A novel wavelength selection strategy for chlorophyll prediction by MWPLS and GA

Haojie Liu[1], Minzan Li[1,2], Junyi Zhang[1], Dehua Gao[1], Hong Sun[1*], Man Zhang[1], Jingzhu Wu[3]

(1. *Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China*; 2. *Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing 100083, China*; 3. *Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China*)

Abstract: The research proposed a novel wavelength selection strategy by the combination of moving window partial least squares (MWPLS) and genetic algorithm (GA) for the chlorophyll content detection of winter wheat canopy using spectroscopy technology. Firstly, the original spectral dataset was pre-processed by wavelet denosing, multiple scatter correction. Then, abnormal data samples were removed by Pauta Criterion and the dataset was divided into modeling set and validation set by SPXY. Finally, the sensitive wavebands were selected using MWPLS method and MWPLS+GA respectively and partial least squares (PLS) models were established for chlorophyll content prediction. For the model established by using all the wavebands in the region of 400-900 nm, its R_c^2 and R_v^2 were 0.4468 and 0.3821 respectively; its modeling root mean square error (RMSEM) and verification root mean square error (RMSEV) were 2.9057 and 1.7589 respectively. For the model established by using 151 wavebands selected by MWPLS, its R_c^2 and R_v^2 were 0.6210 and 0.5901 respectively; its RMSEM and RMSEV were 2.4007 and 1.6408 respectively. For the model established by using 36 wavebands selected by MWPLS+GA, its R_c^2 and R_v^2 were 0.7805 and 0.7497 respectively; its RMSEM and RMSEV were 1.8504 and 1.1315 respectively. The results show that wavelength selection can remove redundant information and improve model performance. The strategy of combining MWPLS with GA has also been proved to work well in selecting sensitive wavebands for chlorophyll content prediction.

Keywords: MWPLS, GA, canopy spectral reflectance, Chlorophyll content prediction

DOI: 10.25165/j.ijabe.20191205.4033

Citation: Liu H J, Li M Z, Zhang J Y , Gao D H , Sun H , Zhang M, et al. A novel wavelength selection strategy for chlorophyll prediction by MWPLS and GA. Int J Agric & Biol Eng, 2019; 12(5): 149–155.

1 Introduction

Chlorophyll content is an important indicator to diagnose crop growth status, and it provide important reference for crop nitrogen management[1,2]. As a fast and non-destructive tool for component analysis, the spectroscopy technology has been widely used for crop nitrogen detection[3,4].

With a modern spectroscopic instrument, the common feature of the obtained data is that there tend to be numerous variables but measured on much fewer samples. Moreover, the canopy spectral characteristics of wheat plants are often influenced by the lighting conditions and the canopy structure. There exist multicollinearity and overlap among spectral data. Thus, variables which contain much information are necessary and important to be selected for generating stable models with better interpretability and lower prediction error[5,6]. A large number of selection methods have been developed and can be mainly divided into two types, the spectral intervals selection and individual wavebands selection[7].

The moving window partial least squares (MWPLS) method is a wavelength interval selection method and each selected interval consists of a number of continuous variables. MWPLS builds a series of partial least squares (PLS) models in a window that moves over the whole spectral region and then locates useful spectral intervals in terms of the least complexity of PLS models reaching a desired error level. The main advantage of MWPLS is that the model is very stable against the interference from non-composition related factors[8]. However, in the process of such interval selection methods, some noises may be retained while some really useful information may be dropped, since they are always mixed together[9]. As a result, the methods that can select the full spectrum point-by-point are still needed to be developed.

