Identification of pesticide residue level in lettuce based on hyperspectra and chlorophyll fluorescence spectra

Sun Jun[1,2*], Zhou Xin[1], Mao Hanping[2], Wu Xiaohong[1], Zhang Xiaodong[2], Gao Hongyan[2]

(1. *School of Electrical and Information Engineering of Jiangsu University, Zhenjiang 212013, China*;
2. *Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang 212013, China*)

Abstract: Fast identification of pesticide residue level in lettuce leaves plays a key role in the test of food safety. In order to identify the different concentrations pesticide residues of lettuce leaves in a fast and nondestructive way, the hyperspectra coupled with chlorophyll fluorescence spectra was used in this research. Transmission electron microscopy (TEM) was used to identify the microstructure changes of lettuce leaves under different concentrations of dimethoate residue. Besides, a method involving wavelet transform and MD-MCCV algorithm (WT-MD-MCCV) was developed for identifying the optimal wavelengths of the spectral data. The hyperspectra and chlorophyll fluorescence spectra data of 150 lettuce leaf samples at five different concentrations of pesticide residues were obtained using hyperspectral data acquisition device and Cary Eclipse Fluorescence Spectrophotometer. The combination of Savitzky-Golay (SG) algorithm and SNV algorithm (SG-SNV) preprocessing algorithms was used to preprocess the raw spectra. In addition, Principal Component Analysis (PCA), Successive Projections Algorithm (SPA) and wavelet transform coupled to MD-MCCV algorithm (WT-MD-MCCV) were applied to identify the optimal wavelengths of raw spectra including hyperspectra data, chlorophyll fluorescence spectra data and hyperspectra coupled with chlorophyll fluorescence spectra data. Support vector regression (SVR) was applied to build the prediction models based on preprocessed spectra feature in characteristic wavelengths coupled with different spectral data. The results showed that with the increase of the concentration of dimethoate pesticide spraying, lettuce chloroplast number of osmiophilic particles increased and the starch granules decreased. Besides, the intercellular space of lettuce leaves increased gradually, with the increase of dimethoate concentration. Different concentrations of pesticide residues of lettuce in the near infrared and fluorescence spectrum have a certain difference. In addition, the related parameters of the three preferably prediction models were R_p^2=0.956 and *RMSEP*=0.018, R_p^2=0.937 and *RMSEP*=0.161, R_p^2=0.987 and *RMSEP* =0.005, respectively, using WT-MD-MCCV algorithm combined with hyperspectra data, chlorophyll fluorescence spectra data and hyperspectra coupled to chlorophyll fluorescence spectra data. WT-MD-MCCV algorithm combined with hyperspectra and chlorophyll fluorescence spectra data performed best among the nine SVR models and the hyperspectra coupled with chlorophyll fluorescence spectra can be used to identify the pesticide residue level in lettuce leaves.

Keywords: lettuce, chlorophyll fluorescence spectra, hyperspectra, modeling, pesticide residue
DOI: 10.3965/j.ijabe.20160906.2519

Citation: Sun J, Zhou X, Mao H P, Wu X H, Zhang X D, Gao H Y. Identification of pesticide residue level in lettuce based on hyperspectra and chlorophyll fluorescence spectra. Int J Agric & Biol Eng, 2016; 9(6): 231－239.

1 Introduction

The pesticides, generated in modern agricultural practices, are used widely to protect crops against pests, diseases and weeds. It is regarded as one of the most serious issues to food safety[1].

More attention has been paid to food safety, particularly the problem of pesticide residues on fruits

Received date: 2016-04-09 **Accepted date:** 2016-08-21
Biographies: Zhou Xin, Master, research interests: signal and information processing, Email: 862218958@qq.com; **Mao Hanping,** PhD, Professor, research interests: applications of signal processing in agricultural, Email: maohp@ujs.edu.cn; **Wu Xiaohong,** PhD, research interests: applications of signal processing in agricultural, Email: wxh419@ ujs.edu.cn; **Zhang Xiaodong,** PhD, research interests: applications of signal processing in agricultural, Email: zxd700227@126.com; **Gao Hongyan,** PhD, research interests: applications of signal processing in agricultural, Email: ghy06@126.com.
Corresponding author: Sun Jun, PhD, Professor, research interests: applications of signal processing in agricultural. Mailing address: School of Electrical and Information Engineering of Jiangsu University, Zhenjiang 212013, China. Tel/Fax: +86-511-88791245, Email: sun2000jun@ujs.edu.cn.

and vegetables[2]. Organic pesticide is commonly used for spraying fruits and vegetables[3]. Lettuce is a suitable crop to pursue flavor improvement as it is widely eaten across the world. Besides, it also contains a range of beneficial secondary plant metabolites, including phenolics, ascorbate, α-tocopherol, lignans, as well as SLs[4]. Lettuce is often used for food directly, without using regular special treatment methods. Therefore, it is necessary for the identification of pesticide residue concentration in lettuce leaves during the food processing and storage.

