

The Multivariate Analysis of Foliar Fertilization by Using Random Forest for Improving Alliin Concentration of Garlic (*Allium sativum* L.)

Qiang Shi¹, Xuxin Liu^{1,*}, Tursunay Dilxat¹, Jun Zhang¹, Qing Hao²

¹Agricultural Research Institute, Xinjiang Agricultural Vocational Technical College, Changji, China

²Institute of Horticulture Crops, Xinjiang Academy of Agricultural Sciences, Urumqi, China

Email address:

lx529@hotmail.com (Xuxin Liu)

*Corresponding author

To cite this article:

Qiang Shi, Xuxin Liu, Tursunay Dilxat, Jun Zhang, Qing Hao. The Multivariate Analysis of Foliar Fertilization by Using Random Forest for Improving Alliin Concentration of Garlic (*Allium sativum* L.). *International Journal of Applied Agricultural Sciences*.

Vol. 4, No. 3, 2018, pp. 52-59. doi: 10.11648/j.ijaas.20180403.11

Received: May 7, 2018; Accepted: May 30, 2018; Published: July 26, 2018

Abstract: This study was for analyzing significance and optimum proportion of components in foliar fertilizer, which aimed on improving alliin concentration of garlic (*Allium sativum* L.) in field condition. The Study with designed experiments implemented with uniform experimental design method U_{30} (30^{13}) ($D=0.1926$). Alliin concentration from garlic clove was measured by HPLC, the analysis of data set engaged by Random Forest for components significance test and optimal formula calculation. The results shown that the RF is an appropriate choice for this kind of study; the top 3 important components of foliar fertilizer were: Glutathione, Ethanethiol and Mg^{2+} ; the application of foliar fertilizer with optimal combination of components had significant effects on improving alliin concentration in garlic clove; there was synergistic interaction between components, combination of different elements can get much better effects. The maximal data of alliin concentration in garlic clove from experiment results obtained 3.5% (mg/g) of fresh weight far more than control 1.5%.

Keywords: Garlic, Alliin, Random Forest, HPLC, Multivariate Analysis

1. Introduction

Garlic (*Allium sativum* L.) is an important crop with dramatically increasing world popularity (<http://www.fao.org>). It is estimated to have been cultivated for over 5, 000 years, and have been used as food, condiment and medicine by many cultures in Asia and the Mediterranean region [1]. The major active principles in crude garlic extracts contain numerous pharmacologically active substances, including organosulfur compounds with varying stability and biological activity [2]. Such as antimicrobial and anticancer properties, as well as effects on cardiovascular diseases and diabetes have been proved [3], which also inhibited the proliferation of cancer cells, and some of them induced apoptosis in tumor cells of different tissue origin [4-6], garlic also can reduce virus infection [7].

Alliin (S-2-propenyl-L-cysteine sulfoxide) is the predominant form of organosulfur found in garlic [8], which

is the primary odorless, sulfur-containing amino acid, and also is a precursor of allicin [9]. Alliin can undergo an enzymatic reaction to produce allicin when garlic tissue is injured, allicin is mother-nature's insecticide with typical odor and shows therapeutic properties [10, 11], allicin is a highly unstable molecule, rapidly transformed into a variety of small molecular components [12]. Allicin is not present in intact garlic and hard to obtain commercially, therefore, garlic possesses high therapeutic and pharmacological properties, which activities is attributed to alliin, the main active principle localized in garlic cloves [13].

From the viewpoint of consumers and pharmaceutical industries, there is a high interest to increase the concentration of bioactive compounds, such as s alliin in garlic [14]. The impact of S fertilization on the alliin concentration of garlic was investigated under controlled greenhouse conditions and field conditions supplied in soil as base fertilization, which best reflected dose/effect relationships, and the alliin

concentration in cloves increased with S fertilization application significantly [14, 15]. For cysteine sulfoxides are synthesized in the leaves and translocated immediately into the cloves [16], it is possible for foliar fertilizer supplied to benefit alliin synthesis in leaves for garlic, and finally increase the alliin concentration in cloves. Until now, there was no report for effects on alliin concentration by using foliar fertilizer on garlic. Thereby, the aim of this trial to investigate the influence of some substance nutrients and their proportion on effecting concentration of alliin for garlic, the trial was carried out under field condition and data analysis conducted by Random forest (RF) that are recognized as one of the powerful machine learning method.

2. Materials and Methods

2.1. Experimental Design

From 2015 to 2017, a multi-factorial field trial was conducted at the experimental station of the Xinjiang Agricultural Vocational Technical College experimental fields in Changji city, Xinjiang, China (E 87.313727°, N 44.012981°). The climate is temperate and characterized by frequent changes in temperature, humidity, and winds. The soil type is Cambisol with loamy sand soil texture (67% clay and 10% sand). China white garlic was grown, cloves of garlic were planted on March 25, with an interspace of 15 cm and a distance between rows of 25cm, when garlic grown at 4 foliar, irrigation was began and carried out one times per 10-15 days. When garlic was bolting, the garlic's sprout would be removed immediately.

