Method for detecting soil total nitrogen contents based on pyrolysis and artificial olfaction

Mingwei Li[1,2], Qinghui Zhu[1,2], He Liu[1,2], Xiaomeng Xia[1,2], Dongyan Huang[1,2*]

(1. *School of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China*;
2. *Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China*)

Abstract: Soil nitrogen is an essential nutrient element for crop growth and development, and an important indicator of soil fertility characteristics. This study proposed a method based on pyrolysis and artificial olfaction to quickly and accurately determine the soil total nitrogen (STN) content. A muffle furnace was used to pyrolyze the soil samples, and ten different types of oxide semiconductor gas sensors were used to construct a sensor array to detect the soil samples' pyrolysis gas. The response curves of the sensors were tested at pyrolysis temperatures of 200°C, 300°C, 400°C, and 500°C and at pyrolysis times of 1 min, 3 min, 5 min, and 10 min to obtain the optimal pyrolysis state of the soil samples. The optimal pyrolysis temperature was 400°C, and the pyrolysis time was 3 min. The response area, maximum value, average differential coefficient, variance value, maximum gradient value, average value, and 8th-second transient value of the sensor response curve were extracted to construct an artificial olfactory feature space of 121×10×7 (121 soil samples, ten sensor numbers, seven extracted eigenvalues). Back-propagation neural network algorithm (BPNN), partial least squares regression algorithm (PLSR), and partial least squares regression combined with back-propagation neural network algorithm (PLSR-BPNN) were used to establish a prediction model of artificial olfactory feature space and STN content. Moreover, coefficient of determination (R^2), root mean square error (RMSE), and the ratio of performance to deviation (RPD) were used as the performance indicators of the prediction results. The test results showed that the R^2 of the PLSR, BPNN, and PLSR-BPNN models were 0.89033, 0.81185, and 0.92186, and the RMSE values were 0.24297, 0.37370, and 0.21781, and the RPD were 2.9964, 1.9482, and 3.3426, respectively. The model established by the PLSR-BPNN algorithm has the highest R^2 and RPD and the smallest RMSE, can achieve the accurate prediction of STN content, and therefore the model is rated as "excellent". The detection method in this study achieves a low-cost, rapid, and accurate determination of STN content, and provides a new reference for the measurement of STN.

Keywords: pyrolysis, artificial olfactory system, soil total nitrogen, gas sensor array, prediction methods
DOI: 10.25165/j.ijabe.20221503.7086

Citation: Li M W, Zhu Q H, Liu H, Xia X M, Huang D Y. Method for detecting soil total nitrogen contents based on pyrolysis and artificial olfaction. Int J Agric & Biol Eng, 2022; 15(3): 167–176.

1 Introduction

Soil nutrients supply the elements necessary for plant growth and development. Soil nitrogen is an important nutrient, and an indicator of soil fertility characteristics, and its abundance and deficiency affect crop growth, quality, and yield[1-4]. The soil nitrogen content in China is quite different. The soil total nitrogen (STN) content of cultivated land is 0.4-3.8 g/kg, however, it shows an uneven spatial distribution and unreasonable nitrogen fertilizer application[5-7]. Therefore, accurate, fast, and low-cost determination of STN content and its changes are significant for precision agriculture and natural ecology conservation.

Traditional STN measuring chemical methods include the Kjeldahl method and Dumars combustion method. Although they can accurately measure STN content and have a wide range of applications, they are time-consuming, complex to operate and generate pollution[8,9]. Spectroscopy technology is widely used for STN content determination because it is fast, non-polluting, and non-destructive[10-13]. Wang et al.[14] measured soil with different particle diameters by spectroscopy and constructed an STN content estimation model with the spectral transformation data. They concluded that the smaller the soil particle diameter, the higher the estimation accuracy of the total nitrogen (TN) content, where the model established by the support vector machine (SVM) is superior to partial least squares regression (PLSR) and stepwise multiple linear regression (SMLR). Nie et al.[15] studied the influence of drying temperatures on soil nitrogen determination by near-infrared spectroscopy. Low or extremely high drying temperatures will adversely affect soil nitrogen detection, and a drying temperature of 40°C and competitive adaptive reweighted squares-backward interval partial least squares-partial least squares (CARS-BIPLS-PLS) method is the best method to improve the soil nitrogen detection accuracy. He et al.[16] studied the influence of soil water content on soil nitrogen detection by the near-infrared sensor. When the soil moisture content is 1.03%, the negative impact on soil nitrogen detection was the smallest, and the detection accuracy was the highest. However, the spectroscopy method is affected by soil moisture, iron oxide, and soil texture[17-19].

Soil gas is caused by the balance between biological activity

Received date: 2021-09-27 **Accepted date:** 2022-03-02
Biographies: Mingwei Li, PhD, research interest: automation of agricultural machinery, Email: lmw271314@163.com; **Qinghui Zhu**, Master candidate, research interest: intelligent agricultural machinery design, Email: ahszqh@126.com; **He Liu**, PhD candidate, research interest: intelligent agricultural machinery design, Email: heliu20@mails.jlu.edu.cn; **Xiaomeng Xia**, PhD candidate, research interest: intelligent agricultural machinery design, Email: xiaxiaomeng4700@163.com.

***Corresponding author: Dongyan Huang**, PhD, Professor, research interest: conservation tillage technology and intelligent agricultural machinery equipment. Mailing address: School of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China. Tel: +86-13610712601, Fax: +86-431-85095760, Email: cchdy760829@sina.com.

and gas transfer in the soil is important for plant growth and soil formation[20]. During microbial degradation, nutrients and energy substrate supply produce many volatile gases in the soil[21-23]. Therefore, volatile gas in the soil correlates with soil nutrients, which can be used to detect soil odor through artificial olfactory technology to obtain soil nutrient content. However, volatile gas content in soil is relatively small and requires high-sensitivity gas sensors. Zhu et al.[24,25] used an artificial olfactory system (AOS) to detect soil organic matter content and constructed three models, PLSR, back-propagation neural networks (BPNN), and support vector regression (SVR), and the SVR model had the highest prediction performance with R^2 of 0.91.

Pyrolysis gas chromatography-mass spectrometry (Py-GC/MS) technology is an effective method that has emerged to study the thermal pyrolysis mechanism. The technology first places the sample in a pyrolyzer under strictly controlled conditions for heating so that the macromolecular compounds are quickly pyrolyzed into volatile small-molecule compounds. Then, the carrier gas carries the small molecule compounds into the gas chromatography/mass spectrometer (GC/MS) for separation and identification. Finally, the composition, structure, and chemical characteristics of the sample are inferred through the GC/MS characteristics[26-30]. Py-GC/MS is fast, highly sensitive, and requires a small sample volume, so it is widely used in polymer science, microbiology, bioengineering, and geochemistry[31-35]. De La Rosa et al.[36] used Py-GC/MS to characterize the effects of wildfires on soil organic matter by analyzing in detail the pyrolysis products of burned and unburned surface soil and conducting qualitative and quantitative analysis on the products of 300°C desorption and 500°C pyrolyzes. Chen et al.[37] used Py-GC-MS/MS to study the fingerprint differences between the five density components of the soil organic matter in the alpine grassland, and performed qualitative and quantitative analyses of 150 pyrolysis products. According to similar chemical properties, pyrolysis products were divided into alkyl compounds, aromatic hydrocarbons, polycyclic aromatic hydrocarbons, lignin, phenolic substances, and polysaccharides, nitrogen-containing compounds, and chitin.

Pyrolysis technology uses the thermal instability of organic matter, and thermal decomposition under anaerobic or anaerobic conditions can cause a small amount of soil to produce a large amount of pyrolysis gas. This gas can be used to achieve rapid and accurate detection of soil nutrients by artificial olfactory technology. The Py-GC-MS/MS technology can accurately, qualitatively, and quantitatively analyze the various nutrients contained in the soil. However, it also has disadvantages such as high equipment purchase cost, specialized personnel required for operation, non-specialized design for STN monitoring, and time-consuming and labor-intensive. Therefore, it is difficult to rapidly measure the TN content for many soil samples.

