DETECTION OF THE NUTRITIONAL STATUS OF PHOSPHORUS IN LETTUCE USING THZ TIME-DOMAIN SPECTROSCOPY

Xiaodong Zhang[1,2], Pei Wang[1,2], Hanping Mao[1,2], Hongyan Gao[1,2], Qinglin Li[1*,2]

[1*]Corresponding author. Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang, China. E-mail: lql@ujs.edu.cn | ORCID ID: https://orcid.org/0000-0001-7091-9512

ABSTRACT

Phosphorus deficiency can lead to serious reductions in crop yield and quality. Rapid and accurate determination of phosphorus stress status in vegetables is essential for ensuring the appropriate application of water and fertiliser, improving the yield and quality of vegetables, avoiding the waste of resources and serious non-point pollution source caused by the abuse of chemical fertilisers, and promoting green production and sustainability of facility agriculture. Chemical analysis and spectral or visual imaging methods for detection of phosphorus nutritional status are not conducive in facility production owing to their low accuracy and damage to plants. Owing to its penetration and fingerprint characteristics, terahertz time-domain spectroscopy (THz-TDS) can distinguish differences in internal components caused by excess phosphorus nutrients and changes in nucleic acids, nuclear proteins, phospholipids, and other macromolecules; thus, it could potentially be used to evaluate the nutritional status of crops. In this study, an innovative THz-TDS-based method was used to detect the nutritional status of phosphorus in lettuce. Lettuce was grown with different phosphorus levels using soilless cultivation and different nutrient solutions. Based on the standard formula of Yamasaki nutrient solution, the phosphorus content in the nutrient solutions was reduced or increased by 20%, 60%, 100%, and 150%, and lettuce samples exposed to each of these phosphorus concentrations were collected. Terahertz spectra are highly sensitive to water; thus, the lettuce samples were freeze-dried to minimise the effect of water and maintain their original quality and bioactivity. The spectra of lettuce were recorded using a TS7400 THz-TDS system; noise and interference were eliminated via normalisation based on Savitzky-Golay smoothing. The correction and validation sets were divided using sample set partitioning based on the joint x-y distance (SPXY). The stability competitive adaptive reweighted algorithm, iterative retention information variable algorithm, and interval combination optimisation algorithm were used to select the terahertz characteristic wavelength, and the successive projection algorithm was then used for secondary optimisation. Finally, a THz-TDS model of lettuce phosphorus was established using the partial least squares method with five principal component variables. The coefficient of determination of the model reached 0.7005, and the root-mean-square error of the predictions was 0.003273, indicating that this method has a high prediction accuracy.

INTRODUCTION

Lettuce is a typical leafy vegetable. China being the largest producer of lettuce, accounts for approximately 50% of the global lettuce production. Phosphorus is one of the three major elements necessary for plant growth (Jiang et al., 2014). It is an important component of various compounds such as nucleic acids, nucleoproteins, phospholipids, ATP, and coenzymes (Sun et al., 2014). Phosphorus plays a role in regulating the metabolism of organisms, accelerating the

initial germination of plants, promoting the growth of root systems, and improving disease and drought resistance of plants. The absence or excess of phosphorus in crops can alter the composition of nucleic acids, nucleoproteins, phospholipids, phytochemicals, adenosine phosphate, and various enzymes. Thus, the appropriate application of phosphate fertiliser is critically important for improving the yield and quality of lettuce (Guo & Li, 2021).

Traditional methods of detecting the phosphorus content in crops are mainly based on chemical determination methods. Although these methods have high detection accuracy, they are unsuitable for large, comprehensive, and rapid diagnosis because they are costly, complex, and time-consuming, and can damage crop samples. With the rapid development of information technology, various spectral and imaging technologies have been used to collect and analyse plant morphological data (Zhang et al., 2019). Compared with chemical analysis methods, these technologies have the advantages of being rapid, inexpensive, convenient, and non-destructive. However, these technologies mainly capture changes in reflective characteristics, such as external colours and textures caused by nutrient deficiencies and are also not able to sense differences in internal tissues or components of crops (Araújo et al., 2001). Therefore, nutrient detection based on obvious characteristics, such as leaf chlorosis caused by nitrogen deficiency, generally leads to better results. Moreover, the detection of phosphorus and potassium is less effective using comparison and requires further improvement before it can be more broadly applied (Liu et al., 2014).

Terahertz detection technology is regarded as "one of the top ten technologies influencing the future of mankind in the 21st century." It is an advanced detection technology that has received increased attention from biological scientists (Coutu et al., 2016). Terahertz radiation has penetration, fingerprint, coherent, and multi-dimensional fusion detection capabilities. Unlike traditional spectral and imaging methods that usually only capture reflection characteristics and distribution (image) information of the detected objects in different characteristic spectral bands, terahertz time-domain spectroscopy (THz-TDS) provides rich material information (Wang et al., 2017). It can be used to identify the characteristic fingerprints of biological macromolecules such as proteins and nucleic acids and can detect changes in structural and morphological differences such as external characteristics, internal microstructure, and biological tissue texture caused by variation in the abundance of nutrients. In addition, it can acquire characteristic information not obtained by other technologies in the fields of online food inspection and quality control, agricultural product quality inspections, biomolecular inspections, and non-destructive inspections (Liu et al., 2019).