As an individual wavebands selection method, the genetic algorithm (GA) was invented by simulating the natural selection and genetic mechanism of the biological evolutionary process, as demonstrated by Charles Darwin's "survival of the fittest". GA has been widely used in wavebands selection[10,11], because of the parallelism, adaptability and global optimization ability. But GA is sensitive to the scale of the search space. When the search space is small, the algorithm is relatively stable and easy to search for the optimal solution. When the search space is large, the

Received date: 2018-09-25 **Accepted date:** 2019-04-21

Biographies: Haojie Liu, PhD candidate, research interests: agricultural informatization, Email: hjliu267@163.com; **Minzan Li**, PhD, Professor, research interests: agricultural engineering, Email:limz@cau.edu.cn; **Junyi Zhang**, PhD candidate, research interests: agricultural informatization, Email: junyizh@cau.edu.cn; **Dehua Gao**, PhD candidate, research interests: agricultural informatization, Email:751723071@qq.com; **Man Zhang**, PhD, Professor, research interests: agricultural informatization, Email: cauzm@cau.edu.cn; **Jingzhu Wu,** PhD, Associate Professor, research interests: agricultural engineering, Email:pubwu@163.com.

***Corresponding author: Hong Sun,** PhD, Associate Professor, research interests: agricultural informatization. Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China. Tel: +86-10-62737838, Email: sunhong@cau.edu.cn.

algorithm can not find the optimal solution due to bad convergence[12,13].

Therefore, the research proposed to use the combination of the MWPLS and GA to select sensitive wavebands for chlorophyll content prediction. Firstly, the MWPLS method was used to select the sensitive waveband intervals from the full spectrum. Secondly, GA was used to select sensitive individual wavebands from the intervals determined by MWPLS. And finally PLS models were established based on the selected individual wavebands to estimate the chlorophyll content in the field.

2 Materials and methods

2.1 Field experiment

The experiment was conducted in Xiaotangshan Experimental Station of China Agricultural University (CAU) in Beijing Changping District on May 13, 2016. The weather was sunny and suitable for spectral reflectance measurement. The wheat was at heading stage. Plant height was approximately 50-60 cm.

The canopy spectral reflectance was measured from 10:00 am to 14:00 pm. The spectral range of the handheld spectrometer (ASD, FieldSpecHH) was 325-1075 nm with 2 nm sampling interval, and the field of view is 25°. The field area was 100 m × 30 m and was divided equally into 70 sample plots. The spectral reflectance measurements were taken randomly at four sites in each plot at a height of 50 cm above the canopy. The average of measurements was used to represent the canopy reflectance of each plot.

After spectral reflectance measurement, the corresponding leaves were cut and packed into sealed bags, and then were carried back to the laboratory for determination of chlorophyll content.

2.2 Chlorophyll content measurement

The chlorophyll content was measured by using an ultraviolet spectrophotometer. Firstly, the surface of winter wheat leaves was cleared and the main stems were removed. Secondly, the leaves were cut and mixed evenly. A sample of 0.4 g leaves was soaked into the 25 mL mixed liquor of 99% acetone and absolute ethanol with the proportion of 2:1 for 24 h. In the process of soaking, the mixed liquor was shaken every 8 h to help the chlorophyll extraction. The absorbance data in 645 nm and 663 nm were determined by spectrophotometer. The chlorophyll content of leaves was calculated according to the following two equations:

$$C_a=12.72A_{663}-2.59A_{643} \quad (1)$$

$$C_b=22.88A_{645}-4.67A_{663} \quad (2)$$

where, A_{645} and A_{663} are the absorbance of 645 nm and 663 nm respectively. C_a and C_b are chlorophyll-a and chlorophyll-b content (mg/L) respectively. The total amount of chlorophyll content C_t equals to C_a plus C_b.

2.3 Data processing

The data were processed following the flow chat, which was shown in Figure 1. The process mainly includes data pretreatment, dataset division, and sensitive wavebands selection and modeling. The pretreatment consists of wavelet denosing, multiple scatter correction (MSC) and abnormal data removal. After pretreatment, the remaining samples were divided into modeling set and validation set by using the SPXY method. The sensitive spectral ranges and wavebands were selected by using MWPLS and MWPLS+GA respectively. And finally, PLS models for chlorophyll content prediction were built to evaluate the performance of the new proposed waveband selection methods.