This study proposed an STN content detection method based on pyrolysis and AOS in response to the above problems. The pyrolysis technology was used to pyrolyze soil samples, and the pyrolysis gas was passed into the gas sensor array to obtain the sensor response curve. Seven features of the response curve were extracted, including response area, maximum value, average differential coefficient, variance value, maximum gradient value, average value, and 8th-second transient value, to construct a characteristic artificial olfactory feature space (AOFS). The STN content prediction model and AOFS were established by BPNN, PLSR, and partial least squares regression combined with back-propagation neural network algorithm (PLSR-BPNN), so that the STN content could be detected quickly, accurately, and at low cost.

2 Materials and methods

2.1 Study area and soil sampling

The study area is in Jilin Province, China (40°50'N-46°19'N, 121°38'E-131°19'E, Figure 1), which belongs to the central part of Northeast China, borders Russia and North Korea, in the geographic center of Northeast Asia on the east side of mid-latitude Eurasia. It has a temperate continental monsoon climate. The average annual temperature is between −5°C and 8.6°C, and the average annual rainfall is 400-600 mm. The soil types are mainly dark-brown soil, chernozem soil, albic soil, meadow soil, black soil, aeolian sandy soil, new soil, and paddy soil.

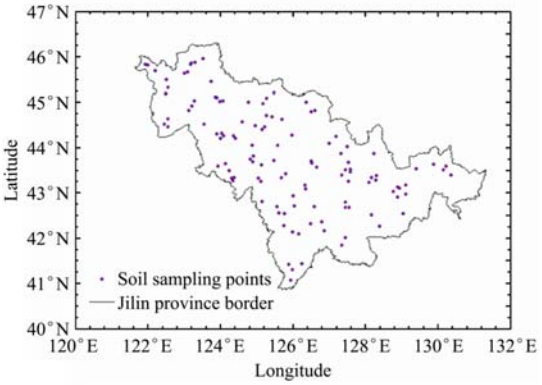
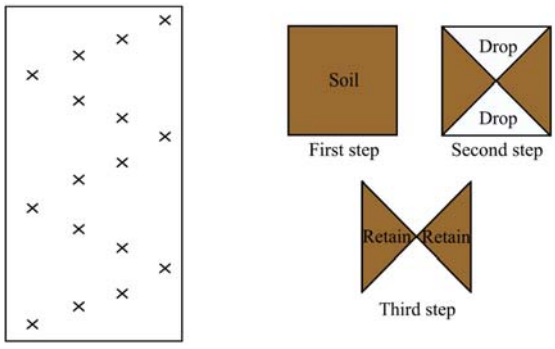


Figure 1 Study area and sampling location

In this study, 121 soil samples were collected at 121 sampling points in the study area (Figure 1). Due to the heterogeneity of the soil, there is a certain degree of variation in each sample. To preserve the actual state of the field soil samples as much as possible, 16 soil samples were collected using an S-shaped route (Figure 2a), with a sampling depth of 0-20 cm. The sampling points avoid fields, roadsides, ditch sides, special terrain, and places where fertilizers have accumulated. Each sample was mixed thoroughly, and 1 kg of soil samples per sampling point was retained according to the quartering method (Figure 2b). According to the requirements of the test, 121 soil samples were dried naturally at a temperature of 23°C, followed by crushing and processing through a 0.25 mm screen. The samples were divided into two parts, respectively used for the Kjeldahl method and detection based on pyrolysis and AOS to determine the STN content.



a. S-shaped sampling point b. Quarter method

Figure 2 Soil sample collection

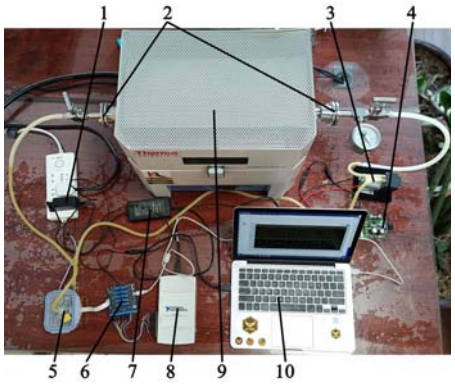
2.2 Kjeldahl Method

The Kjeldahl method was established by the Danish chemist Kedal in 1883 and is a common method for analyzing the nitrogen

content of organic compounds. The theoretical basis of the Kjeldahl method is that the nitrogen content of proteins usually accounts for about 16% (12%-19%) of their total mass[38,39]. Therefore, by measuring the nitrogen content in the substance, the total protein content in the substance can be estimated. The main principle of this method is to digest with concentrated sulfuric acid, use catalysts and warming agents to accelerate the decomposition of organic matter, convert organic nitrogen into ammonia in the solution, and finally titrate the distilled ammonia with standard acid[40-43].

2.3 Pyrolysis and artificial olfactory testing method

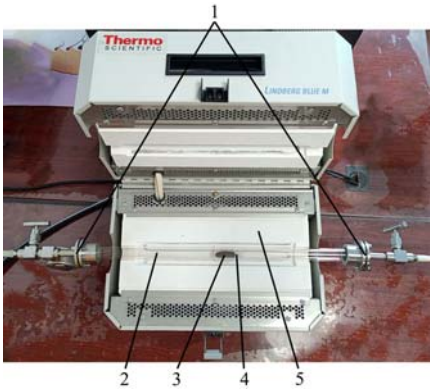
An STN detection system based on pyrolysis and AOS was designed, as shown in Figure 3. The system consists of a pyrolysis furnace, a gas sensor array (installed in a closed reaction chamber), a signal processing circuit, an NI data acquisition card, a portable computer, a PWM speed control module, a micro vacuum pump, a pyrolysis chamber, and a vacuum flange.



1. 12 V power supply 2. Vacuum flange 3. Vacuum pump 4. PWM speed control module 5. Gas sensor array 6. Signal processing circuit 7. 12 V power supply 8.NI data acquisition card 9. Pyrolysis furnace 10. Portable computer

Figure 3 Soil total nitrogen content detection system based on pyrolysis and artificial olfactory system

The quartz boat was placed in the center of the quartz tube to hold the soil sample, and the vacuum flange kept the quartz tube closed to complete the pyrolysis of the soil sample, as shown in Figure 4. The pyrolysis furnace was the Lindberg/Blue M Mini-Mite tube (Thermo Fisher Scientific). The pyrolysis furnace uses a microprocessor-based self-adjusting PID controller and was insulated with Moldatherm® materials, with a precise response temperature of 100°C-1100°C. The gas sensor array is the core component of AOS, responsible for producing a specific response to gas after soil pyrolyzing, and is the data basis for building the system. The products after soil pyrolyzing mainly include alkanes, alkenes, aromatics, nitrogen-containing compounds, fatty acids, lignin, phenolic substances, polysaccharides, and chitin[37,44]. This study selected ten oxide semiconductor gas sensors produced by Figaro (TGS826, TGS2602, TGS2610, TGS2620, TGS821, TGS2603, TGS2611, TGS823, TGS2600, and TGS2612) to form the sensor array and their specific parameters as shown in Table 1. The signal conversion circuit was used to supply power to the sensor array and perform conversion processing on the data generated by the sensor. The sensor array and the signal processing circuit were connected through an FFC cord. The data acquisition card was a multifunctional I/O device USB-6210 produced by National Instruments (NI), connected to the signal conversion circuit through a DuPont cable for sensor data acquisition. The collected data was transferred to the computer through the USB data cable and stored in the LabVIEW detection program. The PWM speed control module was used to adjust the flow rate of the micro vacuum pump, used to circulate air to the entire system.