Wu et al. (2019) used THz-TDS and chemometric methods to qualitatively and quantitatively detect glucose and sucrose in milk powder. These theoretical methods are important for the study of other biological macromolecules. Ge et al. (2015) used terahertz absorption spectra to distinguish eight wheat varieties and established a regression model for wheat variety discrimination using the partial least squares method, with a cross-validated correlation coefficient (R^2) and cross-validated root mean square error (RMSECV) of 0.985 and 1.162, respectively. Karain (2016) performed

terahertz spectroscopy of protein-solvent interaction energy based on the Wiener-Khinchin method and discovered that the spectroscopy spectrum could characterise the spectral range of macromolecular protein mode shapes. Li et al. (2017) showed that THz-TDS could be used in qualitative and quantitative research of d-anhydrous glucose, and they provided a reference for rapid glucose detection in fruits, vegetables, foods, and pharmaceuticals using terahertz spectroscopy. Pagano et al. (2019) studied the feasibility of an advanced terahertz-quantum cascade laser (THz-QCL) instrument to determine the absolute water content of purple coral leaves. In the graph of tau L-A as the M-w function, the best-fit regression line R^2 is always greater than 0.85, indicating that the method can be combined with plant water stress indicators to better understand leaf water management (Pagano et al., 2019). Therefore, THz-TDS can detect differences in the characteristics of biological sample structures and macromolecular components and can be used to detect nutrients, in which characteristics and structures would have been altered owing to nutritional deficiency. In addition, it can detect changes in their internal macromolecules and tissue components.

Owing to its unique advantages in crop detection, THz-TDS was used to predict the phosphorus content of lettuce. The efficacy of the stability competitive adaptive reweighted sampling algorithm (SCARS), iteratively retained information variable algorithm (IRIV), and interval combination optimization algorithm (ICO) were compared. Thereafter, partial least squares regression (PLSR) was used to establish the quantitative detection models of various lettuce phosphorus contents. Comparisons of the lettuce phosphorus content in the best model were made based on the frequency bands of terahertz spectroscopy.

MATERIAL AND METHODS

Lettuce cultivation

The experiment was carried out at the Jiangsu University Provincial Key Laboratory of Modern Agricultural Equipment (China, 32.2 °N, 119.5 °E). The test material used was an Italian moss-resistant half-head lettuce. When the lettuce grew to "five leaves and one heart" after germination and seedling cultivation in plug trays, seedlings with similar growth conditions were transplanted into flowerpots and distributed into batches. Perlite was used as the substrate. The samples were allocated to four treatments: 20%, 60%, 100%, and 150% (all mass fractions) P in the Yamazaki nutrient solution formula. During the culture period, the nutrient solution was watered every morning and evening to ensure that the lettuce grew and that the P concentration remained stable. Lettuce can be placed under nutritional stress in a variety of ways, and interference factors can be removed. A total of 80 samples were included; 20 samples were used at each processing level.

Terahertz spectrum information acquisition

After approximately 30 days of growth, fresh lettuce samples were harvested and freeze-dried. The dryer first froze the water in the sample and then sublimed to a gaseous state. Consequently, most of the molecules in the sample were stored, and the effects of moisture differences between samples during the collection of terahertz information were eliminated.
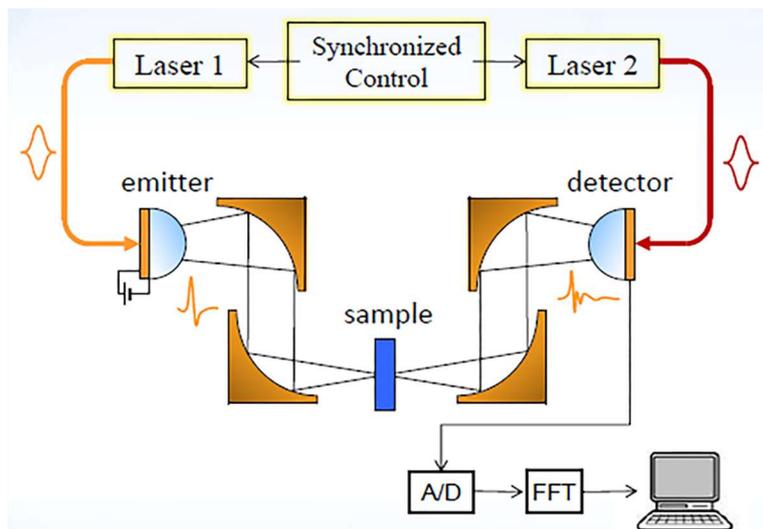
FIGURE. 1 THz-TDS working principle.

Figure 1 shows how THz-TDS works. Under the action of a beam splitter, the laser pulse is divided into two mutually perpendicular laser beams (Jiang et al., 2018). The pump light passes through the terahertz transmitter and reflector through the sample being tested. Through crystal transmission to a terahertz detector, the detector transmits the difference between the two laser beams to the terahertz detector.

A TS7400 THZ-TDS instrument (Advantest Corporation, Japan) was used to collect the terahertz spectra of the samples. The test chamber was dehumidified with compressed air before sampling, and the relative humidity (RH) in the chamber was controlled within 10% RH, which greatly reduced the effect of environmental humidity on the test results. The dried sample was placed on a stage, and the measurement area was calibrated to determine the detection boundary. Background measurements were performed in advance to eliminate the effects of background errors. The spectral data for each of the 80 samples were collected five times, and the average value was used for each sample.