The field trial was performed in a uniform experimental design, with 6 factors and 30 levels total, U30 (30¹³)

(D=0.1621). Each treatment had 3 replicates, and every factor had 10 levels. The first application of foliar fertilizer was carried out 8 weeks after planting, when foliar were developed and cloves began growth. After first fertilization, there were two sprayings every two weeks. Samples were taken every 2 weeks until withering of foliar, when garlic were suitable for harvest, 6-8 plants were harvested per plot, at least 5 cloves and foliar were sampled every plot. The plant material was immediately shock-frozen in liquid N and subsequently stored in freezer at -80 degree Celsius. At the final stage of harvest, only cloves were collected and sampled.

2.2. Chemicals Choosing and Solution Mixing of Foliar Fertilizer

We chose 6 chemicals as components for foliar fertilizer, glutathione (Sigma product), cysteine (Sigma product), ethanethiol (CAS:75-08-1, product of HaoChem, <http://www.haochem.com/en/index.asp>), carbamide (product of Xinjiang Chemicals, 46% N), MnSO₄, Fe chelate (product of Shanghai Yuanye Co.), abscisic acid (ABA, product of Shanghai Yuanye Co.). Ethanethiol with its special characteristic (the water solubility of ethanethiol reaches amount of 1.5g/100mL (0.241mol/L) at room temperature), has strong alliaceous odor, we restricted the maximum amount of ethanethiol, the amount of ethanethiol reaches 0.0625g/100ml (0.01mol/L) used in this trial, which could dramatically reduce the odor of ethanethiol. All that chemicals made solution mixture of 100ml with the concentrations planed (Table 1). Chemicals mixture spraying was operated from the beginning of cloves growing, total three times spraying of foliar fertilizers before final harvest. The spray time was chosen at toward evening of sunny day.

Table 1. The concentration levels of elements in foliar fertilizer (mmol/L).

Factor Level	Ethanethiol	L (+)-Cysteine	Glutathione	ABA	Mn ²⁺	N (Carbamide)	Fe ²⁺
	X1	X2	X3	X4	X5	X6	X7
1	0	0	0	0	0	0	0
2	1.2	1	1.2	0.004	1.2	10	0.5
3	2.4	2	2.4	0.008	2.4	20	1
4	3.6	3	3.6	0.012	3.6	30	1.5
5	4.8	4	4.8	0.016	4.8	40	2
6	6	5	6	0.02	6	50	2.5
7	7.2	6	7.2	0.024	7.2	60	3
8	8.4	7	8.4	0.028	8.4	70	3.5
9	9.6	8	9.6	0.032	9.6	80	4
10	10.8	9	10.8	0.036	10.8	90	4.5

2.3. Instruments Choice and Analysis

The HPLC (high-performance liquid chromatography) system used for the study consisted of a LC10A series (Shimadzu, Japan), with the following specifications: a quaternary pump, an auto-sampler, a column heater and a diode-array detector. The data treatment was performed by the Re-analysis software (Shimadzu, Japan). The column used Carbon column (250 3 mm, Kromasil, AkzoNobel). The condition of HPLC was optimized according to the former reports [11]. The mobile phase contained methanol and H₂O at

ratio of 2:98, and added NaH₂PO₄ (3.6g/L) for protonation, was adjusted to PH 6.8, the acidification of the mobile phase is often suggested in HPLC, for reason the protonation state of the compounds affects their retention. The flow rate was 0.5ml per minute, the column room temperature was set 40 degree Celsius.

Garlic extract was prepared from fresh cloves (harvested and put in liquid nitrogen, stored at -80 degree Celsius), 20g of cloves were carefully peeled and boiled in 200 ml of deionized water during 15 min, then crushed and boiled for 1 minutes. The preparation was then centrifuged and filtered.

Nylon syringe filters of 0.20 μm were provided by Macherey-Nagel. The filtered solution was measured by HPLC for alliin concentration. Alliin was detected at 337 nm by ultraviolet (UV) detection. The methods for measuring alliin concentration followed the former data [11].

2.4. Data Analysis by RANDOM Forest (RF)

RF is an ensemble machine learning algorithm for classification and regression which is constructed by a multitude of full depth decision trees without pruning [17], it is a nonparametric tree-based approach that merges the ideas of adaptive nearest neighbors with bagging for effective data adaptive inference [18]. RF has been successfully employed in dealing with various biological prediction problems [19-22], such microarray data microarray data classification [23-25], SNPs (single nucleotide polymorphisms)[26], chemical toxicity prediction [27], DNA-binding proteins identification [28, 29], it is regarded as a useful alternative to capture the complex interaction effects among the many factors.

We used RF package 4.6-10 of R program [30], such strategies could effectively decrease the risk of overfitting when applied to our problem with large dimension. The RF package in R program can be downloaded from <https://cran.r-project.org/web/packages/randomForest/index.html>, where further explanation of the package is available. The RF package provides various options for design of integrated ensemble predictive models, each with the following significant features: the value of NMSE (Normalized Mean Squared Error) to test the predictive effectiveness of different models, the equation (1) not only considers the adaptability but also the validation of algorithm, which makes model more powerful in discriminating importance of inputs.

$$\text{NMSE} = \frac{\overline{(y-\hat{y})^2}}{(\overline{y-\hat{y}})^2} = \frac{\sum (y-\hat{y})^2}{\sum (\overline{y-\hat{y}})^2} \quad (1)$$

\bar{y} represents mean value, \hat{y} represents predicative value from model training by training data set for data analysis.