1. Vacuum flange 2. Quartz tube 3. Soil sample 4. Quartz boat 5. Pyrolysis furnace

Figure 4 Soil samples in a pyrolysis furnace

Table 1 Gas sensor model and parameters

Senor number	Model	Monitoring gas type	Measuring range/mg·m^{-3}
S1	TGS826	Ammonia	30-300
S2	TGS2602	Air pollutants (ammonia, volatile organic compounds (VOC), hydrogen sulfide, etc.)	1-30
S3	TGS2610	Butane, LP gas	500-10 000
S4	TGS2620	Ethanol, organic solvent	50-5000
S5	TGS821	Hydrogen	100-1000
S6	TGS2603	Trimethylamine, methyl mercaptan, etc.	1-10
S7	TGS2611	Methane, natural gas	500-10 000
S8	TGS823	Ethanol	50-300
S9	TGS2600	Hydrogen, alcohol, etc.	1-30
S10	TGS2612	Methane, propane, isobutane	3000-9000

When the system was working, a quartz boat was used to hold 2 g of the soil sample, placed in the center of the quartz tube, and sealed with a vacuum flange, then, placed in a pyrolysis furnace for pyrolyzing. To obtain the best pyrolysis state of the soil sample and to improve the accuracy of data collection in the monitoring method, pyrolysis was performed at 200°C, 300°C, 400°C, 500°C for 1 min, 3 min, 5 min, and 10 min, respectively. The pyrolysis chamber, the reaction chamber, and the vacuum pump were connected by a rubber tube, forming a closed gas path. After pyrolysis, the LabVIEW detection program was started, the flange opened, and the vacuum pump flow was set to 1 L/min. The pyrolysis gas in the pyrolysis chamber entered the reaction chamber equipped with the sensor array, and the response data were collected for 60 s. After data collection, the rubber tubes at both ends of the flange were removed, and the quartz tube and the quartz boat were removed, cleaned, and dried. The flow rate of the vacuum pump was set to 3 L/min to clean the reaction chamber and the connecting pipes with clean air for 2 min. The collection of soil sample data was performed once. The above process was repeated once for the collection of other sample data.

2.4 Feature selection

Extracting appropriate features from the response curve of the sensor array was conducive to a prediction model with strong generalization ability and a high coefficient of determination. This study extracted the mean differential coefficient value (V_{mdc}), the 8th-second transient value (V_8), the response area value (V_{rav}), the maximum gradient value (V_{mgv}), the maximum value (V_{max}), the mean value (V_{mean}), and the variance value (V_{vav}), to construct a feature space, where the formulas of V_{mdc}, V_{rav}, V_{mgv}, and V_{vav} are as

follows:

$$V_{\mathrm{mdc}}=\frac{1}{N-1}\sum_{i=1}^{N-1}\frac{X_{i+1}-X_i}{\Delta t} \tag{1}$$

$$V_{\mathrm{rav}}=\sum_{i=1}^{N-1}X_i\Delta t \tag{2}$$

$$V_{\mathrm{mgv}}=\frac{X_{i\max}-X_0}{i} \tag{3}$$

$$V_{\mathrm{vav}}=\frac{\sum_{i=1}^{N}(X_i-\bar{X})}{N} \tag{4}$$

where, X_i is the ith data collected by the sensor; Δt is the interval time between two adjacent collection points, taking 0.1 s; N is the total number of collected data; $X_{i\max}$ is the maximum value of the collected data; X_0 is the initial response value of the collected data; i is the time corresponding to the maximum value in the collected data; $\bar{X}$ is the average value of the collected data.

A total of 121 soil samples were collected. The sensor array was composed of ten gas sensors. The response curve of each sensor extracted seven characteristic values, combined into a 121×10×7 artificial olfactory feature space (AOFS). Due to the different dimensions of the extracted eigenvalues, features with a large order of magnitude occupy a larger proportion in the modeling, which is not conducive to building an accurate prediction model. The z-score standardization method can make the feature space meet the standard normal distribution, suitable for the case of outlier data beyond the range of values in the sequence. To eliminate the influence of the order of magnitude and dimension on the modeling, this study used the z-score standardization method to normalize the extracted features, and the equation is

$$z=(x-u)/s \tag{5}$$

where, z is the standardized data; x can respectively represent the original values of the seven extracted features; u is the average value of each feature; s is the standard deviation of each feature.

2.5 Training set and test set division

To build a model that can accurately predict the STN content, the normalized AOFS needs to be divided into two parts: a training set and a test set. The training set was used for fit data samples, train the model's parameters, and reduce the generalization error of the model so that the model can reflect reality, and then predict future or other unknown information. The test set was used to evaluate the prediction performance of the model. The Kennard-Stone method was used to set the ratio of the training set of the test set to 7:3[45].

3 Pattern recognition algorithm

To improve the detection accuracy of this method, PLSR, BPNN, and PLSR-BPNN algorithms were used to establish the prediction model between AOFS and STN content, to find the optimal prediction model of STN content.

3.1 PLSR model

Partial least squares regression analysis is a new multivariate statistical data analysis method that combines principal component analysis and multivariate linear analysis statistics. When there is a high linear correlation between variables, PLSR can establish a high prediction model. To overcome the collinearity problem between predictors, PLSR decomposed the independent variables and dependent variables sets through linear combination to extract the principal component factor (PCF), and established a regression model based on the PCF[46,47]. Therefore, an appropriate number of PCFs is one of the effective methods to make full use of the information of the gas sensor array, filter out noise, and effectively avoid over-fitting or under-fitting of the model[48]. The leave-one-cross-validation method was used to determine the number of PCF retained in the PLSR model, and the PCF impact on the model performance was evaluated by the root-mean-square error of cross-validation (RMSECV)[49,50].

3.2 BPNN model

BP neural network is a multi-layer feed-forward network trained by error back-propagation. The structure consists of the input layer, hidden layer, and output layer. Each part is connected by weight and threshold[51]. The basic BPNN algorithm revises the weights of each unit repeatedly through two processes of signal forward propagation and error back-propagation, to approach the target value infinitely. The number of hidden layer neurons greatly influences the BP neural network performance. If the number of hidden layer neurons is small, the neural network cannot fully describe the relationship between output and input variables. On the contrary, if the number of hidden layer neurons is large, it will lead to a longer learning time of the network, such as over-fitting problems. The selection equation of the optimal number of neurons in the hidden layer is as follows:

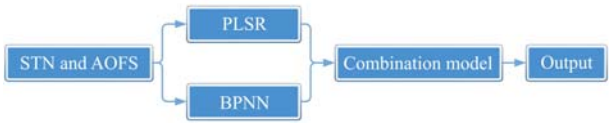
$$l=\sqrt{(m+n)}+\alpha \tag{6}$$

where, n is the extracted eigenvalue; m is the output TN content; α is a constant from 0 to 10.

In addition, the neuron activation function, the number of training iterations, and the target error can also affect the prediction effect of the model, but they are usually set to fixed values.

3.3 PLSR-BPNN model

The causal relationship between soil pyrolysis gas and STN content is complex. The relationship between STN and AOFS can be either linear or nonlinear. Suppose that PLSR and BPNN are organically combined to form the PLSR-BPNN model, PLSR's linear modeling, and BPNN's nonlinear mapping ability. In that case, both algorithms can be fully utilized to their fullest extent to improve the prediction ability of the artificial olfactory detection method for STN content. Therefore, PLSR and BPNN were organically combined in parallel mode, as shown in Figure 5.



Note: STN: Soil total nitrogen; AOFS: Artificial olfactory feature space; PLSR: least squares regression algorithm; BPNN: Back-propagation neural network.

Figure 5 Structure diagram of PLSR-BPNN combined model

The modeling and prediction process of the PLSR-BPNN model was as follows:

1) The PLSR and BPNN models were established separately using the training set;

2) The established PLRS and BPNN models were predicted using the test set, and the predicted values y_p and y_b of the two models were obtained;

3) The weighted sum of y_p and y_b provided the final prediction result $\hat{y}_i$, which equation is

$$\hat{y}_i=k_1\cdot y_p+k_2\cdot y_b \tag{7}$$

where, k_1 and k_2 are the weighting coefficients.

4) A_i represents the prediction precision sequence of the model, and E and σ represent the mean value and mean square deviation of A_i, respectively.