Nowadays, hyperspectral image technology has been more and more widely used with the advantages of being rapid, non-destructive and efficient[5]. It was successfully applied to the non-destructive testing of fruits and vegetables[6-9]. In addition, chlorophyll fluorescence originates from deep within the photosynthetic machinery of the leaf, in the photosystems of the thylakoid membranes[10] and related material transformation may reflect the chlorophyll fluorescence spectra of the plant. Therefore, chlorophyll fluorescence spectra was used for correlation detection of fruits and vegetables[11-13]. However, there were few reports referring to the identification of pesticide residue concentration in lettuce leaves using the hyperspectral imaging technology coupled with chlorophyll fluorescence spectra.

Hyperspectral imaging technology coupled with chlorophyll fluorescence spectra was used to identify the pesticide residue concentration in lettuce leaves in this study. A method involving wavelet transform and MD-MCCV algorithm (WT-MD-MCCV) was developed for identifying the optimal wavelengths of the spectral data. Because the spectral data has the scope of large distribution and a great deal of information, it is difficult to process the spectral data directly. Therefore, spectral preprocessing algorithm (Savitzky-Golay smoothing), feature extraction algorithm (Principal Component Analysis, PCA and Successive Projections Algorithm, SPA) and an optimal prediction model (Support Vector Regression, SVR) for identifying the pesticide residue concentration of lettuce leaves will be found in this study.

2 Materials and methods

2.1 Sample preparation

All the samples of lettuce (Cold butter lettuce seeds Co. Ltd., Shanghai, China) were cultivated under non-soil environment using perlite bag cultivation mode. Planting place was in a Venlo greenhouse at the Laboratory of Modern Agricultural Equipment in Jiangsu University (Zhenjiang of China, 32.11°N, 119.27°E). At the rosette stage, the samples were sprayed using the dimethoate (40% EC, DuPont Chemical Group Co. Ltd., Qingdao, China) solution of five different concentrations, respectively, the ratios of dimethoate pesticides and water were 1:500, 1:800, 1:1000, 1:1200, 1:1600, respectively. Besides, 24 h later after spraying the pesticide, the test samples of lettuce were picked in the same leaf position with full leaf mesophyll due to the existence of safety interval. There were 30 pieces of leaves from 5 different concentrations. Besides, a total of 150 samples of lettuce were stored in different plastic bags with different labels placed in the plant preservation box (at temperature –7°C) and carried to the laboratory for spectral acquisition after being picked off. In addition, the distance between the breeding site and the spectrum acquisition site was 500 m, and the entire cost of time was 5 min.

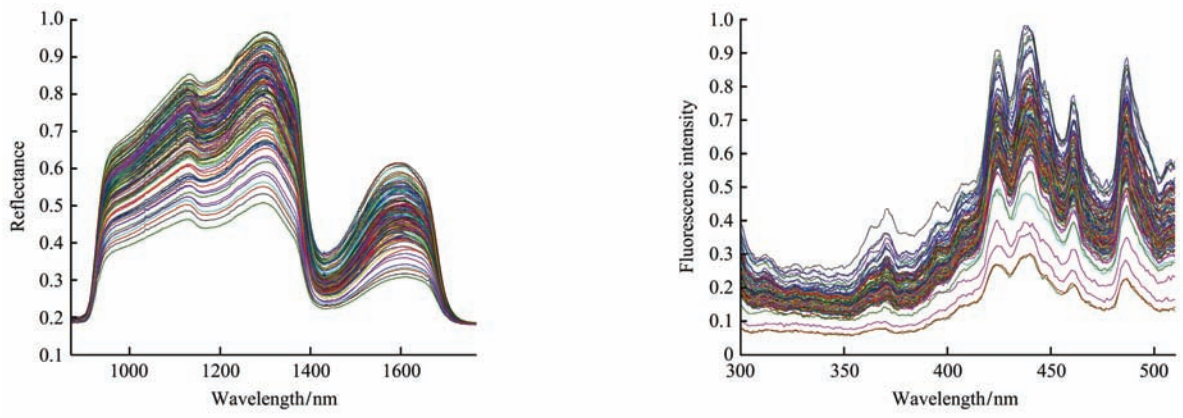
2.2 Hyperspectra acquisition

Hyperspectral data acquisition device consists of an imaging spectrograph with spectral resolution of 2.8 nm (ImSpector V10E, Spectral Imaging Co. Ltd., Oulu, Finland), an illumination unit consisting of two 150 W fiber optical halogen lamps (2900-ER +9596-E, Illumination, USA), camera obscura (SC100, Beijing optical instrument factory, China) and electric displacement platform (MTS120, Beijing optical instrument factory, China). In order to reduce the error, the test of standard reflective plate and a black background were conducted before the acquisition of lettuce sample spectral information. There were 256 spectral channels ranging from 870 nm to 1780 nm (±4.42 nm, FWHM). In this experiment, pixel size of 128×128 blade uniform rectangular area was selected as the region of interest (ROI) and the average spectra of ROI were set as the value of the lettuce sample. The