2.5. Feature Selection for Performance Evaluation

There are two methods to evaluate the importance of effect for components in foliar fertilizer calculated by different models, one is Increase in Mean Square Error (%IncMSE), it is constructed by permuting the values of each variable of the test set, recording the prediction and comparing it with the unper-muted test set prediction of the variable (normalized by the standard error); another is Increase in Rate of Node Purity (IncNodePurity), which measures the quality (NodePurity) of a split for every variable (node) of a tree by means of the Gini Index, the value of them represents the importance of factors we measured, a higher %IncMSE value represents a higher variable importance, a higher IncNodePurity value represents a higher variable importance, i.e. nodes are much 'purer' [31].

3. Results

The method in this trail provided a method for separation of alliin in 15 min with the mobile phase (PH 6.8), before the measuring samples, the validation test was carried out, the accuracy criteria (trueness and precision) for concentrations were checked by using standard alliin sample (Table 2 and Table 3). After the injections of samples, the data were processed, and the chromatogram obtained for the separation of alliin with the optimized method was shown in Figure 1. The major peak (5.6 min) was identified as peak of alliin for the garlic extracts. The method was able to provide a good method for separation of alliin from aqueous garlic extracts. All samples were conducted under same condition, and the final concentration of alliin was calculated by Re-analysis soft program provided by Shimadzu company products.

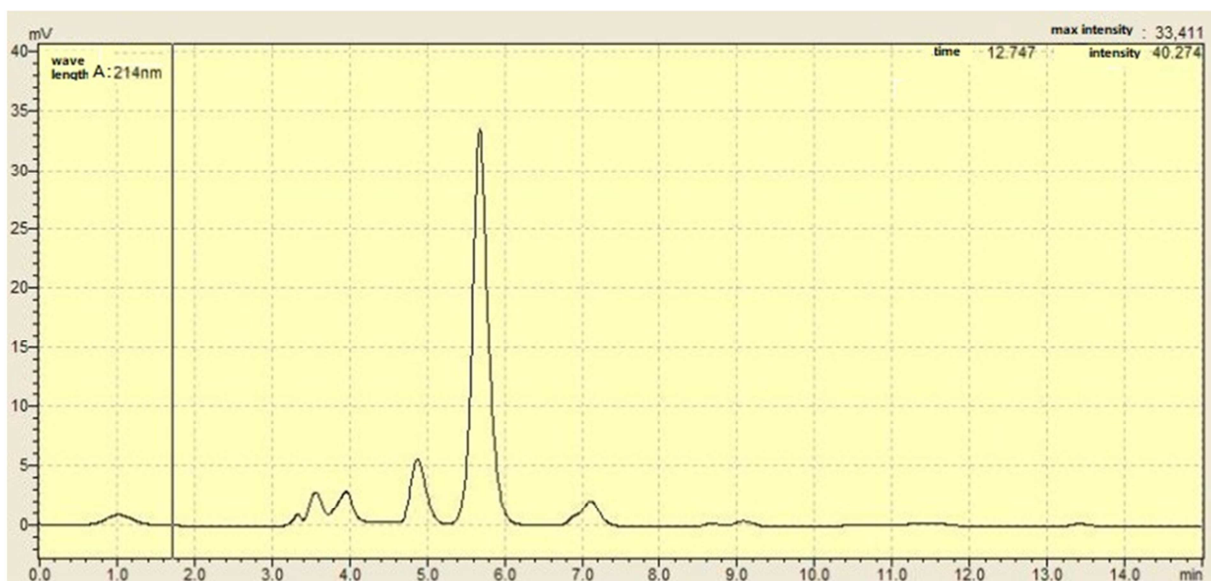


Figure 1. The analysis of Alliin concentration in garlic (*Allium sativum* L.) by HPLC. The peak of alliin appeared at 5.6 min.

Table 2. The trueness test of alliin concentration measurement with standard sample analyzed by HPLC.

Number	Retention time/min	Area	The content of Alliin
1	5.605	99.2254	0.0144803
2	5.599	99.4819	0.0144804
3	5.599	99.1856	0.0144803
4	5.602	100.3363	0.0144806
5	5.606	100.0674	0.0144805
standard deviation	5.6022	99.6563±0.5170	0.01448±0.00000
variation coefficient			0.0009472%

Table 3. The calibration for precision of alliin concentration measurement with standard sample analyzed by HPLC.