5) S represents the validity of the model. Therefore S can be defined as,

$$S=E\cdot(1-\sigma) \tag{8}$$

The larger the S is, the higher the prediction accuracy of the model is. To determine k_1 and k_2 values, S_p and S_b were used to represent the validity of the PLRS model and BPNN model, respectively. S_p and S_b were normalized as values of k_1 and k_2, respectively, namely as,

$$k_1 = \frac{S_p}{S_p + S_b},\ k_2 = \frac{S_b}{S_b + S_p} \tag{9}$$

3.4 Model evaluation index

To evaluate the quality and reliability of the above models, the coefficient of determination (R^2), the root mean square error (RMSE), and the ratio of performance to deviation (RPD) were used as performance indicators. The equations for R^2, RMSE, and RPD can be expressed as the follows equations:

$$R^2 = \frac{(\sum_{i=1}^{n}(f_i - \frac{1}{n}\sum_{i=1}^{n} f_i)(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i))^2}{\sum_{i=1}^{n}(f_i - \frac{1}{n}\sum_{i=1}^{n} f_i)^2 \sum_{i=1}^{n}(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i)^2} \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2} \tag{11}$$

$$\mathrm{RPD} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i)^2}{\sum_{i=1}^{n}(f_i - y_i)^2}} \tag{12}$$

where, n is the number of samples in the training set or test set; y_i is the actual value of STN of the ith sample; f_i is the predicted value of STN of the ith sample.

The closer R^2 is to 1, the better the fitting effect of the model. RMSE was used to represent the error between the predicted and the measured value of the model. The smaller RMSE is, the higher the prediction accuracy of the model. Vohland et al.[52] reported a detailed prediction and evaluation method for soil properties, where the predictive performance of the model can be evaluated as "excellent" with accurate quantitative analysis when R^2>0.90 and RPD>3.0. When R^2 is between 0.82 and 0.90 and RPD is between 2.5 and 3.0, the model's prediction ability can be defined as "good", and the model has a basic quantitative prediction ability. When R^2 is between 0.66 and 0.81 and RPD is between 2.0 and 2.5, the model performance index can be defined as "qualified", indicating that the model can be approximately quantitatively predicted. When R^2 is 0.50-0.65 and RPD is 1.5-2.0, the model is judged as "poor" and can only be used to distinguish between high and low values. When R^2 is lower than 0.5 and RPD is lower than 1.5, the calibration of the model is "unsuccessful", and the prediction results have no reference.

4 Results and analysis

4.1 Kjeldahl method results of soil total nitrogen content

The measurement results of the Kjeldahl method were statistically analyzed by SPSS 24 software. As shown in Figure 6, the K-S test statistic value is 0.001 (p>0.05), and the total nitrogen content of the soil sample approximately obeys normal distribution. The total nitrogen content of the sample ranges from 0.20-4.10 g/kg, the average value is 1.6789 g/kg, the variance is 0.626 g/kg, and the coefficient of variation is 47.11%. It shows that the distribution of total nitrogen content of the sample presents a large spatial variability, which provides a more comprehensive data support for the establishment of a prediction model of STN content.

4.2 System influence factor test results

To obtain the best pyrolysis state of soil samples, the pyrolysis gas with a strong response of gas sensor array is obtained.

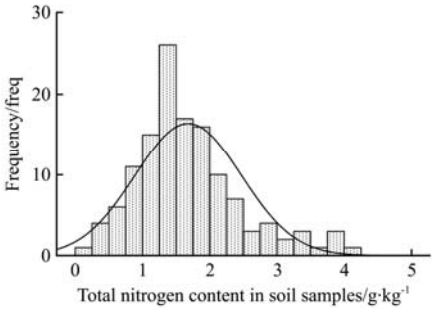


Figure 6 Statistical analysis results of total nitrogen content in soil samples

Pyrolysis temperatures of 200°C, 300°C, 400°C, and 500°C and the pyrolysis times of 1 min, 3 min, 5 min, and 10 min were tested to find the best pyrolysis temperature and time. The soil samples with the lowest, highest, and average TN content were selected for experiments. The test results showed that the sensor array response curve trend of soil samples with different TN contents was roughly the same, and as the TN content of the soil increased, the response intensity of the sensor had an upward trend. The soil sample with a TN content around the average value of the total soil sample has been used as an example to test the pyrolysis temperature and time. The sensor response strength was taken as the test standard, and parallel samples were tested five times.

4.2.1 Pyrolysis temperature test

To test whether the pyrolysis temperature affects the sensor array, a temperature sensor was added to the reaction chamber. The results showed that the reaction chamber temperature did not change significantly after the soil pyrolysis gas generated at different temperatures entered the reaction chamber through the conduit and remained at about room temperature. The pyrolysis temperature does not affect the reaction chamber temperature or the response of the sensor array.

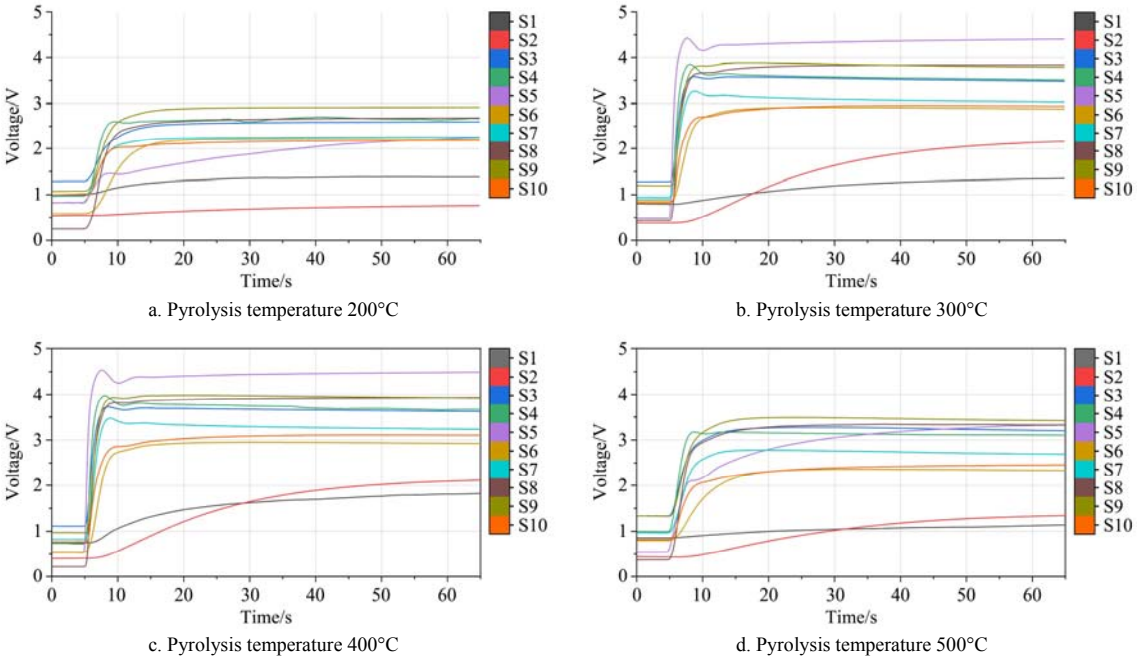
The pyrolysis temperature directly affects the pyrolysis products of the soil sample. 2 g soil samples were measured, the pyrolysis time was 10 min, the data collection time was 1 min, and the pyrolysis temperatures were 200°C, 300°C, 400°C, and 500°C, respectively. The gas flow rate was 3 L/min for washing, and the sensor array response curve was obtained, as shown in Figure 7.