Determination of phosphorus content

Phosphorus content in digestive juice was determined via antimony molybdenum colorimetry using an AA3 continuous-flow analyser. Figure 2 shows the relationship between phosphorus content and photometric value. Phosphorus content positively correlated with the measured photometric values; consequently, this relationship can be used to calculate the phosphorus content of lettuce leaves exposed to different phosphorus concentrations.
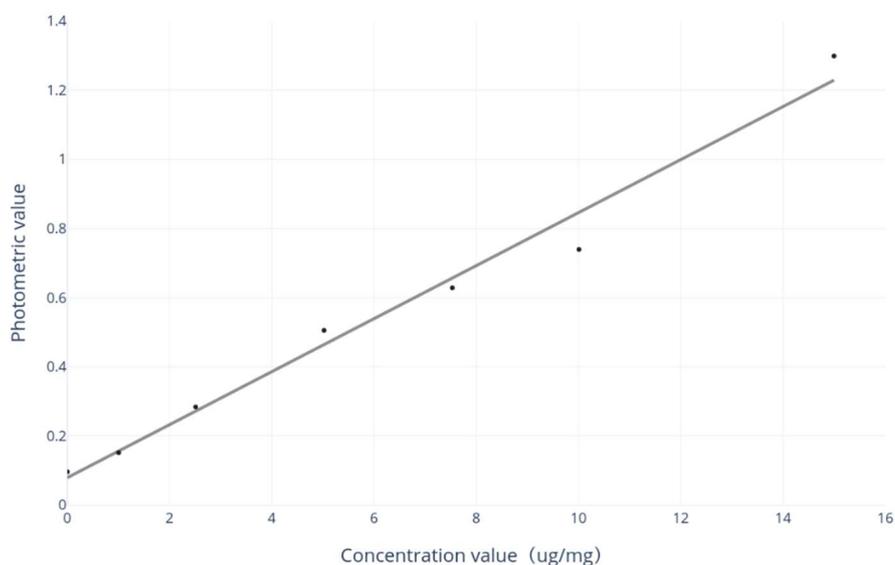

FIGURE. 2 Relationship between phosphorus content and photometric value.

Figure 3 shows the terahertz reflection spectra of leaves exposed to different concentrations of phosphorus. As the phosphorus content increased, the reflective index of the terahertz reflection spectra also increased. As the THz-TDS frequency band increased, the spectral reflection intensity of the different samples increased, and the relationship between the two was significant.
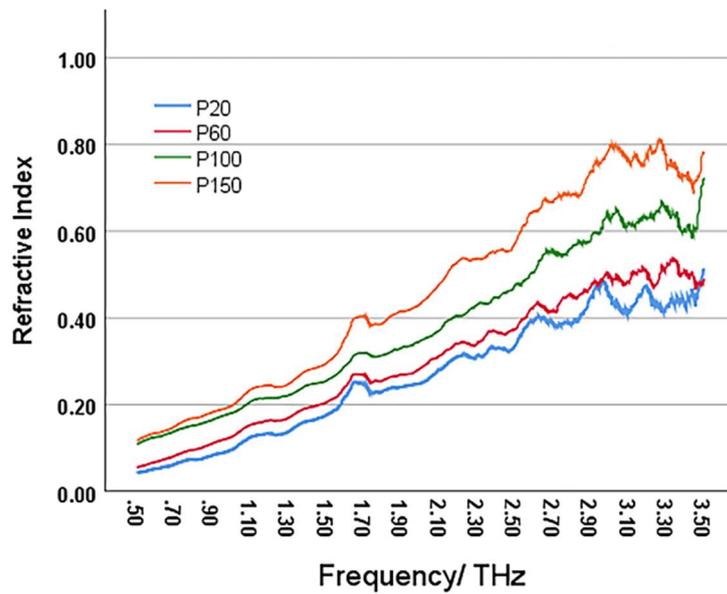
FIGURE. 3 THz reflection spectra of samples exposed to different phosphorus concentrations.

RESULTS AND DISCUSSION

Figure 4 shows that the phosphorus content of lettuce leaves was significantly different under different treatment conditions. However, under 20% and 60% phosphorus contents, the phosphorus content was similar, which may stem from the regulation and adaptation mechanisms of lettuce.
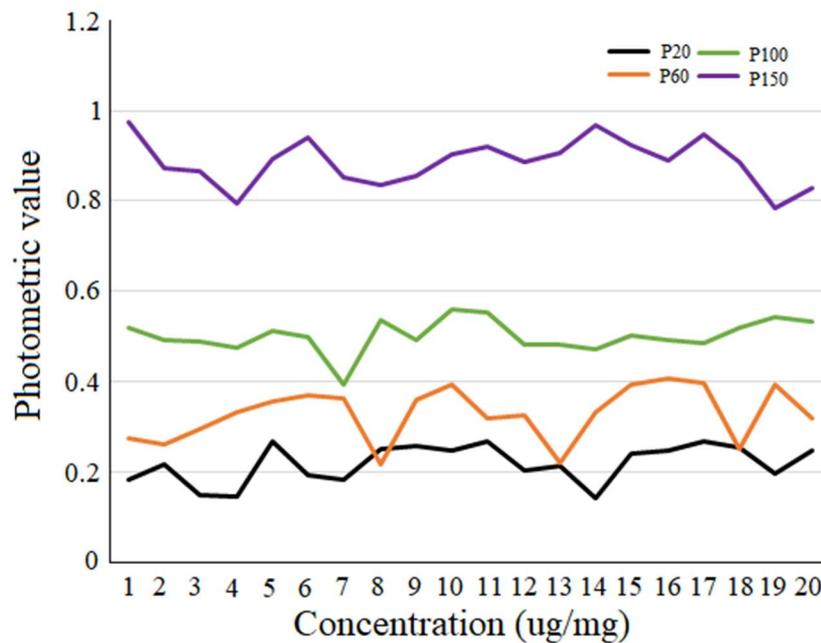