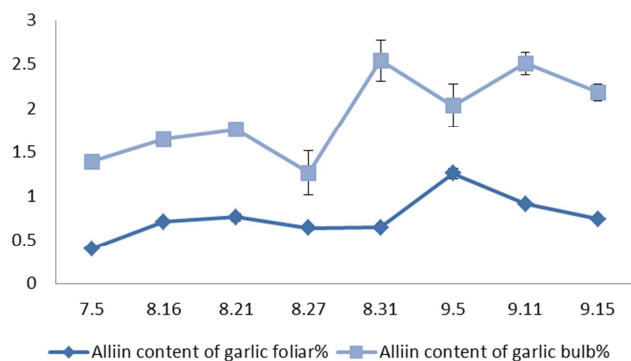
Alliin concentration in samples Mg ml ⁻¹	Standard sample concentration Mg ml ⁻¹	Number of test	Value after test Mg ml ⁻¹	Recovery rate
45	50	3	97.8±0.12	102.9
45	200	3	257.7±1.08	105.2
45	500	3	565.2±1.20	103.7

We compared with several approaches for machine learning and prediction, they were Random Forest (RF), Model-Based Boosting (MBoosting), Support Vector Machine (SVM), Recurrent Neural Networks (RNN), K-nearestneighbor (Kknn), Linear Regression (LR), all those approaches have been training and testing with our dataset, the results shown the correlation coefficient between predictive and actual

results with approaches by using the 5-fold cross validation performance for our dataset, the value of NMSE show RF was proper and optimal model for our calculation (Table 4). With some other advantages of RF, time-saving and robust performance for different parameter values, therefore, we chose the RF algorithm as proper approach for our data set analysis.

Table 4. The comparison of different models by 5-fold cross-validation method and interaction for 7 factors.

model		RF	mboost	svm	nnet	kknn	lm
The comparison for different models	NMSE	0.9423	1.245	1.1698	3.164	1.0897	12768100
The multivariate analysis for different models with 7 factors	NMSE	1.4656	1.4558	1.5339	2.6809	1.5707	1.5069

**Figure 2.** The data show of alliin concentration in garlic clove and foliar within growing period of garlic (clove growing period). The two fluctuation points at 31/8 and 11/9 in clove, one point at 5/9 in foliar; spraying application of foliar fertilizer twice at 26/8 and 1/9. Ordinate value (mg/g, fresh weight).

The application of foliar fertilizer on garlic was useful for improving the alliin concentrations in foliar and cloves of garlic. We measured the alliin concentration of garlic foliar and clove during its growing period with spraying of foliar fertilizer by using optimal formula of fertilizer mixture, which we summarized from results last year (in 2016), the concentration of alliin in foliar and clove showed several peaks subsequently after application of foliar fertilizer (Figure 2). The reaction of alliin concentration increasing in cloves and foliar was directly affected to the foliar fertilizer application, during the period of garlic growing, alliin in clove and foliar presented in uniform growth and fluctuation

tendency, but the two points at 31/8 and 11/9 for clove, one point at 5/9 for foliar, with which date the concentration of alliin shown the increasing points, before the date of peak points we applied foliar fertilizer spraying twice at 26/8 and 1/9, consequently have the alliin concentration in garlic foliar and cloves increased. The tendencies directly reflected in corresponding changes between foliar fertilizer supply and alliin concentration at clove growing stage, irrespective of the irrigation. This finding also supported the need for a sufficiently high S substance supply was enough to improve crop quality and revealed that application of S fertilizer was a suitable method to increase the alliin concentration of garlic [14].

Different components in foliar fertilizer had different effectiveness for alliin concentration. The correlation between alliin concentration in cloves and components of foliar fertilizer were evaluated by the value of %IncMSE and IncNodePurity, which results shown in Figure 3, the value of %IncMSE shown that the sequence of importance for single element, which followed the order of x3 (glutathione)>x1 (ethanethiol), others components (x6 (N), x4 (ABA), x5 (Mn²⁺), x2 (L (+)-cysteine), x7 (Fe²⁺)) were shown meaningless or had no significant for our analysis because of value of %IncMSE under zero; the value of IncNodePurity shown that the sequence for elements followed the order of x3 (glutathione)>x5 (Mn²⁺)>x1 (ethanethiol)>x4 (ABA)>x6 (N)>x2 (L (+)-cysteine)>x7 (Fe²⁺). Two indicators shown that the most important components for effecting the concentration of alliin in cloves was glutathione (x3), ethanethiol (x1) and Mn²⁺ (x5), which three components were in first three

positions, but x2 (L(+)-cysteine) did not stand in front position in results of importance analysis, because of serine is possible original substrate for S-allyl-cysteine biosynthesis in vivo [32], so, external supply of cysteine contributed less for alliin biosynthesis.

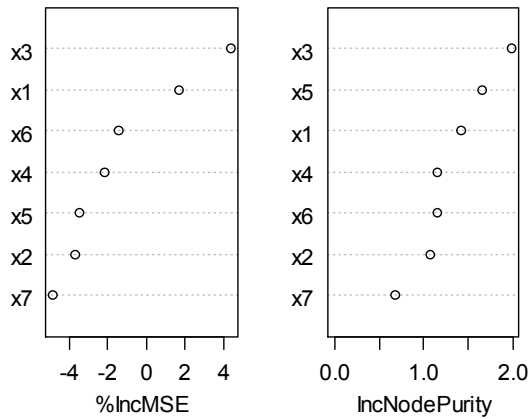


Figure 3. The significant analysis of single component. A higher %IncMSE value represents a higher variable importance. A higher IncNodePurity value represents a higher variable importance. x3 was the most importance factor in source data set.