Figure 7 shows that when the pyrolysis temperature is 200°C, all sensors have evident responses except sensor S2. When the pyrolysis temperature is 300°C, the response of the sensor array is significantly improved, and the response of sensor S2 is very strong, indicating that the pyrolysis temperature of 300°C increases the pyrolysis products of soil samples and the concentration of pyrolysis gas, resulting in ammonia, VOC, hydrogen sulfide, and other gases. When the pyrolysis temperature is 400°C, the sensor S1 response is stronger than before, while the responses of other sensors do not change significantly compared with 300°C, indicating that more ammonia is pyrolyzed at this temperature, which makes the sensor S1 respond strongly. When the pyrolysis temperature is 500°C, the sensor array response is greatly reduced because the higher pyrolysis temperature allows for complete pyrolysis, and changes the pyrolysis gas composition. The gas detected by the sensor array is reduced, leading to a decrease in the sensor response intensity. Therefore, the optimal pyrolysis temperature of 400°C was selected in this study.

4.2.2 Pyrolysis time test

The pyrolysis time may affect the response of the sensor array, so the pyrolysis times of 1 min, 3 min, 5 min, and 10 min were

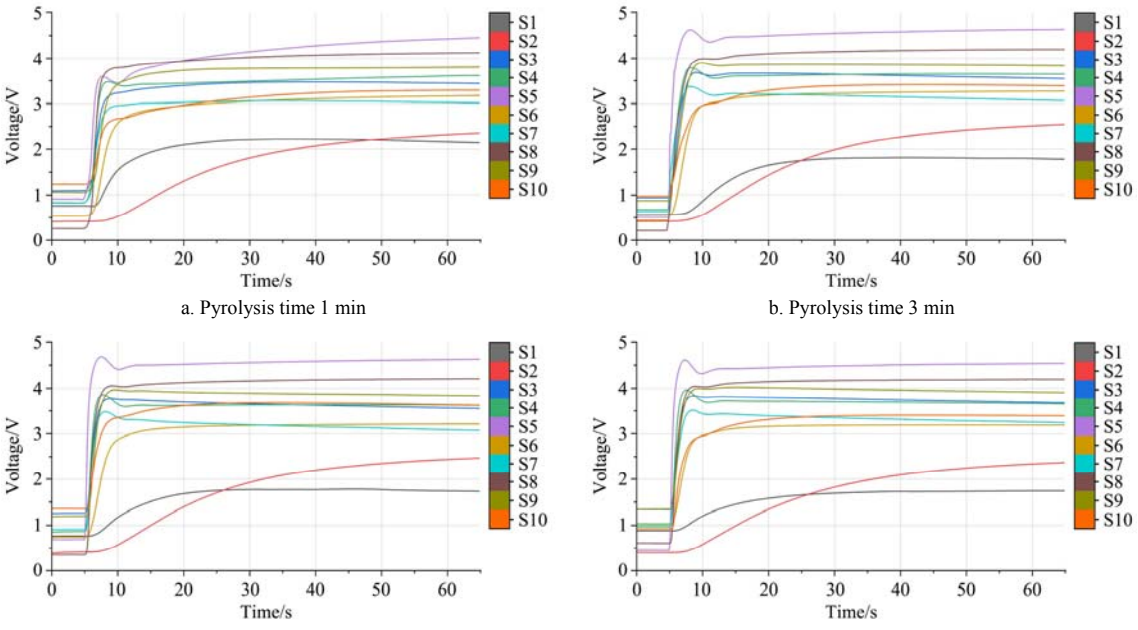
selected for testing. 2 g soil samples were measured, the pyrolysis temperature was 400°C, the data collection time was 1 min, the gas flow rate was 3 L/min for washing, and the response curve of the sensor array was obtained as shown in Figure 8.



a. Pyrolysis temperature 200°C b. Pyrolysis temperature 300°C

c. Pyrolysis temperature 400°C d. Pyrolysis temperature 500°C

Figure 7 Influence of pyrolysis temperature on sensor response



a. Pyrolysis time 1 min b. Pyrolysis time 3 min

c. Pyrolysis time 5 min d. Pyrolysis time 10 min

Figure 8 Influence of pyrolysis time on sensor response

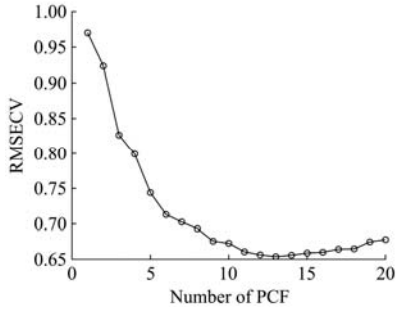
As shown in Figure 8, when the pyrolysis temperature is 400°C and the pyrolysis time is 1 min, the sensor array has an evident response. Especially sensor S1 has a stronger response than other pyrolysis times, but the initial response of other sensors within 5-10 s is relatively low. The sensor array responds strongly and changes little when the pyrolysis time is 3 min, 5 min, and 10 min. To improve the working efficiency of the system and reduce the acquisition time, the pyrolysis time selected was 3 min.

4.3 Modeling soil total nitrogen content

4.3.1 PLSR model establishment

The number of principal component factors (PCF) affects the PLSR model prediction accuracy, and an appropriate number of PCFs effectively improve the model's prediction performance. Here, a 20-component and one-retention cross-validation method was used to find the appropriate number of PCF, and the robustness of the model was verified by the root-mean-square error of cross-validation (RMSECV). The relationship between RMSECV and PCF was obtained, as shown in Figure 9. When the number of PCFs is 10-15, the RMSECV is small. To obtain the most appropriate number of PCFs, 10-15 PCF numbers were used to establish the PLSR model, and the influence of the PCFs number on the prediction performance of the PLSR model training and test sets was obtained as shown in Table 2. As the number of PCF increases, the R^2 and PRD of the training set increase, while RMSE decreases. However, the prediction performance of the test set is different from that of the training set. When the number of PCF is 12, R^2 and PRD reach the maximum, RMSE is the minimum, and then the prediction performance decreases. Therefore, as the

number of PCFs increases, the PLSR model has overfitting. A smaller number of PCFs can reduce the complexity of the model. Therefore, this study used 12 PCFs to establish a PLSR prediction model.



Note: PCF: Principal component factor.

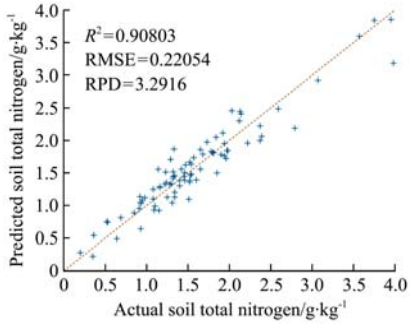
Figure 9 Relationship between the number of PCFs in PLSR and RMSECV

Table 2 Influence of the number of PCFs on the prediction performance of the PLSR model

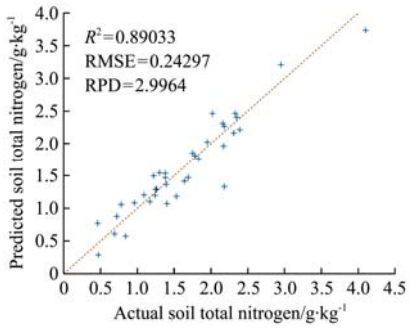
Number of PCF	Training set			Test set		
	R^2	RMSE	RPD	R^2	RMSE	RPD
10	0.89847	0.23187	3.1308	0.88157	0.24773	2.9388
11	0.90387	0.22543	3.2201	0.88235	0.24632	2.9498
12	0.90803	0.22054	3.2916	0.89033	0.24297	2.9964
13	0.91313	0.2145	3.3842	0.88338	0.24682	2.9497
14	0.91619	0.21066	3.4459	0.88506	0.24909	2.9228
15	0.91649	0.2102	3.4534	0.88423	0.25142	2.9344

Note: PCF: Principal component factor.

The PLSR prediction model was established using the training set divided into 12 PCFs and AOFS. The model prediction performance was verified using the test set, as shown in Figure 10. The model built with the training set is $R^2=0.90803$, RMSE=0.22054, RPD=3.2916 (Figure 10a), and the test set $R^2=0.89033$, RMSE=0.24297, RPD=2.9964 (Figure 10b). The model prediction results showed that the R^2 of training and test sets of the STN prediction model established by the PLSR algorithm was



a. Training set



b. Test set

Figure 10 Prediction results of the PLSR model

greater than 0.89, RMSE was less than 0.24297, and RPD was greater than 2.9964. The model's prediction ability could be defined as "good", with basic quantitative prediction ability and good generalization ability.