hyperspectra of all samples (870-1780 nm) is shown in Figure 1a.

2.3 Chlorophyll Fluorescence Spectra acquisition

Chlorophyll fluorescence spectra were collected using fluorescence spectrophotometer (Cary Eclipse, Varian Technology Co. Ltd., USA). Instrument adopts flashing xenon lamp as excitation light. The optimum fluorescence spectrum information was obtained by adjusting the measuring condition of the instrument blade clamping device. In addition, the fluorescence spectrophotometric parameters testing meter was set as follow: the excitation light source wavelength was set to 245 nm, the range of emission wavelength was 300-510 nm (±1.07 nm, FWHM), the excitation and emission slits were 5 nm, scanning speed was set as medium, voltage was set to 600 V. The selected area of the fluorescence spectra of lettuce was the same as the ROI of the hyperspectral data. Besides, the measurement of each selected area was conducted with three replications and the average value was taken as the final measurement results. The fluorescence spectrum of all samples (300-510 nm) is shown in Figure 1b.



a. Hyperspectra of all the lettuce leaf samples (870-1780 nm) b. Chlorophyll fluorescence spectra of all the lettuce leaf samples (300-510 nm)

Figure 1 Original spectra of all the lettuce leaf samples

2.4 Detection mechanism and chemical detection

The near infrared and fluorescence diffuse reflectance spectra of the plants were related to the microstructure of the plants[14]. The microstructure changes of lettuce leaves under different concentrations of dimethoate residue were investigated by TEM (JEM-1400, Japan Electronic Co., Ltd., Japan). In the TEM instrument, the acceleration voltage was 40-100 kV, the amplification factor was 50-1 000 000 times, the point resolution reached 0.38 nm, the line resolution reached 0.2 nm. Besides, the TEM detection experiment was carried out after the completion of the experiment of spectrum acquisition. Liquid chromatography with tandem mass spectrometry (LC-MS/MS) was used to the detection of dimethoate pesticide residues in lettuce leaves. The specific detection steps of LC-MS/MS were referred to the national standard GB/T 20769-2008[15]. In addition, the five different concentrations of pesticide residues in lettuce (the ratio of dimethoate pesticides and water were 1:500, 1:800, 1:1000, 1:1200, 1:1600) detected by LC-MS/MS method were 1.02 mg/mL, 0.64 mg/mL, 0.51 mg/mL, 0.42 mg/mL, 0.32 mg/mL, respectively.

2.5 Spectral preprocessing method

In the process of spectral analysis, choosing the appropriate smoothing algorithm can effectively eliminate the noise interference. The Savitzky-Golay (SG) smoothing algorithm is a low-pass filter well adapted to smoothing noisy data[16]. Standard Normalized Variable (SNV) algorithm is able to effectively correct the spectral difference between samples due to scattering[17]. The combination of SG and SNV algorithms (SG-SNV) can effectively achieve the standard normalization of spectral data and the removal of noise, and better retain the effective information in the spectral information. Therefore, SG-SNV was used to process the spectral data including hyperspectra and chlorophyll fluorescence spectra in this research.

2.6 PCA algorithm

Hyperspectra and chlorophyll fluorescence spectra

data were a set of multivariate data, including redundant information. The PCA, spectral data dimensionality reduction algorithm, was used to reduce the dimension of spectral data and extract the characteristic bands in this research[18]. The main viewpoint of PCA was to use fewer variables to explain most of the variation in the original data, and many highly relevant variables were changed into those independent or uncorrelated[19].

2.7 SPA algorithm

The successive projections algorithm (SPA) was used for topic modeling, document clustering and hyperspectral image unmixing[20]. Successive projections algorithm was proposed to select a subset of variables for multivariate calibration[21] and it was commonly used to select characteristic bands for high dimensional data. Thus, SPA was used for feature selection in the spectral data including hyperspectra and chlorophyll fluorescence spectra in this research.