The results of NMSE calculated by using RF for the multivariate analysis (number of factors ≤ 5) with consideration of synergistic effects shown that the combination of x1*x1*x1*x2*x3 had obvious effects in performance, and there were ten combinations (%IncMSE > 13) in front position with better performances than following others (Figure 4). The combination of x1*x1*x1*x2*x3, represent cubic x1 with x2 and x3 have synergistic effects on alliin concentration, x1 (ethanethiol) performed critical function and more sensitive than other two elements, x2 (L(+)-cysteine) and x3 (glutathione). There were almost no N (x6) appearing in top ten best combinations, it shown that N (x6) did not significantly affect the alliin concentration, which result agreed with what the former report stated [14]. The value tendency of %IncMSE for multivariate analysis shown there were better fitting effectiveness and optimized formulas with some other effective factors added, in other combinations, such as x1*x2*x3*x3*x7 (2nd place), which shown the pattern that those elements in combination for foliar fertilizer

would have synergistic effects on alliin concentration.

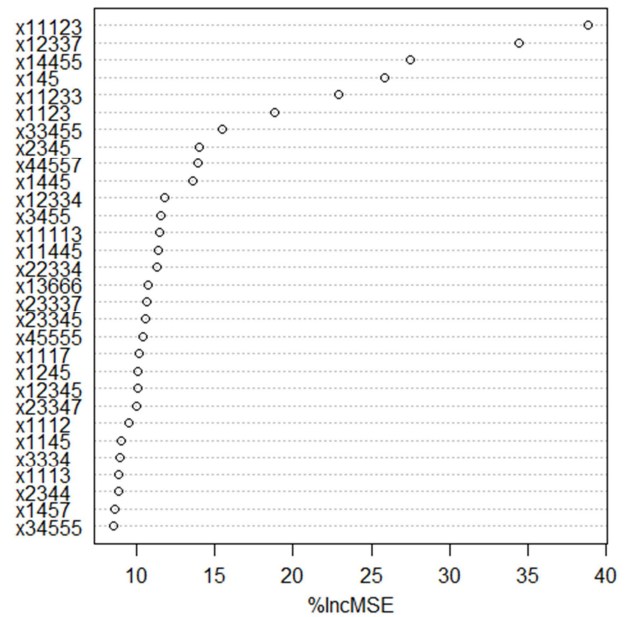


Figure 4. The multivariate analysis of importance and interaction factor effect of 5 elements. The most important types of combination is x11123, which represented factor 1 has 3 times with interaction with factor 2 and 3.

RF provided a function for calculation of optimized combination; we presented ten optimal proportions (with 6 components, no N (x6)) from the results calculated by RF model with our dataset (Table 5), which shown the alliin concentration in one-to-one relationship with each component. The concentration of alliin in garlic cloves reached 3.5~3.6% (w/w) by using ten formulas in front position, the highest concentration of alliin rise up to 36 (mg/g, fresh weight), which was double the figure of 8-15 (mg/g) from samples bought from local market, in those formulas, three components (x1, x3, x5), which concentration almost reached the highest value in our trial. The garlic with application of best formula also yielded the highest alliin concentration in foliar and cloves; it was still potential to accumulate alliin concentration in cloves by application of more S-contented substance in foliar fertilizer.

Table 5. The optimized combination of 6 factors (mmol/L).

Number	x1	x2	x3	x4	x5	x7	Content of Alliin in clove of Garlic % (mg/g, fresh weight)
1	9.6	2	9.6	0.008	9.6	2	3.5994
2	9.6	2	9.6	0.008	9.6	2.5	3.6025
3	9.6	2	9.6	0.008	9.6	3	3.6025
4	9.6	2	9.6	0.008	9.6	3.5	3.6051
5	9.6	2	9.6	0.008	9.6	4	3.6105
6	9.6	2	9.6	0.008	9.6	4.5	3.6136
7	9.6	2	10.8	0.008	9.6	4	3.5969
8	9.6	2	10.8	0.008	9.6	4.5	3.6001
9	9.6	3	9.6	0.008	6	3.5	3.5998
10	9.6	3	9.6	0.008	6	4	3.5998

The prediction results by using RF modules with various features were presented in Figure 5, the data shown relationship between foliar fertilizer components and alliin concentration could be fit by using RF model, the predictive value that fit with measured values well ($R^2=0.9294$), which might get predictive value with many condition of components as inputs. The results indicated that the combination of all features was capable of capturing more information for predictive value.