4.3.2 BPNN model establishment

The neural network toolbox in MATLAB 9.7.0.1190202 (R2019b) software was used to establish the BPNN prediction model. The tansig function was used as the transfer function of the hidden layer, linear function purelin was used as the transfer function of the output layer, and newff function was used to create the neural network. The learning rate was set as 0.01, the target error was set as 0.001, the number of training iterations was set as 1000, the training set was brought into the neural network for training by train function, and finally, the BPNN model which had been trained was simulated and predicted by the sim function.

In BPNN modeling, the number of hidden layer neurons directly affects the model's predictive performance. According to Equation (6), the number of neurons was 8-18. The minimum, maximum, and average values of the determination coefficients of training and test sets corresponding to the program running 15 times were selected as evaluation indexes. The obtained results are listed in Table 3. A too-small number of neurons leads to inadequate model fitting and low model accuracy. An excessive number of neurons leads to over-fitting and affects the model's generalization ability. Table 3 shows that when the number of neurons is 11, the training set's min, mean, and three indicators in the test set are the highest. Therefore, the model parameters were identified as collateral 70-11-1 (70 inputs, 11 hidden layer neurons, and 1 output), and the predicted results of its training set and test set are shown in Figure 11.

Table 3 Influence of neuron number on BPNN model prediction performance

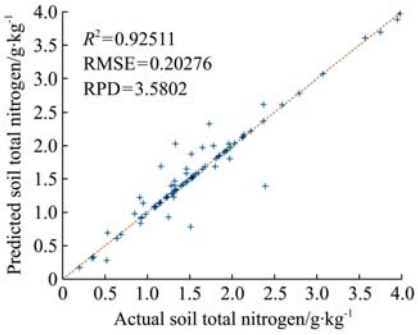
Number of neurons	Training set R^2			Test set R^2		
	Min	Max	Mean	Min	Max	Mean
8	0.74444	0.91141	0.82797	0.44527	0.72187	0.62905
9	0.74573	0.90545	0.85386	0.45987	0.77321	0.65920
10	0.74806	0.9114	0.83811	0.43088	0.72338	0.59701
11	0.76955	0.92511	0.86931	0.59020	0.81185	0.69744
12	0.55438	0.90864	0.83039	0.50233	0.71754	0.63657
13	0.65971	0.90746	0.84156	0.51564	0.72467	0.64891
14	0.58961	0.92549	0.81381	0.56518	0.70789	0.63549
15	0.70981	0.91982	0.81419	0.55491	0.73597	0.59987
16	0.69181	0.89771	0.82116	0.56941	0.74164	0.61094
17	0.54249	0.88540	0.80200	0.46971	0.71697	0.60159
18	0.53100	0.91056	0.82469	0.44569	0.70163	0.59483

The model prediction results show that the model established by the BPNN algorithm has $R^2=0.92511$, RMSE=0.20276, RPD=3.5802 in the training set (Figure 11a), and an $R^2=0.81185$, an RMSE=0.37370, and an RPD=1.9482 in the test set (Figure 11b). The model has high accuracy and fitting effect in the training set, but the R^2 and RPD in the test set have dropped significantly. It shows that the STN prediction model established by the BPNN algorithm is lacking in generalization ability.

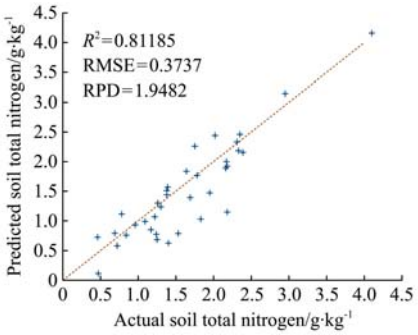
4.3.3 PLSR-BPNN model establishment

From the PLSR and BPNN model prediction results established above, the PLSR model has basic quantitative prediction capabilities and good generalization capabilities, but its prediction accuracy needs to be improved. The BPNN model has good accuracy in the training set; however, its prediction performance in the test set is significantly reduced, and the model's

generalization ability is insufficient. To improve the accuracy and generalization ability of the model, PLSR and BPNN models were combined to establish an STN prediction model based on the PLSR-BPNN algorithm. The parameter settings of the two parallel combined branch algorithms, PLSR and BPNN, are the same as the modeling parameters of the single PLSR and BPNN models, respectively. The PCF of PLSR is 12, and the number of neurons in the hidden layer of BPNN is 11. According to Equation (9), the weighted coefficients of PLSR and BPNN can be calculated as k_1=0.72 and k_2=0.28, respectively, and the prediction results of the training set and test set of the PLSR-BPNN model are obtained as shown in Figure 12.

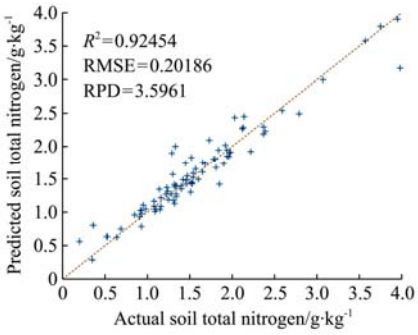


a. Training set

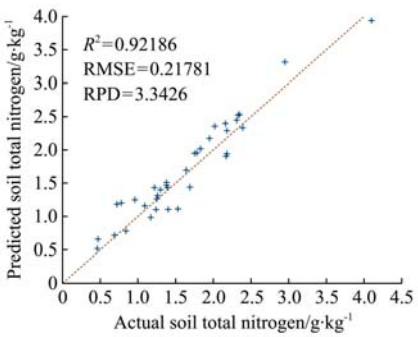


b. Test set

Figure 11 Prediction results of the BPNN model



a. Training set



b. Test set

Figure 12 Prediction results of the PLSR-BPNN model

Figure 12 shows that the prediction model established by the PLSR-BPNN algorithm has an R^2=0.92454, an RMSE=0.20186, and an RPD=3.5961 in the training set (Figure 12a), and R^2= 0.92186, RMSE=0.21781, and RPD=3.3426 in the test set (Figure 12b). The PLSR-BPNN model training and test sets showed an R^2>0.93 and an RPD>3.73; therefore, the prediction performance of the model can be rated as "excellent", with accurate quantitative analysis capabilities and good generalization capabilities.

4.4 Comparative analysis of STN content models

To improve the detection method accuracy, PLSR, BPNN, and PLSR-BPNN algorithms were used to test the training and test sets of AOFS to find the optimal relationship model. Among the prediction models of the PLSR, BPNN, and PLSR-BPNN training sets, the prediction effect of the PLSR-BPNN model was the best (the R^2 and RPD are the largest, and the RMSE is the smallest), followed by BPNN, and PLSR, as shown in Figures 10a, 11a, and 12a. However, the R^2 of the three models are all greater than 0.90, and the RPD is greater than 3.29, indicating that the prediction performance of the three models in the training set can be rated as "excellent", and they have accurate quantitative analysis capabilities.

The test set was brought into the already trained BPNN, PLSR, and PLSR-BPNN models to obtain the prediction results in Figures 10b, 11b, and 12b. For a more direct comparison, a scatter plot of the prediction results of the three models, and the actual STN content was drawn as shown in Figure 13, and the relevant prediction performance of the model test set is listed in Table 4. The prediction result of the BPNN model has a large deviation from the actual value, consistent with its lower R^2, as shown in Figure 13. Compared with the BPNN model, the accuracy of the model established by the PLSR algorithm is significantly improved, and it can quantitatively predict the STN content. For the PLSR-BPNN model, except that the predicted values of soil samples 3 and 6 are significantly different from the actual STN content, the rest of the test set samples can predict the STN content accurately. The results showed a high correlation between soil pyrolysis gas and STN content.