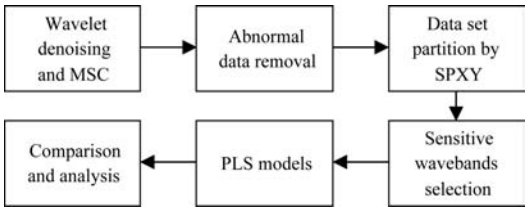

Figure 1 Flow chart of data processing

2.3.1 Wavelet denoising and MSC

The basic idea behind wavelet denoising is that the wavelet transformation leads to a sparse representation for real-world signals. Signals are always concentrated in a few large-magnitude wavelet coefficients, while typically noise corresponding to the wavelet coefficients which are small in value. It means, if an appropriate threshold is set, the wavelet coefficients which are greater than the threshold corresponds to the true signal and should be retained; the wavelet coefficients which are less than the threshold corresponds to the noise and should be set to 0. Thus the noise can be eliminated by reconstructing the signal using the inverse wavelet transform. The process of wavelet denoising is shown in Figure 2.

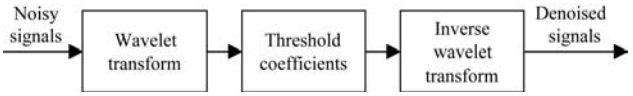

Figure 2 Flow chart of wavelet denoise

In the previous research[14], the wavelet denoising effect by different wavelet basis functions and different decomposition layers has been quantitatively evaluated using two indicators, relative signal-to-noise ratio (RSNR) and curve smoothness (CS). The calculations are shown as Equations (3) and (4) respectively. The results showed that the Sym6 wavelet under 5th layer decomposition had the best performance in denoising. This paper continues to use this conclusion.

$$RSNR=\frac{1}{N}\left(\sum_{j=1}^{N}10\times\log_{10}\left(\sum_{i=1}^{L}\frac{x_{ij}'^{2}}{(x_{ij}'-x_{ij})^{2}}\right)\right) \quad (3)$$

$$CS=\left\{\sum_{i=1}^{L}(x_{i+1,j}'-x_{i,j}')^{2}\right\}/\left\{\sum_{i=1}^{L}(x_{i+1,j}-x_{i,j})^{2}\right\} \quad (4)$$

where, x_{ij} denotes the original spectral absorbance of sample j at wavelength of i nm, x_{ij}' denotes the corresponding filtered spectral absorbance, N denotes the number of samples, and L denotes the number of wavebands.

To reduce the influence of the canopy structure with leaves, wheat ears and wheat awns at heading stage, the MSC algorithm was used to correct the spectral reflectance of winter wheat canopy. The calculation processes are shown in Equations (5) and (6).

$$X_i=a_i+b_i\overline{X}_i \quad (5)$$

$$X_{i\_msc}=(X_i-a_i)/b_i \quad (6)$$

where, a_i denotes linear translation coefficient; b_i denotes tilt migration coefficient; X_i denotes the reflectance spectrum of the i-th sample; $\overline{X}_i$ denotes average reflection spectrum of all samples; X_{i_msc} denotes the corrected reflectance spectrum.

2.3.2 Abnormal sample removal by Pauta Criterion

According to the Pauta Criterion[15], the probability for one datum being greater than $\mu+3\delta$ or smaller than μ-3δ is smaller than 0.003, which was shown as Equation (7).

$$P(|x-u|>3\delta)\leq 0.003 \quad (7)$$

where, μ is mathematical expectation, δ is standard deviation. At this point, the experimental data which are not in the range of ($\mu-3\delta$, μ+3δ) could be regarded as abnormal data and should be

removed. Standardized values (*z*-score) can be used to help identify abnormal data. Thus, the data whose *z*-score lower than –3 or higher than 3 will be regarded as abnormal data and be removed.

2.3.3 Data set division by SPXY

Kennard-Stone (KS) was the most commonly employed sample set partitioning method. The Euclidean distance was used in the method as the reference condition for selection, whose calculation formula is as Equation (8). As can be seen, KS selects samples only based on spectral features without considering the influence of property variables[16].