2.8 WT-MD-MCCV algorithm

In this research, WT-MD-MCCV was proposed for extracting features after spectral preprocessing. Furthermore, there are effective combination of wavelet transform (WT) and MD-MCCV. MD-MCCV is algorithm for calculating the related parameters of Mahalanobis distances between samples based on the multi modeling of Monte Carlo cross validation algorithm[22]. WT-MD-MCCV is a relatively simple algorithm which can be summarized as the following steps:

Step 1: Input spectral preprocessing sample set $S=\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$ and concentration value set $L=\{(s_1,v_1),(s_2,v_2),...,(s_n,v_n)\}$, where $x_i \in X$, $y_i \in Y$, $s_i \in S$, $v_i \in V$, X is the value of sample wavelength. Y is the spectral preprocessing values of the sample data. V is concentration values. $v_i \in \{1.02, 0.64, 0.51, 0.42, 0.32\}$, $i=1,2,3,...,n$.

Step 2: The matrix T is obtained for the average processing of each gradient sample set in the original sample set S. Where $T=\{(st_1,vt_1),(st_2,vt_2),...,(st_m,vt_m)\}$, $m=1,2,3,4,5$. $st_i \in ST$, $vt_i \in VT$, ST is the sample after mean treatment. VT is the average value of different concentration sample.

Step 3: The matrix T was conducted seven layers of wavelet decomposition based on db6 (Daubechies 6) wavelet basis function. The high-frequency wavelet coefficient of the seven layers was stored in a matrix CD. Where $CD=\{(x_1,c_1),(x_2,c_2),...,(x_7,c_7)\}$, $x_i \in X$, $c_i \in C$, $i=1,2,3,...,7$. C is the high-frequency wavelet coefficient of the different layers.

Step 4: The matrix S was conducted seven layers of wavelet decomposition based on db6 (Daubechies 6) wavelet basis function. The high-frequency wavelet coefficient of the seven layers was stored in a matrix HD and the low-frequency wavelet coefficient of the seven layers was stored in a matrix LD.

Step 5: Obtain the best decomposition layer and extract the best wavelength.

(1) The Markov distance was calculated between samples of different decomposition layers in the matrix CD. The formulation of Markov distance is as following:

$$MD_i^2 = (x_i - \overline{x})C^{-1}(x_i - \overline{x})^T \quad (1)$$

$$C = \frac{1}{(m-1)}(X_c)^T(X_c) \quad (2)$$

where, MD_i is the Markov distance of the i set of matrix; x_i is the i set of matrix; $\overline{x}$ is the mean of matrix in Equation (1). Moreover, C is the covariance matrix, X_c is the centralized data matrix, m is the number of concentration in Equation (2).

(2) Calculate the mean and variance of the Markov distance of each sample in step 5 (1) on different decomposition level. In addition, the best decomposition layer DL was obtained by evaluating and sorting the maximum variance corresponding decomposition layer.

(3) Under the optimal decomposition level DL, MD-MCCV was used to the corresponding Markov distance of each sample in different wavelengths based on the matrix CD. The formulation of Markov distance is same as Equations (1) and (2). Besides, m is the cycle times of MCCV in Equation (2). The mean and variance of the Markov distance between samples of different decomposition layers were calculated and the best wavelengths were extracted by evaluating and sorting it.

(4) The original set will be replaced by the new data ND in matrix LD corresponding to the optimal

decomposition level DL. The best wavelengths data will be extracted from the new data ND.