FIGURE. 4 Variation in the phosphorus content under exposure to different phosphorus concentrations.

Spectral data preprocessing

Terahertz spectroscopy requires the establishment of an analytical model, based on a sample database that covers the components of the tested sample, to detect the components and contents of unknown samples. The accuracy of the model directly affects the analytical accuracy.

Noise is inevitable for the spectral data generated by the instrument. However, preprocessing can help improve data quality, validity, and accuracy of the model. Data smoothing can eliminate random errors and instrumental noise. The idea is to ignore small deviations in the data and retain large deviations so that the important information can be largely retained. Figure 5 shows the effects before and after smoothing filtering.

To obtain the best filtering effect, the Savitzky-Golay (SG) smoothing algorithm, SG derivative algorithm, data centralisation, and data normalisation were used to preprocess the original data.
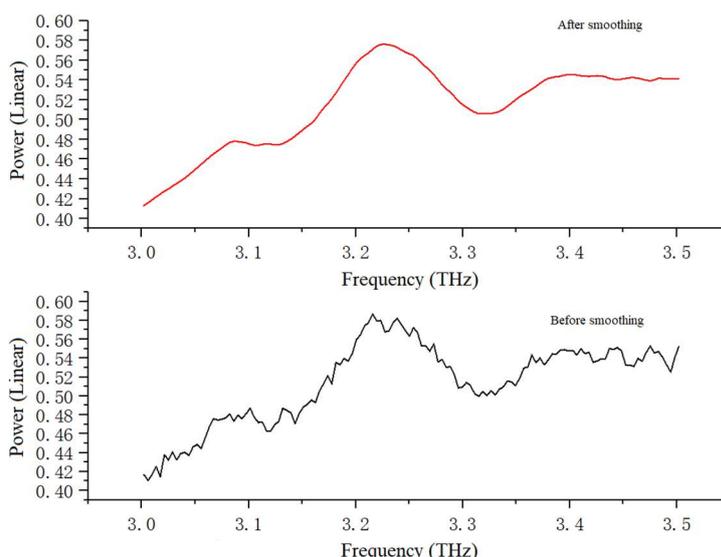
FIGURE. 5 Spectra before and after data smoothing.

Table 1 shows a comparison of the precision obtained using different preprocessing methods. The smoothing and derivative algorithms used window widths of 5, 7, 11, and 15.

The PLSR of the entire THz-TDS spectrum was conducted using the measured phosphorus content, and the R^2 and RMSEC were compared to identify the best window width.

TABLE 1. Regression coefficient R^2 and RMSEC of different preprocessing methods.

Serial number	Preprocessing method	R^2	RMSEC
1	5-point-width smoothing	0.9432	2.941%
2	7-point-width smoothing	0.9142	3.611%
3	9-point-width smoothing	0.8759	4.343%
4	11-point-width smoothing	0.8261	5.140%
5	Normalization	0.9246	3.386%
6	5-point-width derivative	0.9484	3.024%
7	7-point-width derivative	0.9226	3.704%
8	9-point-width derivative	0.9103	4.222%
9	11-point-width derivative	0.8466	4.825%
10	Centralization	0.8080	5.402%

The smoothing and derivative preprocessing methods with a 5-point window width outperformed the algorithms with 7-, 9-, and 11-point window widths. Considering the complementarity among the various algorithms, the two algorithms were combined to obtain a higher accuracy. Table 2 shows the fitting regression coefficients and RMSECs calculated using the two methods together.

TABLE 2. Regression coefficient and root mean square error calculated by combining the two methods.

Serial number	Preprocessing method	R^2	RMSEC
1	Derivative on the basis of smoothing	0.9453	2.763%
2	Centralization on the basis of smoothing	0.9226	3.485%
3	Normalization on the basis of smoothing	0.9510	2.695%
4	Centralization on the smoothing-based derivative	0.9465	2.757%
5	Normalization on the smoothing-based derivative	0.9479	2.819%

The accuracy of the single processing method was 0.8080-0.9484, with an average of 0.8920 and a maximum of 0.9484. The accuracy of the two methods used together was 0.9226-0.951, with an average of 0.9427. The combination of the two methods during preprocessing can effectively improve the accuracy of the PLSR. The modelling had the best normalisation effect based on 5-window-width smoothing (determination coefficient R^2: 0.9510; RMSEC, 2.695%).