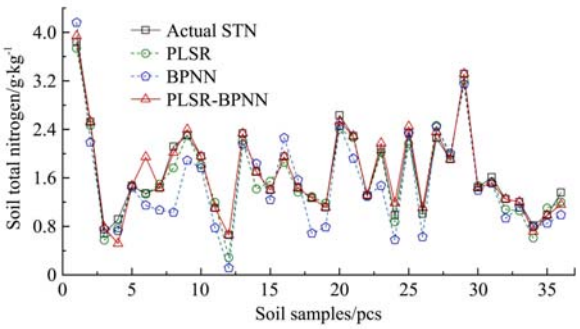


Figure 13 Comparison of model prediction results and actual soil total nitrogen content

Table 4 Comparison of prediction performance of model test set

Models	R^2	RMSE	RPD
PLSR	0.89033	0.24297	2.9964
BPNN	0.81185	0.37370	1.9482
PLSR-BPNN	0.92186	0.21781	3.3426

Note: PLSR: least squares regression algorithm; BPNN: Back-propagation neural network.

The comparison of the prediction performance of the test set (Table 4) shows that the R^2 of the three models are all greater than

0.81, and the RMSE<0.38, indicating that all models have good prediction capabilities. However, in terms of model rating, the BPNN algorithm prediction model can barely be defined as "qualified", the PLSR model can barely be defined as "good", and only the PLSR-BPNN model can be defined as "excellent", with accurate quantitative analysis capabilities. The modeling results showed that the PLSR-BPNN-algorithm STN-prediction model could effectively improve the prediction accuracy of the PLSR model, solve the problems of the insufficient generalization ability of the BPNN model, and provide a new reliable method for STN measurement. The reason is that there is a certain degree of linear and nonlinear correlation between the STN content and the AOFS. The PLSR-BPNN model compensates for the lack of nonlinear and linear relationships between the PLSR and BPNN models to establish a more accurate STN prediction model.

5 Conclusions

This study proposed a method of soil total nitrogen (STN) content detection based on pyrolysis and an artificial olfaction system. The method overcomes the disadvantages of the traditional STN content determination methods, such as time-consuming, complicated operation, and pollution. The spectroscopic method for determining STN is affected by soil moisture, soil texture, and iron oxide. The disadvantages of the Py-GC/MS method are that the equipment acquisition cost is high, it cannot be used exclusively to determine STN, and it is time-consuming and difficult to rapidly determine the TN content in many soil samples. The detection method in this study achieves a low-cost, rapid, and accurate determination of STN content.

The gas sensor array used in this study has specificity and cross-sensitivity to produce specific responses to different soil pyrolysis gases, which lays the foundation for establishing STN prediction models. The high prediction accuracy of the three models proved that the soil pyrolysis gas has a high correlation with the STN content. The optimal pyrolysis temperature of 400°C and the pyrolysis time of 3 min were used for soil samples pyrolysis, and ten different types of gas sensors were used to build an artificial olfactory system to complete pyrolysis gas collection, and the PLSR, BPNN, and PLSR-BPNN algorithms were used to construct the STN prediction, achieving the accurate STN content prediction.

The detection accuracies of PLSR, BPNN, and PLSR-BPNN models in the STN content detection method based on pyrolysis and artificial olfactory system were compared and analyzed. The relationship between artificial olfactory feature space and STN content was established using these three models. R^2, RMSE, and RPD were used as performance indexes to compare the prediction ability of the models. The modeling results showed that all models had good prediction ability for STN content, but the PLSR-BPNN model had the highest R^2 and RPD and the lowest RMSE. The reason is that there is a certain degree of linear and nonlinear correlation between STN content and AOFS. The PLSR-BPNN model can make up for the lack of nonlinear and linear relationships between PLSR and BPNN models respectively, to establish a more accurate prediction model of STN content. The method based on pyrolysis and artificial olfactory system was stable and effective for determining STN content and provided a new basis for rapid and accurate measurement of STN content.

Acknowledgements

This work was financially supported by the Jilin Science and Technology Development Plan (Grant No. 20200502007NC).

[References]

[1] Guo P T, Li M F, Luo W, Lin Q H, Tang Q F, Liu Z W. Prediction of soil total nitrogen for rubber plantation at regional scale based on environmental variables and random forest approach. Transactions of the CSAE, 2015; 31(5): 194–202. (in Chinese)

[2] Li Y, Wang R H, Guan Y L, Jiang Y L, Wu X Q, Peng Q. Prediction analysis of soil total nitrogen content based on hyperspectral. Remote Sensing Technology and Application, 2017; 32(1): 173–179. (in Chinese)

[3] Zhou P, Yang W, Li M Z, Deng L H, Chen Y Q. Soil total nitrogen content prediction based on gray correlation-extreme learning machine. Transactions of the CSAM, 2017; 48(S1): 271–276. (in Chinese)

[4] Song Q. Study on composition and property of organic nitrogen in several soils of China. Acta Pedologica Sinica, 1988; 25(1): 95–100. (in Chinese)

[5] Guan L Z. General soil science, 2nd ed. Beijing: China Agricultural University Press, 2016; 382p. (in Chinese)

[6] Wang M H. Development of precision agriculture and innovation of engineering technologies. Transactions of the CSAE, 1999; 15(1): 1–8. (in Chinese)

[7] Hou Y P, Han L G, Kong L L, Yi C X, Qin Y B, Li Q, et al. Nutrient absorption, translocation in rice and soil nitrogen equilibrium under different nitrogen application doses. Journal of Plant Nutrition and Fertilizer, 2015; 21(4): 836–845. (in Chinese)

[8] Qin L, Huang S Q, Zhong L L, Zhou H, Zhao S, Xiang B, Lei S R. Comparison of Dumas combustion and Kjeldahl methods for determining total nitrogen content in soil. Soil and Fertilizer Sciences in China, 2020; 4: 258–265. (in Chinese)

[9] Zhang J J, Tian Y C, Yao X, Cao W X, Ma X M, Zhu Y. Estimating model of soil total nitrogen content based on near-infrared spectroscopy analysis. Transactions of the CSAE, 2012; 28(12): 183–188. (in Chinese)

[10] Ben-Dor E, Banin A. Near-Infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal, 1995; 59(2): 364–372.

[11] Yao X, Ren H, Cao Z, Tian Y, Cao W, Zhu Y, Chen T. Detecting leaf nitrogen content in wheat with canopy hyperspectrum under different soil backgrounds. International Journal of Applied Earth Observation & Geoinformation, 2014; 32: 114–124.

[12] An X F, Li M Z, Deng L H, Liu Y M, Zhang Y J. Performance of portable soil TN detector based on NIR spectroscopy. Transactions of the CCSAM, 2012; 43(S1): 283–288. (in Chinese)

[13] Zhao Y D, Pi T T. Spectral prediction model of soil total nitrogen content of clay loam soil in Beijing. Transactions of the CSAM, 2016; 47(3): 144–149. (in Chinese)

[14] Wang H J, Liu F, Yunger J A, Cui J, Ma L. Fitting model of soil total nitrogen content in different soil particle sizes using hyperspectral analysis. Transactions of the CSAM, 2019; 50(2): 202–211. (in Chinese)

[15] Nie P C, Dong T, He Y, Xiao S P, Qu F F, Lin L. The effects of drying temperature on nitrogen concentration detection in calcium soil studied by NIR spectroscopy. Applied Sciences, 2018; 8(2): 269. doi: 10.3390/app8020269.

[16] He Y, Xiao S P, Nie P C, Dong T, Qu F F, Lin L. Research on the optimum water content of detecting soil nitrogen using near infrared sensor. Sensors, 2017; 17(9): 2045. doi: 10.3390/s17092045.

[17] Stoner E R, Baumgardner M F. Characteristic variations in reflectance of surface soils. Soil Science Society of America Journal, 1981; 45(6): 1161–1165.

[18] Bowers S A, Hanks R J. Reflection of radiant energy from soil. Soil Science, 1965; 100(2): 130–138.

[19] Liu W D, Baret F, Gu X F, Tong Q X, Zheng L F, Zhang B. Relating soil surface moisture to reflectance. Remote Sensing of Environment, 2002; 81(2): 238–246.