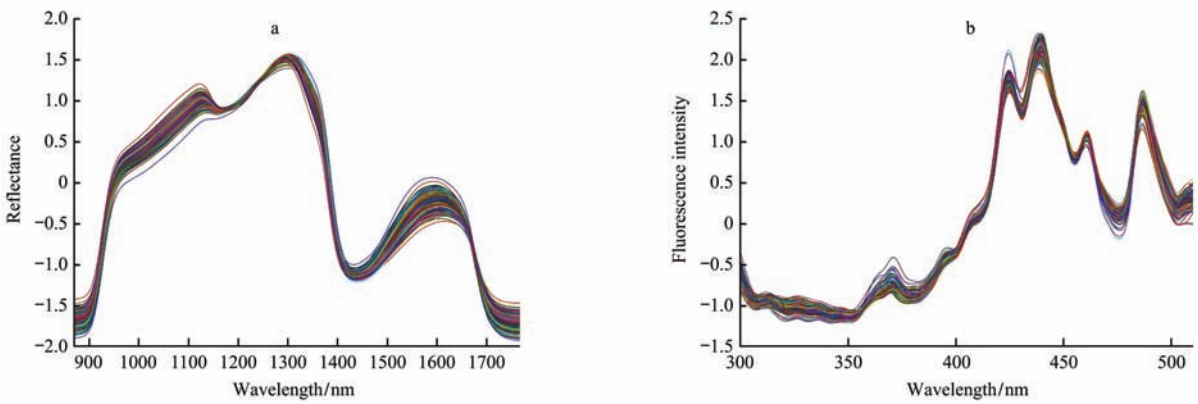
2.9 Modeling algorithm

Support vector regression (SVR), a regression version of support vector machines (SVM), was proposed to solve nonlinear regression problems with a maximum margin algorithm[23]. It is an efficient and robust method and provides high generalisability and performance[24]. Therefore, the SVR algorithm was used to establish the analysis model using the characteristic wavelengths in this research. Furthermore, the performances of the proposed model are evaluated by using the following indices: coefficient of determination for calibration R_c^2, root mean square error for calibration RMSEC, coefficient of determination for validation R_{CV}^2, root mean square error for validation RMSECV, coefficient of determination for prediction R_p^2, root mean square error for prediction RMSEP. Moreover, in the case that the smaller the predicted residual sum of squares and the root mean square error are, as well as the greater the determination coefficient is, the better the performance of the model will be.

3 Results and discussion

3.1 Spectral pretreatment

In order to realize the normalization of spectral data, eliminate spectral noise and improve signal to noise ratio (SNR), SG-SNV algorithm was used to deal with spectral data and the processed spectral curves are shown in Figure 2. This method not only can improve the SNR and realize the normalization of spectral data, but also can keep the useful information of the spectral information.



Note: (a) and (b) show the hyperspectra and the chlorophyll fluorescence spectra of all the lettuce leaf samples preprocessed by SG-SNV, respectively.

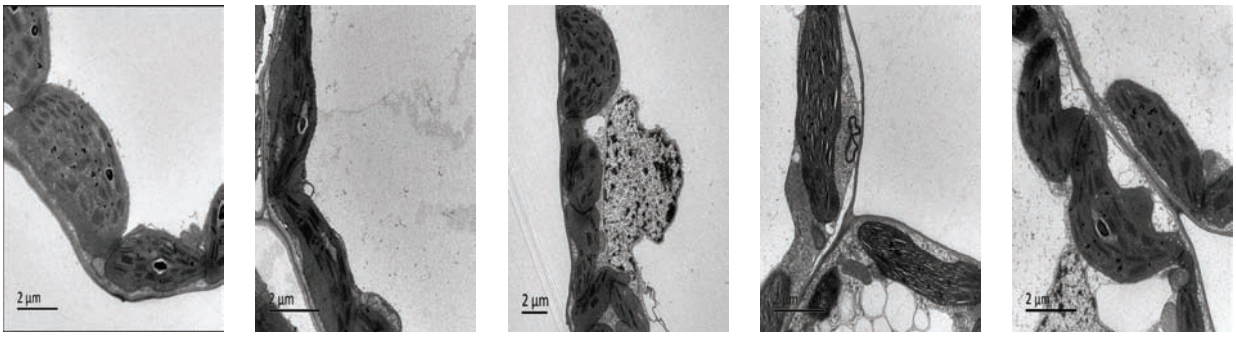
Figure 2 Spectral curves of lettuce leaf samples after SG-SNV preprocessing method

3.2 Microstructure of lettuce samples

Lettuce leaf samples need to be washed, fixed, dehydrated, soaked, embedded, aggregated and ultrathin sections before the experiment of TEM. First, 1 mm× 2 mm lettuce leaf samples were washed three times by using pH 7.0 phosphate buffer, each time was 20 min. Then, the lettuce leaves were placed in 2% osmic acid solution until the color black. After that, these samples were cleaned again with the pH 7.0 phosphate buffer, each time was 20 min. Besides, acetone was used to deal with the dehydration of lettuce samples using four concentrations (30%, 50%, 70% and 90%) and 100% acetone was used to deal with the dehydration of lettuce samples twice, each time was 30 min. Moreover, the embedding agent was used to soak the samples of the lettuce leaves, and the embedding was carried out in a constant temperature chamber for overnight polymerization using three different temperatures (37°C, 45°C, and 60°C). Finally, the 50-70 nm thickness of lettuce leaves was obtained by using the RMC thin slice machine (PowerTome-XL, American RMC Corporation, United States), and the lettuce leaves were stained with uranium lead method. Finally, the microstructure of 1 mm × 2 mm lettuce leaves was observed by TEM and the structure was observed as shown in Figure 3.