Division of the correction set and verification set

The division of sample sets is a key factor that affects the applicability of the model. The predictive ability of the model can be enhanced, provided the selected sets have good representation (Kennard & Stone et al., 1969). The random sampling (RS) algorithm, Kennard-Stone (KS) algorithm, and sample set partitioning based on the joint X-Y distances (SPXY) algorithm were used to classify the sample set. Typically, calibration sets contain 50–75% of the total sample size. In this study, each target tested contained 80 samples: the calibration set contained 60 samples, and the prediction set contained 20 samples. The sample set division method was optimised by comparing and analysing its advantages and disadvantages.
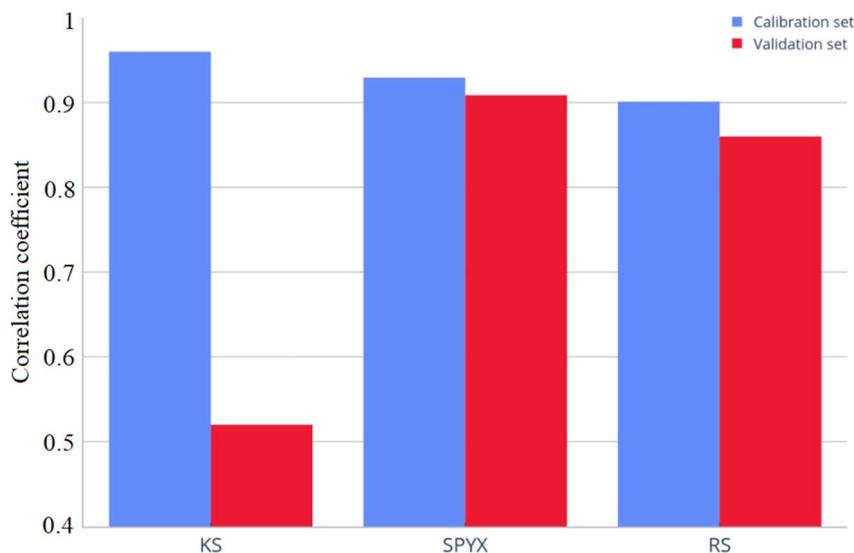

FIGURE 6. Division of the correction sets and verification sets for different algorithms.

Figure 6 shows the classification effect of the sample sets obtained using the three methods above. The correlation coefficient between the calibration and verification sets obtained using the SPXY algorithm classification was the highest; therefore, SPXY was used to divide the sample set.

Feature frequency extraction

Stability competitive adaptive reweighted sampling (SCARS)

The total number of phosphorus samples was 80, and the number of SCARS runs was initially set to 30, 40, and 50. Figure 7 shows the results of the operation set to 50 times. After running up to 40 times, the sample number basically stabilised; thus, the parameter was set to 50 times to obtain stable runs and characterise subsequent trends. During the entire feature variable screening, the number of retained variables gradually decreased as the number of runs increased. The number of sampled variables decreased sharply in the first five sampling processes and then slowly decreased, demonstrating exponentially decreasing function (EDF). This pattern also reflects the preliminary selection and final selection of the algorithm. The principle of SCARS is based on the stability of the variable. Greater stability indicates a greater possibility of the variable being selected using the adaptive weighted sampling technology. EDF selects frequency band points with a large absolute value of regression coefficient in PLSR, removes frequency band points with small weight, and uses interactive verification to select subset with the lowest RMSECV, which can effectively identify the optimal combination of variables.
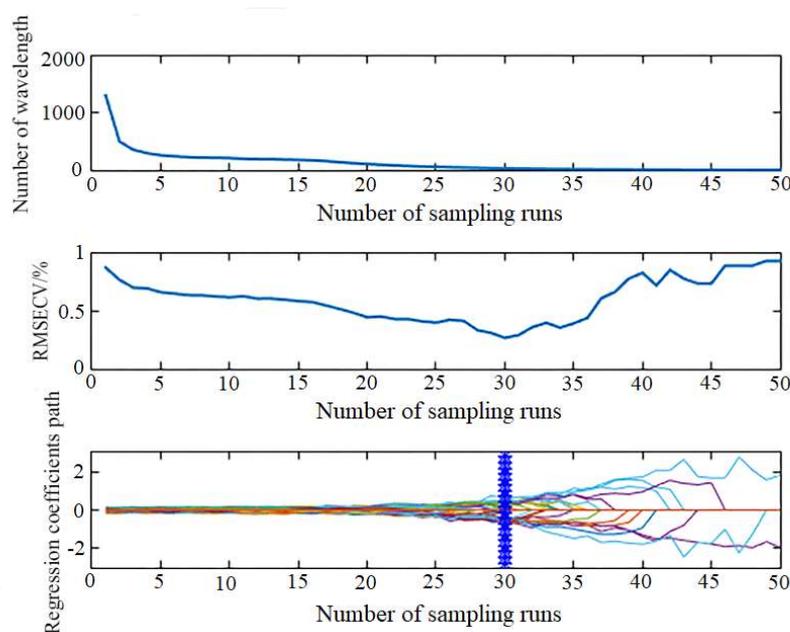
FIGURE. 7 Results of the SCARS algorithm.

