cu}+0.359y_{Zn}$
$F_{62}=0.129y_N-0.141y_P+0.01y_K-0.184y_{Ca}-0.493y_{Mg}-0.244y_{Fe}+0.701y_{Mn}+0.028y_{Cu}+0.369y_{Zn}$
$F_{63}=-0.401y_N+0.083y_P-0.295y_K+0.059y_{Ca}+0.492y_{Mg}-0.477y_{Fe}+0.119y_{Mn}+0.415y_{Cu}+0.293y_{Zn}$

Table 8. Principal Component Regression Prediction Function Expressions of Five Varieties

Feature function equation
$F_{\text{O'Neal}}=-36.71x_N-1.531x_P-605.365x_K+0.066x_{Ca}-389.262x_{Mg}+0.386x_{Fe}-1.545x_{Mn}-1.513x_{Cu}-1.17x_{Zn}+3.242x_{PH}+2.972x_{\text{Organic}}-477.35$
$F_{\text{Duke}}=25.208x_N+1.724x_P-408.599x_K-158.597x_{Ca}-14.904x_{Mg}+0.715x_{Fe}+1.119x_{Mn}-1.407x_{Cu}+4.097x_{Zn}-29.486x_{PH}+19.083x_{\text{Organic}}-1228.12$
$F_{\text{Legacy}}=140.608x_N-1.566x_P+5.541x_K-47.022x_{Ca}+704.681x_{Mg}-0.008x_{Fe}+0.973x_{Mn}-1.451x_{Cu}+2.773x_{Zn}+8.984x_{PH}+4.231x_{\text{Organic}}-127.004$
$F_{\text{Bluecrop}}=-8.965x_N+295.86x_P+237.027x_K+15.506x_{Ca}-312.26x_{Mg}+0.693x_{Fe}+0.961x_{Mn}-1.326x_{Cu}+20.405x_{Zn}+0.43x_{PH}-26.223x_{\text{Organic}}-466.281$
$F_{\text{Northland}}=-0.177x_N-0.045x_P-4.952x_K+24.61x_{Ca}+49.366x_{Mg}+0.051x_{Fe}-0.11x_{Mn}-1.207x_{Cu}+0.562x_{Zn}+1.642x_{PH}-3.417x_{\text{Organic}}-1.503$

The PCA diagram was used to calculate the load vector for each principal component of each variety, as shown in **Table 6**. According to the load vector of the principal components and the eigenvalues of each component found the coefficient vector of each variable in each principal component, and finally obtained the principal component characteristic function equations of different varieties, just as **Table 7**. From the table of principal component equations, it could be known that the proportion of N, P, Mg, Fe, and Mn elements were relatively large, therefore, it could perform the construction of the principal component regression equation between soil and leaves.

Construction of principal component regression model for five kinds of blueberry varieties

Based on the contribution values of different principal components, the multiple linear regression equation expressions between five leaf-soil varieties of five blueberry varieties were substituted into an integrated statistical function, and finally the principal component regression prediction function expressions between leaves and soil were obtained, as shown in **Table 8**.

From the prediction function, it was found that the trends in the N, P, K, Ca, Fe, Mn, Cu, and PH elements of O'Neill and Northland were consistent, the proportion of K, Mg, Ca, and organic matter were relatively large. This might be due to the inconsistent effect of soil fertility after large-scale improvement of experimental plots; the proportions of N, P, K, Ca, Mg, and organic matter of Duke, Legacy and Bluecrop breeds were relatively large, the remaining elements were more balanced with the corresponding to proportions in O'Neill and Bluecrop.

CONCLUSION

(1) The nutrients of blueberry leaf growth depended largely on the content of soil elements. According to the MLR regression equation of different blueberry varieties, the N and Mg elements in the soil of 6 blueberry varieties had a greater effect on the N element in the leaves. Therefore, in the actual fertilization, the intake of such fertilizer should be considered, and the intake ratio of Ca, Mn and Zn elements should be ensured. In addition, the correlations between, $\mathrm{Soil}_{\mathrm{K,Mg}} \rightarrow Leaf_P$, $Soil_{\mathrm{N,Mg}} \rightarrow Leaf_K$, $Soil_{\mathrm{N,K,Mg}} \rightarrow Leaf_{\mathrm{Ca}}$, $Soil_{\mathrm{N,K,Mg}} \rightarrow Leaf_{\mathrm{Mg}}$, $Soil_{\mathrm{N,K,Mg,Ca}} \rightarrow Leaf_{\mathrm{Fe}}$, $Soil_{\mathrm{K,Mg,Organic,PH}} \rightarrow Leaf_{\mathrm{Mn}}$ and $Soil_{\mathrm{K,Mg}} \rightarrow Leaf_{\mathrm{Zn}}$ were relatively high. Therefore, in the application of actual fertilization control, it was necessary to grasp the corresponding influence relationship between different nutrient elements. Only in this way, an efficient fertilization plan can be more effective.