Lettuce cell mainly consists of chloroplast, osmiophilic particles and cell wall. After spraying different concentrations of dimethoate pesticide, lettuce chloroplast morphology and internal structure were not occurred larger changed, suggesting that the organization of internal structure of dimethoate infringement has certain resistance. From Figures 3a-3d, it can be seen

that the cell lines were clear and rectangular in shape. With the increase of the concentration of dimethoate pesticide spraying, lettuce chloroplast number of osmiophilic particles increased and the starch granules decreased. Besides, the intercellular space of lettuce leaves gradually increased, with the increase of dimethoate concentration. Therefore, different concentrations of pesticide residues of lettuce have a certain difference in the near infrared and fluorescence spectrum.



a b c d e

Note: a, b, c, d and e represent the cell structure of lettuce leaves sprayed with 1:1600, 1:1200, 1:1000, 1:800 and 1:500 concentrations of dimethoate, respectively.

Figure 3 Cell structure of lettuce leaves with different concentrations of dimethoate

3.3 Identification of optimum wavelengths

In this research, three different algorithms were used to identify the optimal wavelengths of the hyperspectra data, the chlorophyll fluorescence spectra data and the hyperspectra data coupled with chlorophyll fluorescence spectra data of all the lettuce leaf samples by PCA, SPA and WT-MD-MCCV, respectively.

PCA is one of the most commonly methods used feature extraction algorithm and its cumulative contribution rate of the hyperspectra, the chlorophyll fluorescence spectra and the hyperspectra data coupled with chlorophyll fluorescence spectra data were shown in Table 1. From Table 1, it can be seen that the principal components of hyperspectra data, the chlorophyll fluorescence spectra data and the hyperspectra data coupled with chlorophyll fluorescence spectra data selected by PCA as 5, 6 and 6, the cumulative contribution rates were 99.218%, 99.0162% and 99.277%. Therefore, in order to reduce the data redundancy and retain the maximum amount of information, the principal components of 5, 6 and 6 were selected for the hyperspectra data, the chlorophyll fluorescence spectra data and the hyperspectra data coupled with chlorophyll fluorescence spectra data.

In recent years, SPA was widely used to select feature according to root mean square error (RMSE). The results of the optimum wavelengths selected by SPA of the hyperspectra, the chlorophyll fluorescence spectra and the hyperspectra data coupled with chlorophyll fluorescence spectra data are shown in Table 2, based on the principle that the RMSE was relatively low and the number of wavelength was relatively small.

WT-MD-MCCV was proposed for extracting features after spectral preprocessing in this research. The optimal decomposition layers and the extracted characteristic wavelengths obtained by using WT-MD-MCCV algorithm were shown in Table 3.

3.4 SVR models

In this research, about 80% of the spectral optimal wavelength data was selected as the calibration set and the remaining 20% will be used as the prediction set. The SVR models were established for the calibration set and the prediction set was used to validate the model. The results of SVR models are presented in Table 4.

Table 1 Cumulative contribution rate using different numbers of principal component

Principal component		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Cumulative contribution rate/%	H	92.7099	97.0160	98.0653	98.7164	99.2180	99.4543	99.6351	99.7827
	C	80.2829	87.6327	91.2172	93.9331	97.4031	99.0162	99.3621	99.5417
	H+C	86.2955	93.1708	96.0544	98.1598	98.9516	99.2770	99.5400	99.6331

Note: H, C and H+C represent the hyperspectra data, the chlorophyll fluorescence spectra data and the hyperspectra data coupled with chlorophyll fluorescence spectra data of all the lettuce leaf samples, respectively.

Table 2 Characteristic wavelengths extracted by SPA

The type of data	RMSE	No.	Characteristic wavelengths/nm
H	0.1037	5	902.13, 931.79, 1207.45, 1433.02, 1673.29
C	0.2064	5	315, 366, 441.06, 460, 483.03
H+C	0.1375	6	375.07, 906.42, 1111.96, 1536.61, 1750.24, 1766.32

Note: (1) H, C and H+C represent the hyperspectra data, the chlorophyll fluorescence spectra data and the hyperspectra data coupled with chlorophyll fluorescence spectra data of all the lettuce leaf samples, respectively. (2) No. is the number of characteristic wavelengths.

Table 3 Characteristic wavelengths selected by WT-MD-MCCV

The type of data	ODL	No.	Characteristic wavelengths/nm
H	4	6	1281.86, 1285.05, 1288.22, 1351.26, 1354.40, 1357.54
C	5	4	383.07, 385.07, 418.93, 425.07
H+C	5	5	416.06, 418.03, 1281.86, 1297.74, 1304.06

Note: (1) H, C and H+C represent the hyperspectra data, the chlorophyll fluorescence spectra data and the hyperspectra data coupled with chlorophyll fluorescence spectra data of all the lettuce leaf samples, respectively. (2) No. is the number of characteristic wavelengths. (3) ODL is the optimal decomposition layer.