The RMSECV of the interactive verification model was 0.340% after 30 runs and then gradually increased, indicating that the algorithm removed the feature variables that had great impact on algorithmic accuracy. Therefore, the subset of feature variables obtained through 30 runs was selected as the best subset, and a total of 19 terahertz feature bands, which was significantly related to phosphorus nutritional status, were initially selected.

Interval combination optimization (ICO)

The local search for each optimised interval was repeated until no new variables could affect the RMSECV of the model. At this time, the optimised interval can be regarded as the wavelength interval that is finally selected by the ICO. A total of 75 frequencies in four intervals were selected as references. The ICO divides the terahertz spectrum of a sample into N parts with equal widths and each part is a spectral interval. Bootstrap sampling was used to generate a subset of random combinations of M different band intervals, and PLS and a 50% cross-test were used to calculate the RMSECV for each interval combination subset. If the inclusion of an adjacent wavelength point causes the RMSECV to decrease, the wavelength point is selected, otherwise, it is excluded.

Iteratively retained information variables (IRIV) algorithm

Table 3 shows the characteristic frequency numbers of the lettuce terahertz spectral features extracted using the three algorithms and the accuracy of PLSR. The model accuracy of SCARS was significantly higher than that of IRIV when the characteristic frequency numbers were approximately the same. However, ICO screening had a higher frequency and lower accuracy than SCARS. Therefore, SCARS was used for feature extraction and subsequent processing.

TABLE 3. Number of frequency bands, regression coefficients, and mean squares screened by different algorithms.

Method	Number of features	R^2_C	RMSEC (%)	R^2_P	RMSEP (%)
SCARS	19	0.897	4.572	0.634	5.824
ICO	75	0.746	5.641	0.537	7.547
IRIV	17	0.797	5.536	0.523	8.352

IRIV based on binary matrix rearrangement filter divides all variables into four types: strong, weak, no information, and interference variables (Yun et al., 2014). IRIV requires multiple iterations. The purpose of each iteration is to retain strong and weak information variables and eliminate the other two types so as to obtain the best variable set. IRIV was used to extract the terahertz spectra, and a total of 17 frequencies were selected.

Establishment of the THz-TDS model for lettuce phosphorus

There were many feature frequency bands obtained using SCARS, many of which contained similar features. To address the multicollinearity among variables, the redundancy and number of variables should be reduced, and the model operation efficiency and application value should be improved. Hence, a successive projection algorithm (SPA) was used to optimise the initial screening results, and a THz-TDS model of lettuce phosphorus was established using PLSR.

Screening of feature variables based on SPA

The spectral values of the 19 characteristic frequency bands obtained using SCARS were used as independent

variables, and the minimum and maximum values of the screening results were set to 1 and 19, respectively. The maximum value was set to 19 instead of the expected interval because the 19 variables were correlated. To reduce the collinearity among the variables, the variable interval was initially set to the maximum value and the dimensionality reduction effect of the stepwise screening method was better in setting the parameter to a smaller expected feature interval. SPA was performed to obtain the six characteristic frequencies of lettuce phosphorus.
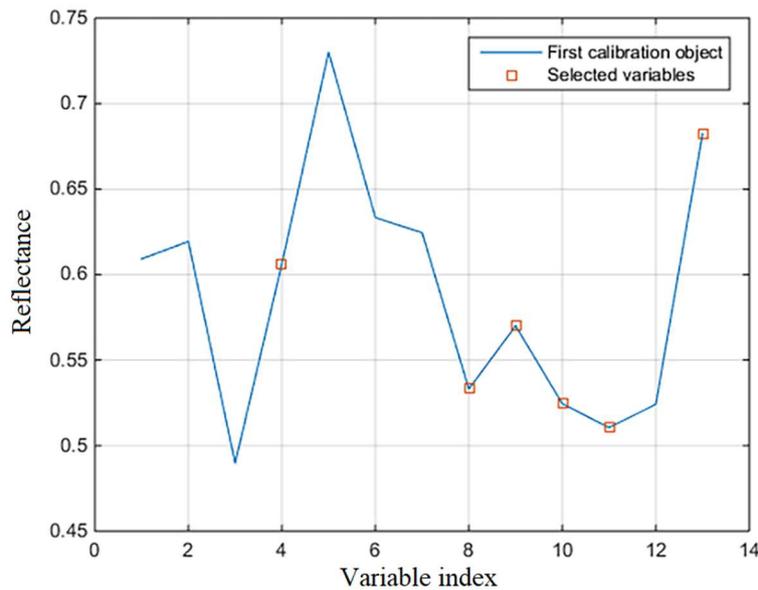

FIGURE. 8 Distribution of the characteristic frequencies of the SPA.

The six selected characteristic frequencies were mostly in the peaks and valleys of the waveform, and the corresponding positions in the entire spectrum were 1.102, 1.934, 2.620, 2.934, 3.311, and 3.38 THz (Figure 8).