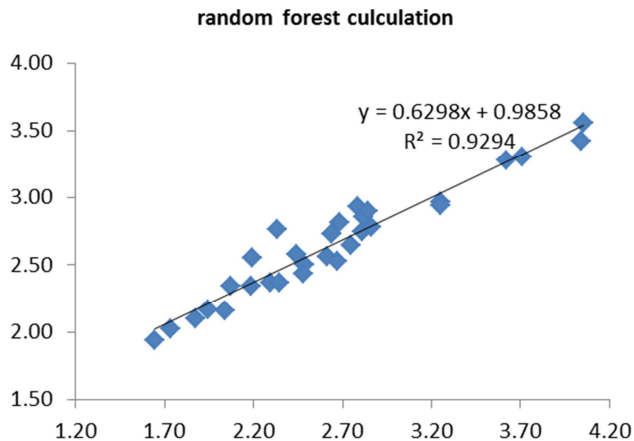


Figure 5. Random Forest model calculation for prediction of concentration of alliin in garlic clove. Ordinate value (mg/g, fresh weight). This is a generalization of a simple linear regression. A continuous metric response variable y (i.e. the shift) gets approximated through a linear combination of n multiple influence quantities and has the form: $y = a_1 * x_1 + \dots + a_n * x_n + b + \epsilon$.

4. Discussion

4.1. The Random Forest (RF) Algorithm was Capable for Importance Analysis of Fertilizer Components

In statistical prediction, the following three cross-validation methods are often used to examine a predictor for its effectiveness in practical application: independent dataset test, subsampling test, and jackknife test [33], in our calculation, 5-fold cross-validation was used to examine our model defectiveness and the prediction quality (Table 4), the results shown RF performed better than other models for our data analysis. RF was confirmed to be a well-suited alternative approach for importance analysis of fertilizer components in our study. The RF as a machine learning method could be used for predicting the best formula of foliar fertilizers for improving alliin concentration in garlic clove in the condition of fields.

4.2. Foliar Fertilization with S Content Improved Garlic Products Quality

Many reports of S content fertilizer could improve the alliin concentration in garlic cloves [13-15, 34], but S materials was often used as base fertilizer with lots of amount. For the approaches of foliar fertilizer, which could be a better way and efficiency method, because of cysteine sulfoxides (including alliin) are synthesized in the leaves [16], some S materials could be absorbed directly for alliin biosynthesis. The alliin content

represented the nutritional and phytopharmaceutical value of the garlic, there are innumerable health products, in which the alliin as a leading substance should be declared to give the consumer the chance to compare products with their phytopharmaceutical value and high quality products [13], therefore, the improving of alliin content in garlic will be considered as improving quality of garlic. We proposed an approach that foliar fertilization could have higher efficiency with lower amount of substances applying than basic fertilization.

4.3. The Components of Foliar Fertilizer Contributed to Improving Alliin Concentration with Different Degree

Glutathione (x3) contributed to improvement of alliin concentration directly and efficiently, the formulas of foliar fertilizer with the highest concentration of glutathione (x3) did increase alliin concentration significantly, it was a key ingredient for alliin biosynthesis in garlic, the glutathione was also considered as a precursor for alliin very close in the course of biosynthesis [32], it could be used as a key component for foliar fertilizer. Ethanethiol (x1) was at second place in significance analysis, which could provide mercaptans for alliin biosynthesis [35], ethanethiol was a sensitive component for effecting alliin concentration from analysis results, but for its strong odor, the amount of ethanethiol should be controlled under certain level, the maximum tolerated dose was 0.0625g/100ml (0.01mol/L) in our study. Fe^{2+} (x7) contributed less in our calculation result, but Mn^{2+} (x5) would be a core role for relative enzymes in biosynthesis and function, it could play a critical role in our analysis results but not needed in large quantity. Two possible biosynthetic pathways leading to S-allyl-cysteine, which is oxidized to alliin that is the "inactive" precursor of allicin, this pathway leads from glutathione to S-allyl-cysteine [32]; in another pathway, cysteine appeared at later stage and derived from serine, and serine is one another possible substrate for S-allyl-cysteine biosynthesis.[36], so, cysteine (x6) did not stand in front line in our analysis, it was able to understand that cysteine added in foliar fertilizer maybe not sensitive for effecting alliin biosynthesis, because of cysteine derived from serine. The plant hormone ABA (x4) is considered the major chemical signal involved in stomatal regulation and stress response, the external ABA source conduct the closure of stomatal guard cell [37, 38], therefore, the substance of foliar fertilizer could be as intake into space of stomatal pore and absorbed by osmotic cell for taking part in alliin biosynthesis path way. In sequence, N (x6) also played less important role in our result, but foliar growth will need certain amount of N, and N played a mediate role for different component, and contributed more for foliar growth, from former study, there were also interactions between N and S: the total S content was decreased with increasing N fertilization, it was same in onion [15, 39].

4.4. Synergistic Interaction of Components Existed in Compounds of Foliar Fertilization

Analysis results (Figure 4) presented several optimal compositions of ingredients which contributed significantly to

influence alliin concentration, there were several single substances which had leading function, such as glutathione (x3), ethanethiol (x1) and Mn^{2+} (x5), but they had remarkable difference in expressivity. The glutathione (x3) presented the performance in quantity, ethanethiol (x1) presented performance in dose effects with more sensitivity, Mn^{2+} (x5) was essential and requisite substance. The composition of $x1*x1*x1*x2*x3$ was one of the best formula, which indicated the leading role of x1 (ethanethiol) that was sensitive on improvement of alliin concentration for garlic clove, it could be conducted that x1 was dose effect in function. In some other formulas (top seven compositions) also had remarkable utility for alliin concentration improvement, the compositions of x33455 ($x3*x3*x4*x5*x5$) stand on 7th position, after which compositions did not have remarkable effectiveness in our analysis.