Table 4 Comparison of the results of SVR for different feature extraction methods

Feature extraction method	The type of data	R_c^2	R_{cv}^2	R_p^2	*RMSEC*	*RMSECV*	*RMSEP*
PCA	H	0.897	0.841	0.689	0.040	0.137	0.267
	C	0.854	0.823	0.631	0.236	0.361	0.380
	H+C	0.932	0.856	0.696	0.030	0.131	0.165
SPA	H	0.935	0.885	0.783	0.138	0.256	0.339
	C	0.927	0.848	0.732	0.058	0.312	0.458
	H+C	0.978	0.901	0.811	0.132	0.227	0.301
WT-MD-MCCV	H	0.983	0.977	0.956	0.004	0.009	0.018
	C	0.978	0.969	0.937	0.010	0.121	0.161
	H+C	0.997	0.990	0.987	0.008	0.005	0.005

Note: H, C and H+C represent the hyperspectra data, the chlorophyll fluorescence spectra data and the hyperspectra data coupled with chlorophyll fluorescence spectra data of all the lettuce leaf samples, respectively.

Table 4 shows that different identify the optimal wavelength algorithms have some effect on the performance of the proposed model. Part of the identification of optimal wavelength algorithm can improve the performance of the model, but also some of the algorithm reduces the performance of the model. From the prediction accuracy, the prediction accuracy of WT-MD-MCCV algorithm of the model is higher than the PCA algorithm, SPA algorithm. In addition, the order of the different spectral is that the hyperspectra data coupled with chlorophyll fluorescence spectra better than hyperspectra data and chlorophyll fluorescence spectra in the establishment of a model. Among them, the establishment of the SVR model has best prediction ability of the concentration of pesticide residues in lettuce using the hyperspectra data coupled with chlorophyll fluorescence spectra.

4 Conclusions

The hyperspectra and chlorophyll fluorescence spectra data of five different concentrations pesticide residues in lettuce leaves were obtained by hyperspectral data acquisition device and Cary Eclipse Fluorescence Spectrophotometer. Besides, the microstructure changes of lettuce leaves under different concentrations of dimethoate residue were investigated by TEM. The results showed that with the increase of the concentration of dimethoate pesticide spraying, lettuce chloroplast number of osmiophilic particles increased and the starch granules decreased. Besides, the intercellular space of lettuce leaves gradually increased, with the increase of dimethoate concentration. Therefore, different concentrations of pesticide residues of lettuce in the near infrared and fluorescence spectrum have a certain difference. A method involving WT and MD-MCCV algorithm was developed for identifying the optimal wavelengths of the hyperspectra, chlorophyll fluorescence spectra data and hyperspectra data coupled with chlorophyll fluorescence spectra data. In this research, the combination of SG-SNV preprocessing

algorithms was used to process the spectral data. Then, PCA, SPA and WT-MD-MCCV were applied to the identification of the optimum wavelengths of spectral data. Finally, nine different combinations of these kinds of spectral wavelengths were used as the input of SVR to establish the model of predicting the concentrations of pesticide residues in lettuce leaves. By comparing the indices of the nine forecasting models, the results showed that the WT-MD-MCCV can identify the optimal wavelengths of spectra and get the best prediction effect compared with PCA and SPA. Among them, the best prediction model was WT-MD-MCCV based on hyperspectra data coupled with chlorophyll fluorescence spectra data (R_p^2=0.987 and *RMSEP*=0.005). These results indicated that WT-MD-MCCV algorithm performed best among PCA algorithm, SPA algorithm in identifying the optimal wavelengths of the hyperspectra, chlorophyll fluorescence spectra data and hyperspectra data coupled with chlorophyll fluorescence spectra data from five different concentrations pesticide residues of lettuce leaves, and the performance of SVR model based on hyperspectra data coupled with chlorophyll fluorescence spectra was better than the hyperspectra and chlorophyll fluorescence spectra.

Acknowledgements

This work was partially supported by National Natural Science Funds Projects (31471413), project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Six Talent Peaks Project in Jiangsu Province (ZBZZ-019), Natural Science Foundation of Jiangsu Province of China (BK20141165), the Key Laboratory of Modern Agricultural Equipment and Technology of Jiangsu University (NZ201306).

[References]

[1] Shamsipur M, Yazdanfar N, Ghambarian M. Combination of solid-phase extraction with dispersive liquid–liquid microextraction followed by GC–MS for determination of pesticide residues from water, milk, honey and fruit juice. Food Chemistry, 2016; 204: 289–297.

[2] He Q, Huang J, Yang X, Yan X, He J, Li S, et al. Effect of pesticide residues in grapes on alcoholic fermentation and elimination of chlorothalonil inhibition by chlorothalonil hydrolytic dehalogenase. Food Control, 2016; 64: 70–76.