Thus, PLSR was used to establish a prediction model for the measured value using the calibration and validation sets of lettuce phosphorus.
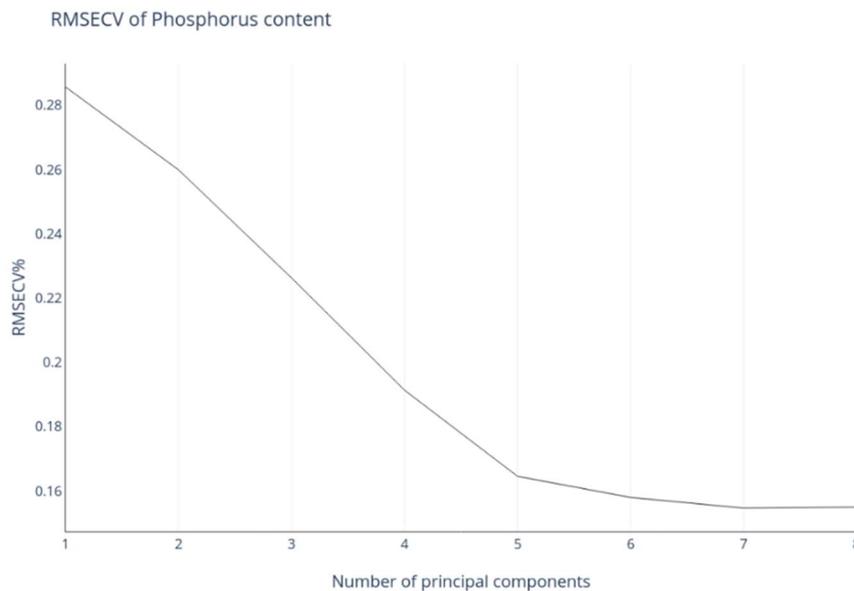

FIGURE. 9 Determination of the optimal number of principal components by PLSR.

The results are presented in Figure 9. Where the number of principal components is five, the RMSECV is small, and the rate at which the error decreases is reduced. The effect of increasing the number of principal components on accuracy was weak. Therefore, the model accuracy and operational efficiency were improved using five principal components. Modelling using PLSR first requires the determination of the principal components. PLSCV was used to determine the principal components based on the principle that a low RMSECV corresponds to a low number of principal components.

PLSR fitting was used. The correction set R^2_c was 0.7705, RMEC was 0.264%, validation set R^2_p was 0.7005, and RMSEP was 0.3273%.
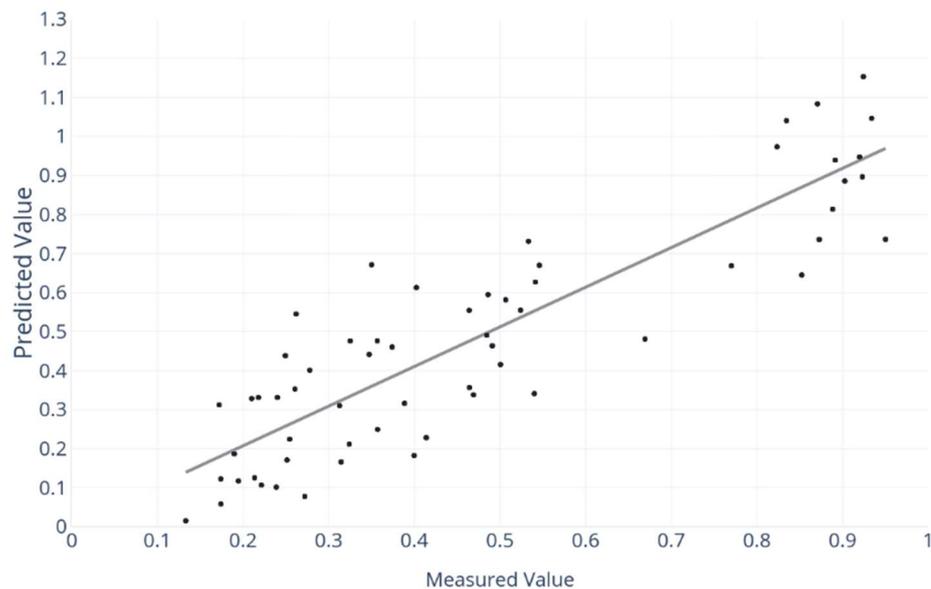
FIGURE. 10 Correlation model between the predicted value and true value of the correction set.
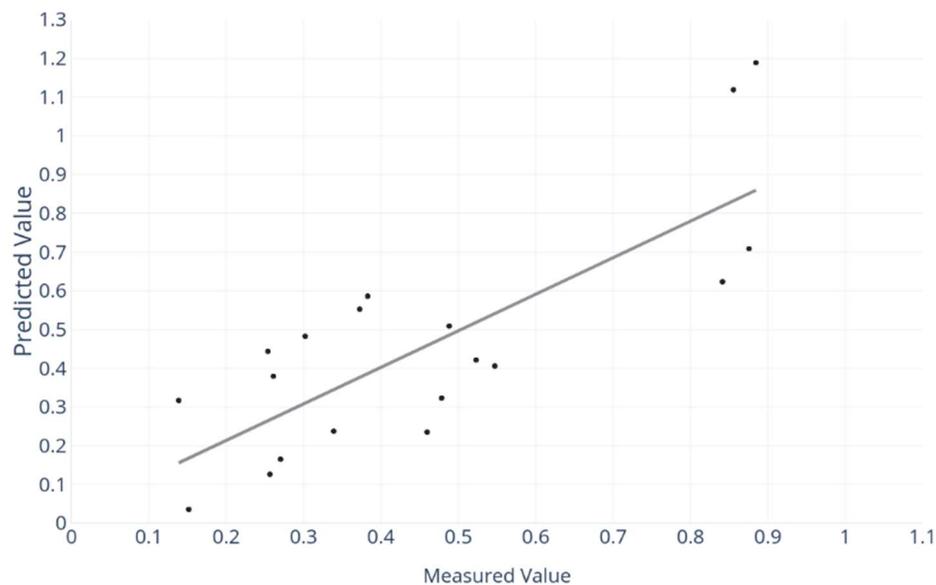

FIGURE. 11 Correlation model between the predicted value and true value of the verification set.