4.5. The Prediction of Alliin Concentration with Features of Fertilizer Components Could be Conducted by RF

RF has very wide range of applications at biological sector [19-22], which can be an approach for prediction of alliin concentration of garlic with input of quantity of substances related to alliin's precursor in biosynthesis pathway, our data (Figure 5) shown that the correlation and regression analysis among inputs of ingredients features and alliin concentration, which was tasted with significance level of $R^2=0.9294$. So, by this results, RF could be considered as a useful model for analysis of relationship between foliar fertilizer components and secondary metabolites according to somehow biological interlink between them.

5. Conclusion

From summary of our data, the highest amount of alliin concentration after foliar fertilizer application reached 3.5 % (mg/g, fresh weight), which was twice as much as control samples 1.5-1.8%. This result might indicate that the application of S-content foliar fertilization with small amount have potential for alliin accumulation. The different component had Synergistic interaction and different role for effectiveness, Glutathione was a main component and contributed directly for alliin biosynthesis, presented the performance in quantity; Ethanethiol was more sensitivity and with dose effects; Mg^{2+} played an essential and requisite role, three components stood in front position in our analysis. The composition of $x1*x1*x1*x2*x3$ was one of the best formula for foliar fertilizer, which obtained the high concentration of alliin in garlic cloves. RF was a suitable method for analysis of effectiveness and utility for foliar components, might be a useful tool for more using in agricultural aspect.

The active compound alliin could be regard as a product of phytopharmaceutical from garlic, and the price per gram alliin would be in the range of 45-50 EUR (estimated exemplary from the selling price of garlic product)[15], the profit of garlic could be enhanced by improving quality with high alliin content. Foliar fertilization with optimal formulas showed more advantages for application, and could give chance for farmer another way to improve quality of garlic product.

Funding

This work was supported by the grants from the Funding of Xinjiang Agricultural Vocational Technical College (No. XJNZYKJ201503), the funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests

The authors have declared that no competing interests exist.

References

- [1] M. Ipek, A. Ipek, S. G. Almquist, P. W. Simon, Demonstration of linkage and development of the first low-density genetic map of garlic, based on AFLP markers, *Theor Appl Genet* 110 (2) (2005) 228-36.
- [2] M. JA, Garlic: its anticarcinogenic and antitumorigenic properties, *Nutr Rev.* (54) (1996) s82-6.
- [3] H. A. Collin, Garlic and cardiovascular disease. In *Functional Foods, Cardiovascular Disease and Diabetes*, Arnoldi, A., Ed.; Woodhead Publishing, Ltd.: Cambridge, U. K. (2004) 240-260.
- [4] V. G. M Dirsch, Alex Vollmar, Angelika, Ajoene, a compound of garlic, induces apoptosis in human promyeloleukemic cells, accompanied by generation of reactive oxygen species and activation of nuclear factor kappaB, 1998.
- [5] J. Pinto, S. Lapsia, A. Shah, H. Santiago, G. Kim, Antiproliferative Effects of Garlic-Derived and Other Allium Related Compounds, 2001.
- [6] G. Sigounas, J. Hooker, A. Anagnostou, M. Steiner, S-allylmercaptocysteine inhibits cell proliferation and reduces the viability of erythroleukemia, breast, and prostate cancer cell lines, 1997.
- [7] A. Hall, A. Troupin, B. Londono-Renteria, T. M. Colpitts, Garlic Organosulfur Compounds Reduce Inflammation and Oxidative Stress during Dengue Virus Infection, *Viruses* 9 (7) (2017).
- [8] S. M. Kawakishi, Y., Sulfur chemistry of onions and inhibitory factors of the arachidonic acid cascade, In *Food Phytochemicals for Cancer Prevention I*, American Chemical Society, 1993, pp. 120-127.
- [9] H. J. T. A. T. A. B, Synthesis of the flavour precursor, alliin, in garlic tissue cultures, *Phytochemistry* 66 (2) (2005) 187-194.
- [10] M. S. Rahman, Allicin and Other Functional Active Components in Garlic: Health Benefits and Bioavailability, *International Journal of Food Properties* 10 (2) (2007) 245-268.
- [11] B. Dethier, M. Laloux, E. Hanon, K. Nott, S. Heuskin, J. P. Wathelet, Analysis of the diastereoisomers of alliin by HPLC, *Talanta* 101 (2012) 447-52.
- [12] S. A. Nasim, B. Dhir, R. Kapoor, S. Fatima, Mahmooduzzafar, A. Mujib, Alliin obtained from leaf extract of garlic grown underin situconditions possess higher therapeutic potency as analyzed in alloxan-induced diabetic rats, *Pharmaceutical Biology* 49 (4) (2011) 416-421.