[3] Cao X, Song H L, Yu C Y, Li X N. Simultaneous degradation of toxic refractory organic pesticide and bioelectricity generation using a soil microbial fuel cell. Bioresource Technology, 2015; 189: 87–93.

[4] Chadwick M, Gawthrop F, Michelmore R W, Wagstaff C, Methven L. Perception of bitterness, sweetness and liking of different genotypes of lettuce. Food Chemistry, 2016; 97: 66–74.

[5] He H J, Sun D W. Hyperspectral imaging technology for rapid detection of various microbial contaminants in agricultural and food products. Trends in Food Science & Technology, 2015; 46: 99–109.

[6] Teena M, Manickavasagan A, Ravikanth L, Jayas D S. Near infrared (NIR) hyperspectral imaging to classify fungal infected date fruits. Journal of Stored Products Research, 2014; 59: 306–313.

[7] Siripatrawan U, Makino Y. Monitoring fungal growth on brown rice grains using rapid and non-destructive hyperspectral imaging. International Journal of Food Microbiology, 2015; 199: 93–100.

[8] Serranti S, Gargiulo A, Bonifazi G. Hyperspectral imaging for process and quality control in recycling plants of polyolefin flakes. J. Near Infrared Spectrosc., 2012; 20: 573–581.

[9] Mo C, Kim G, Lim J, Kim M S, Cho H, Cho B K. Detection of lettuce discoloration using hyperspectral reflectance imaging. Sensors, 2015; 15: 29511–29534.

[10] Atherton J, Nichol C J, A. Porcar-Castell. Using spectral chlorophyll fluorescence and the photochemical reflectance index to predict physiological dynamics. Remote Sensing of Environment, 2016; 176: 17–30.

[11] Gameiro C, Utkin A B, Cartaxana P, Marques da S J, Matos A R. The use of laser induced chlorophyll fluorescence (LIF) as a fast and non-destructive method to investigate water deficit in Arabidopsis. Agricultural Water Management, 2016; 164: 127–136.

[12] Analia I, Gavin D, Alicia F C, Maria G L. Effect of arsenic on reflectance spectra and chlorophyll fluorescence of aquatic plants. Chemosphere, 2015; 119: 697–703.

[13] Jochem V, Christiaan V T, Federico M, Neus S, Juan P R, Gina M, et al. Evaluating the predictive power of sun-induced chlorophyll fluorescence to estimate net photosynthesis of vegetation canopies: A SCOPE modeling study. Remote Sensing of Environment, 2016; 176: 139–151.

[14] Ragupathi C, Vijaya J J, Narayanan S, Jesudoss S K, John K L. Highly selective oxidation of benzyl alcohol to benzaldehyde with hydrogen peroxide by cobalt aluminate

catalysis: A comparison of conventional and microwave methods. Ceramics International, 2015; 41(2): 2069–2080.

[15] GB/T 20769-2008. Determination of 450 pesticides and related chemicals residues in fruits and vegetables-LC-MS-MS method, China. (in Chinese)

[16] Staggs J E J. Savitzky–Golay smoothing and numerical differentiation of cone calorimeter mass data. Fire Safety Journal, 2005; 40: 493–505.

[17] Guo Y, Ni Y N, Kokot S. Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2016; 153: 79–86.

[18] Oh J Y, Nojun K. Generalized mean for robust principal component analysis. Pattern Recognition, 2016; 54: 116–127.

[19] Sun J, Jin X M, Mao H P, Wu X H. Identification of lettuce storage time based on spectral preprocessing technology and PCA+SVM. Journal of Pure and Applied Microbiology, 2013; 7: 747–752.

[20] Tomohiko M. Robustness analysis of preconditioned successive projection algorithm for general form of separable NMF problem. Linear Algebra and its Applications, 2016; 497: 1–22.

[21] Liu K, Chen X J, Li L M, Chen H L, Ruan X K, Liu W B. A consensus successive projections algorithm–multiple linear regression method for analyzing near infrared spectra. Analytica Chimica Acta, 2015; 858: 16–23.

[22] Sun J, Zhou X, Wu X H, Zhang X D, Li Q L. Identification of moisture content in tobacco plant leaves using outlier sample eliminating algorithms and hyperspectral data. Biochemical and Biophysical Research Communications, 2016; 471: 226–232.

[23] Kang P, Kim D, Cho S. Semi-supervised support vector regression based on self-training with label uncertainty: An application to virtual metrology in semiconductor manufacturing. Expert Systems with Applications, 2016; 51: 85–106.

[24] Volkan U, Huseyin S. Quantitative prediction of peptide binding affinity by using hybrid fuzzy support vector regression. Applied Soft Computing, 2016; 43: 210–221.