Figures 10 and 11 show scattered plots of the correlation between the predicted values and the true values of the correction and validation set models for lettuce phosphorus content, respectively. The predicted sample points of the dataset were compared. A good distribution on both sides of the fitted line indicates that the prediction model can better predict the phosphorus content in lettuce leaves.

CONCLUSIONS

The quantitative relationship between THz-TDS and the chemical measurement of phosphorus content in lettuce leaves was studied by culturing lettuce samples in nutrient solutions with varying concentrations of phosphorus. The effects of different algorithms on raw data preprocessing and sample set classification were compared, and the optimal combination of algorithms was determined.

The SCARS algorithm combined with the SPA algorithm was used to extract the characteristic frequency of phosphorus, ensuring the accuracy and improving the efficiency of the model. PLSCV was combined with PLSR to establish a detection model for nutritional status of phosphorus in lettuce. The model verified that R^2_p reached 0.7005, and RMSEP was 0.3273%, indicating that it could accurately detect phosphorus in lettuce.

ACKNOWLEDGMENTS

REFERENCES

Araújo MCU, Saldanha TCB, Galvão RKH, Chame HC, Visani V (2001) The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics & Intelligent Laboratory Systems 57(2):65-73.

Coutu RA, Medvedev IR, Petkie DT (2016) Improved sensitivity mems cantilever sensor for terahertz photoacoustic spectroscopy. Sensors 16(2):251

Ge HY, Jiang YY, Lian FY, Zhang Y, Xia SH (2015) Characterization of wheat varieties using terahertz time-domain spectroscopy. Sensors (Basel, Switzerland) 15(6):12560-12572.

Guo LY, Li ZL (2021) Comprehensive Evaluation of Salt Tolerance of Lettuce at Germination Stage Based on Principal Component Analysis. Molecular Plant Breeding 11(8):1569-1573.

Jiang YY, Ge HY, & Zhang Y (2018) Quantitative determination of maltose concentration in wheat by using terahertz imaging. Spectroscopy and Spectral Analysis 38(10):3017-3022.

Jiang YY, Ge HY, Lian FY, Zhang Y, Xia SH (2014) Application of Thz technology to nondestructive detection of agricultural product quality. Spectroscopy and Spectral Analysis 34(8):2047-2052.

Karain W (2016) Thz frequency spectrum of protein–solvent interaction energy using a recurrence plot-based wiener–khinchin method. Proteins Structure Function & Bioinformatics 84(10):1549-57

Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11(1): 137-148.

Li B, Long Y, Liu HS, Zhao CJ (2017) The determination of glucose based on terahertz spectroscopy. Spectroscopy and Spectral Analysis 37(7):2165-2170.

Liu GH, Xia RS, Jiang H, Mei CL, Huang YH (2014) A wavelength selection approach of near infrared spectra based on scars strategy and its application. Spectroscopy & Spectral Analysis 34(8):2094-2097.

Liu W, Zhao PG, Wu CS, Liu CH, Yang JB, Zheng Lei (2019) Rapid determination of aflatoxin B1 concentration in soybean oil using terahertz spectroscopy with chemometric methods. Food Chemistry 293:213-219.

Pagano M, Baldacci L, Ottomaniello A, Dato GD, Corona P,Chianucci F, Masini L, Carelli G, Toncelli A,Storchi P, Tredicucci A, Corona P (2019) Thz water transmittance and leaf surface area: an effective nondestructive method for determining leaf water content. Sensors 19(22): 4838-4849.

Sun J, Jin XM, Mao HP, Wu XH, Zhu WJ, Zhang XD, Gao HY (2014) Detection of nitrogen content in lettuce leaves based on spectroscopy and texture using hyperspectral imaging technology: Transactions of the Chinese Society of Agricultural Engineering 30(10):167-173.

Wang KQ, Sun DW, Pu HB, Wei QY (2017) Principles and applications of spectroscopic techniques for evaluating food protein conformational changes: a review. Trends in Food Science & Technology 67:207-219.

Wu J, Liu YD, Li B, Hu J (2019) Qualitative and quantitative detection of glucose and sucrose in milk powder based on terahertz spectroscopy. Spectroscopy and Spectral Analysis 39(8):2568-2573.

Yun YH, Wang WT, Tan ML, Liang YZ, Li HD, Cao DS, Lu HM, Xu QS (2014) A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. Analytica Chimica Acta 807:36-43.

Zhang DR, Fang H, Yong HE (2019) Research of Crop Disease Based on Visible/Near Infrared Spectral Image Technology: A Review. Spectroscopy and Spectral Analysis 39(6):1748-1756.