(2) The leaf nutrient prediction function established by soil nutrient elements could easily, quickly and scientifically show the high or low conditions of leaf nutrient. For the Duke variety, we could appropriately increase the content of N, P, Fe or reduce K, Ca, Mg to improve the nutritional quality of the leaves; as for Bluecrop breed, we could increase appropriately the content of the P, K, Ca and other elements or reduce the content of Mg, Cu and organic matter to improve the nutritional quality of the leaves. Taken together, reasonable control of different nutrient elements' intake was one of the effective measures for actual fertilization.

ACKNOWLEDGEMENTS

We thank the laboratory of Fruit Quality of Horticultural College in Jilin Agricultural University for this work.

FUNDING

Funding provided by the National Natural Science Foundation Project of China (41601454), Jilin Province Science and Technology Development Plan Project in China (20160519014JH).

AVAILABILITY OF DATA AND MATERIALS

We declared that materials described in the manuscript, including all relevant raw data, will be freely available to any scientist wishing to use them for non-commercial purposes, without breaching participant confidentiality.

AUTHORS' CONTRIBUTIONS

LJ gived the research ideas and application models. WM-yue and ZM sorted out datas and calculated model. LJ and CG drafted the first version of the manuscript. WL collected the data acquisition. All authors have read and approved the final manuscript.

REFERENCES

Afshin M, Hiua D, Loghman A, Aram F (2014) Comparison of QSAR models based on combinations of genetic algorithm, stepwise multiple linear regression, and artificial neural network methods to predict K d of some derivatives of aromatic sulfonamides as carbonic anhydrase II inhibitors. Sci. Russian Journal of Bioorganic Chemistry, 40: 61–75.

Istek N, Gurbuz O (2017) Investigation of the impact of blueberries on metabolic factors influencing health. Sci. Journal of Functional Foods, 38: 298–307.

Lei L, Yang H, Chen J, Zhao M, Chen B, Zhou J (2017) Development and utilization of blueberry pomace: a review. China Brewing in Chinese, 36(10): 17-22.

Li Z (2017) Effect of Bio-organic Fertilizer on Soil Microbial Properties and Nutrients in Blueberry Root Zone. Research of Soil and Water Conservation in Chinese, 24(2): 36-42.

Mannozzi C, Tylewicz U, Chinnici F, Siroli L, Rocculi P, Dalla Rosa M, Romani S (2018) Effects of chitosan based coatings enriched with procyanidin by-product on quality of fresh blueberries during storage. Sci. Food Chemistry, 251: 18–24.

Mir JI, Ahmed N, Singh DB, Padder BA, Shafi W, Zaffer S, Hamid A, Bhat HA (2017) Diversity evaluation of fruit quality of apple (Malus × domestica Borkh.) germplasm through cluster and principal component analysis. Sci. Indian Journal of Plant Physiology, 22: 221–226.

Qu H, Drummond F (2018) Simulation-based modeling of wild blueberry pollination. Sci. Computers and Electronics in Agriculture, 144: 94–101.

Riccardi M, Mele G, Pulvento C, Lavini A, d'Andria R, Jacobsen S-E (2014) Non-destructive evaluation of chlorophyll content in quinoa and amaranth leaves by simple and multiple regression analysis of RGB image components. Sci. Photosynthesis Research, 120: 263–272.

Sakhanokho HF, Rinehart TA, Stringer SJ, Islam-Faridi MN, Pounders CT (2018) Variation in nuclear DNA content and chromosome numbers in blueberry. Sci. Scientia Horticulturae, 233: 108–113.

Shen H (2010) Influence of soil Microelement on Plant. Forestry Science and Technology Information in Chinese, 42(4): 12-14.

Shi Y, Wang X, Li Y (2017) Study on Causes Analysis and Preliminary Reform of Low-yield Blueberry Farm. South China Agriculture in Chinese, 11(20): 27-28.

Wu L (2016) Thirty-five Years of Research and Industry Development of Blueberry in China. Journal of Jilin Agricultural University in chinese, 38(1): 1-11.