- [13] E. B. S. H. E. Schnug, Influence of Nitrogen and Sulfur Fertilization on the Alliin Content of Onions and Garlic, *Journal of Plant Nutrition* 27 (10) (2004) 1827-1839.
- [14] S. H. ELKE BLOEM, AND EWALD SCHNUG, Influence of Fertilizer Practices on S-Containing Metabolites in Garlic (*Allium sativum L.*) under Field Conditions, *J. Agric. Food Chem.* 58 (2010) 10690–10696.
- [15] E. B. S. H. E. Schnug, Influence of nitrogen and sulfur fertilization on the alliin content of onions and garlic *J. Plant Nutr* 27 (2004) 1827-1839.
- [16] M. Mutsch-Eckner, Isolierung, Analytik und biologische Aktivität von Aminosäuren und Dipeptiden aus *Allium sativum L.*, ETH: Zürich; Dissertation Nr (1991) 9462.
- [17] L. Breiman, Random forests, *Mach. Learn* 45 (1) (2001) 5-32.
- [18] L. Breiman, Bagging predictors, *Mach. Learn* 24 (2) (1996) 123-140.
- [19] L. H. Wu JS, Duan XY, Ding Y, Wu HT, Prediction of DNA-binding residues in proteins from amino acid sequences using a random forest model with a hybrid feature., *Bioinformatics* 25 (2009) 30-35.
- [20] P.-A. S. Dehzangi A, Dehzangi O Using Random Forest for Protein Fold Prediction Problem: An Empirical Study, *Journal of Information Science and Engineering* 26 (2010) 1941–1956.
- [21] W. L. Liu ZP, Wang Y, Zhang XS, Chen LN., Prediction of protein-RNA binding sites by a random forest method with combined features, *Bioinformatics* 26 (2010) 1616–1622.
- [22] C. K. Kandaswamy KK, Martinetz T, Moller S, Suganthan PN., AFP-Pred: A random forest approach for predicting antifreeze proteins from sequence-derived properties, *Journal of Theoretical Biology* 270 (2011) 56–62.
- [23] M. S. M. Kohbalan Moorthy, Random forest for gene selection and microarray data classification, *Bioinformation* 7 (3) (2011) 142-146.
- [24] T. Shi, D. Seligson, A. S. Beldegrun, A. Palotie, S. Horvath, Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma, *Modern Pathology* 18 (4) (2004) 547-557.
- [25] A. Anaissi, P. J. Kennedy, M. Goyal, D. R. Catchpoole, A balanced iterative random forest for gene selection from microarray data, *BMC Bioinformatics* 14 (2013) 261.
- [26] V. Botta, G. Louppe, P. Geurts, L. Wehenkel, Exploiting SNP correlations within random forest for genome-wide association studies, *PLoS One* 9 (4) (2014) e93379.
- [27] B. A. Tuulaikhuu, H. Guasch, E. Garcia-Berthou, Examining predictors of chemical toxicity in freshwater fish using the random forest technique, *Environ Sci Pollut Res Int* (2017).
- [28] V. N. Uversky, W.-Z. Lin, J.-A. Fang, X. Xiao, K.-C. Chou, iDNA-Prot: Identification of DNA Binding Proteins Using Random Forest with Grey Model, *PLoS ONE* 6 (9) (2011) e24756.
- [29] J. Wu, H. Liu, X. Duan, Y. Ding, H. Wu, Y. Bai, X. Sun, Prediction of DNA-binding residues in proteins from amino acid sequences using a random forest model with a hybrid feature, *Bioinformatics* 25 (1) (2009) 30-35.
- [30] M. W. Andy Liaw Classification and Regression by randomForest, *R News* 2 (3) (2002) 18-22.
- [31] S. Kuhn, B. Egert, S. Neumann, C. Steinbeck, Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction, *BMC Bioinformatics* 9 (1) (2008) 400.
- [32] J. Borlinghaus, F. Albrecht, M. C. Gruhlke, I. D. Nwachukwu, A. J. Slusarenko, Allicin: chemistry and biological properties, *Molecules* 19 (8) (2014) 12591-618.
- [33] Z. C. Chou KC, Review: Prediction of protein structural classes, *Critical Reviews in Biochemistry and Molecular Biology* 30 (1995) 275–349.
- [34] G. G. M. Freeman, N., Influence of sulphate nutrition on the flavour components of garlic (*Allium sativum*) and wild Onion (*A. vineale*). *J. Sci. Fd Agric* (22) (1971) 330-334.
- [35] B. Granroth, Biosynthesis and decomposition of cysteine derivatives in onion and other *Allium* species, *Ann. Acad. Sci. Fenn. Chem.* (154) (1970) 4–71.
- [36] B. Granroth, Biosynthesis and decomposition of cysteine derivatives in onion and other *Allium* species., *Ann. Acad. Sci. Fenn. Chem.* (154) (1970) 4-71.
- [37] J. P. Comstock, Hydraulic and chemical signalling in the control of stomatal conductance and transpiration, *Journal of Experimental Botany* 53 (367) (2002) 195-200.
- [38] R. Finkelstein, Abscisic Acid synthesis and response, *Arabidopsis Book* 11 (2013) e0166.
- [39] W. M. K. Randle, D. E.; Kopsell, D. A, Sequentially reducing sulfate fertility during onion growth and development affects bulb flavor at harvest., *HortScience* 37 (1) (2002) 118-121.