[20] De Cesare F, Di Mattia E, Pantalei S, Zampetti E, Vinciguerra V, Canganella F, et al. Use of electronic nose technology to measure soil microbial activity through biogenic volatile organic compounds and gases release. Soil Biology & Biochemistry, 2011; 43(10): 2094–2107.

[21] Sheppard S K, Lloyd D. Direct mass spectrometric measurement of gases in soil monoliths. Journal of Microbiological Methods, 2002; 50(2): 175–188.

[22] Mcneal K S, Herbert B E. Volatile organic metabolites as indicators of soil microbial activity and community composition shifts. Soil Science Society of America Journal, 2009; 73(2): 579–588.

[23] Fernando W G D, Ramarathnam R, Krishnamoorthy A S, Savchuk S C.

Identification and use of potential bacterial organic antifungal volatiles in biocontrol. Soil Biology & Biochemistry, 2005; 37(5): 955–964.

[24] Zhu L T, Li M W, Xia X M, Huang D Y, Jia H L. Soil organic matter detection method based on artificial olfactory system. Transactions of the CSAM 2020; 51(3): 171–179. (in Chinese)

[25] Zhu L T, Jia H L, Chen Y B, Wang Q, Li M W, Huang D Y, et al. A novel method for soil organic matter determination by using an artificial olfactory system. Sensors, 2019; 19(15): 3417. doi: 10.3390/s19153417

[26] Al-Rahbi A S, Onwudili J A, Williams P T. Thermal decomposition and gasification of biomass pyrolysis gases using a hot bed of waste derived pyrolysis char. Bioresource Technology, 2016; 204: 71–79.

[27] Kim Y, Oh J I, Lee S S, Lee K H, Lee J, Kwon E E. Decontamination of petroleum-contaminated soil via pyrolysis under carbon dioxide atmosphere. Journal of Cleaner Production, 2019; 236: 117724. doi: 10.1016/j.jclepro.2019.117724.

[28] Li H L, Qu R H, Li C, Guo W L, Han X M, He F, et al. Selective removal of polycyclic aromatic hydrocarbons (PAHs) from soil washing effluents using biochars produced at different pyrolytic temperatures. Bioresource Technology, 2014; 163: 193–198.

[29] Becker J N, Dippold M A, Hemp A, Kuzyakov Y. Ashes to ashes: Characterization of organic matter in Andosols along a 3400 m elevation transect at Mount Kilimanjaro using analytical pyrolysis. Catena, 2019; 180: 271–281.

[30] Girona-Garcia A, Badia-Villas D, Nicasio T, Jimenez-Morillo N T, Gonzalez-Perez G A. Changes in soil organic matter composition after Scots pine afforestation in a native European beech forest revealed by analytical pyrolysis (Py-GC/MS). Science of the Total Environment, 2019; 691: 1155–1161.

[31] Jackie J, Chua C K, Chong D C Y, Lim S Y, Li S F Y. Rapid and sensitive direct detection of endotoxins by Pyrolysis-Gas Chromatography-Mass Spectrometry. ACS Omega, 2021, 6(23): 15192–15198.

[32] al Sandouk-Lincke N A, Schwarzbauer J, Hartkopf-Froeder C, Volk H, Fuentes D, Young M. The effect of different pyrolysis temperatures on organic microfossils, vitrain and amber-A comparative study between laser assisted- and Curie Point-pyrolysis-gas chromatography/mass spectrometry. Journal of Analytical & Applied Pyrolysis, 2014; 107: 211–223.

[33] Stewart C E. Evaluation of angiosperm and fern contributions to soil organic matter using two methods of pyrolysis-gas chromatography-mass spectrometry. Plant and Soil, 2011; 351(1-2): 31–46.

[34] Wei S Y, Pintus V, Schreiner M. Photochemical degradation study of polyvinyl acetate paints used in artworks by Py-GC/MS. Journal of Analytical & Applied Pyrolysis, 2012; 97: 158–163.

[35] Plum A, Engewald W, Rehorek A. Rapid qualitative pyrolysis GC-MS analysis of carcinogenic aromatic amines from dyed textiles. Chromatographia, 2003; 57(S1): S243–S248.

[36] De La Rosa J M, Faria S R, Varela M E, Knicker H, Gonzalez-Vila F J, Gonzalez-Perez J A, Keizer J. Characterization of wildfire effects on soil organic matter using analytical pyrolysis. Geoderma, 2012; 191(S1): 24–30.

[37] Chen Q Y, Wu Y Q, Lei T Z, Si G C, Zhang G X. Study on the fingerprints of soil organic components in alpine grassland based on Py-GC-MS/MS Technology. Acta Ecologica Sinica, 2018; 38(8): 2864–2873. (in Chinese)

[38] Stafilov T, Spiric Z, Glad M, Barandovski L, Andonovska K, Sajn R, et al. Study of nitrogen pollution in the Republic of North Macedonia by moss biomonitoring and Kjeldahl method. Journal of Environmental Science and Health, Part A, 2020; 55(6): 759–764.

[39] Hakoda A, Li Y, Naito S, Suzuki T, Yasui A. Determination of crude protein in macaroni products by the combustion method and comparison with the Kjeldahl method: interlaboratory study. Japanese Society for Food Science and Technology, 2011; 17(3): 227–232.

[40] Kiełczewska K, Dąbrowska A, Jankowska A, Wachowska M, Kowalik J. The effect of high-pressure treatment and skimming on caprine milk proteins. Applied Sciences, 2021; 11(13): 5982. doi: 10.3390/app11135982.

[41] Qi N L, Li P W, Zeng X H, Huang H H, Yang Z M, Gong X. Comparison of Kjeldahl and the elemental analysis methods for determination of nitrogen content in raw natural rubber. Advanced Materials Research, 2013; 815: 722–726.

[42] Silva T E D, Detmann E, Franco M D O, Palma M N N, Rocha G C. Evaluation of digestion procedures in Kjeldahl method to quantify total nitrogen in analyses applied to animal nutrition. Acta Scientiarum: Animal Sciences, 2016; 38(1): 45–51.

[43] Ates F, Kaya O. The relationship between iron and nitrogen concentrations based on Kjeldahl method and SPAD-502 readings in grapevine (*Vitis vinifera* L. cv. 'Sultana Seedless'). Erwerbs-Obstbau, 2021; 63(S1): 53–59.

[44] De la Rosa J M, Faria S R, Varela M E, Knicker H, Gonzalez-Vila F J, Gonzalez-Perez J A, et al. Characterization of wildfire effects on soil organic matter using analytical pyrolysis. Geoderma, 2012; 191(S1): 24–30.

[45] He Z H, Ma Z H, Li M C, Zhou Y. Selection of a calibration sample subset by a semi-supervised method. Journal of Near Infrared Spectroscopy, 2018; 26(2): 87–94.

[46] Qi H J, Paz-Kagan T, Karnieli A, Jin X, Li S W. Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. Soil and Tillage Research, 2018; 175: 267–275.

[47] Wold S, Ruhe A, Wold H, Dunn W J. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J. Sci. Stat. Comput. 1984; 5(3): 735–743.

[48] Rossel R A V, Walvoort D J J, Mcbratney A B, Janik L J, Skjemstad J O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 2006; 131(1-2): 59–75.

[49] Ji W J, Shi Z, Huang J Y, Li S. In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy. PloS One, 2014; 9: e105708. doi: 10.1371/journal.pone.0159785

[50] Li B B, Julian M, Martin E B. Model selection for partial least squares regression. Chemometrics and Intelligent Laboratory System, 2002; 64: 79–89.

[51] LI X F, Xiang S Y, Zhu P F, Wu M. Establishing a dynamic self-adaptation learning algorithm of the BP neural network and its applications. International Journal of Bifurcation and Chaos, 2015; 25(14): 1540030. doi: 10.1142/S0218127415400301.

[52] Vohland M, Besold J, Hill J, Frund H C. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. Geoderma, 2011; 166(1): 198–205.