Taking both spectral space and property space of samples into account simultaneously, Galvao firstly proposed SPXY sample division method on the basis of KS method and Content Gradient[17]. The distance used in Content Gradient method was calculated by Equation (9). According to the principle of SPXY, in order to ensure that the distance has the same weight in the *X* and *Y* space respectively, $d_x(p,q)$ and $d_y(p,q)$ were divided by the maximum value in respective data space. The SPXY distance between the samples is calculated as Equation (10).

$$d_x(p,q)=\sqrt{\sum_{j=1}^{J}[x_p(j)-x_q(j)]2};\ p,q\in[1,N] \tag{8}$$

$$d_y(p,q)=\sqrt{(y_p-y_q)^2};\ p,q\in[1,N] \tag{9}$$

$$d_{xy}(p,q)=\frac{d_x(p,q)}{max_{p,q}\in[1,N]d_x(p,q)}+\frac{d_y(p,q)}{max_{p,q}\in[1,N]d_y(p,q)} \tag{10}$$

2.3.4 MWPLS (Moving window partial least squares)

MWPLS works as a selection method to locate informative regions through the whole spectra. Briefly, in MWPLS, a spectral window that starts at the *i*th spectral waveband and ends at the $(i+w-1)^{th}$ spectral waveband is constructed, of which *w* denoted the window width. For simplicity, the starting position of each window is used for denoting the window position. The window is moved over the whole spectral region. At each position, a PLS model for chlorophyll content prediction. This research employed the Root Mean Squared of Cross Validation (RMSECV) indicator to evaluate the performance of each PLS model, which is calculated according to Equation (11).

$$RMSECV=\sqrt{\sum_{i=1}^{n}(y'-y)^2/n} \tag{11}$$

where, y' is the predictive value and y is the measured value; n is the number of samples in the data set.

2.3.5 GA (Genetic algorithm)

GA was used to further select sensitive individual wavebands from the intervals determined by MWPLS. GA generates high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection. The specific procedures used to optimize the sensitive wavebands for chlorophyll content detection are as follows:

1) Encode the waveband variables

In a genetic algorithm, a population of candidate individuals to an optimization problem is evolved toward better individuals. In the study, an individual was represented in binary as a string of 0 s and 1 s of which the length is *S* (*S* is the number of wavebands in a candidate individual). The character 0 and 1 in the string represent "not selected" and "selected" respectively. For example, the string "01001100" indicates that the second, fifth, sixth wavebands are selected, and the rest are not selected.

2) Fitness function design

In each generation, the fitness of every individual in the population is evaluated; the fitness is usually the value of the fitness function in the optimization problem being solved. The more fit individuals are stochastically selected from the current population. The smaller of the RMSECV of the prediction model is, the better the model is. Therefore, the study chooses the calculation equation of RMSECV as the fitness function.

3) Initializing groups

The initial population of this study consists of 200 individuals randomly generated by a computer, each individual consisting of *S* 0/1 characters.

4) Genetic operation design

① Selection operator

During each successive generation, a portion of the existing population is selected to breed a new generation. In the present study, the selection operator adopts the most commonly used method, called Rotary Method. Individuals are selected through a fitness-based process, where fitter individuals (as measured by a fitness function) are typically more likely to be selected.

② Crossover operator

The key process of biological evolution is the crossover of genetic genes, which refers to the replacement and recombination of genetic genes from the parents to produce new individuals. Through crossover, the search ability of GA is improved. As shown in Figure 3, the crossover operator used in the study is a single point crossover, that is, randomly select the point to do the crossover in two individuals (strings).

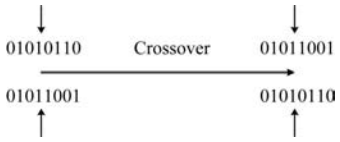

Figure 3 Sketch map of single point cross

③ Mutation operator

The mutation operator is the process of introducing new individuals into the population. The purpose of mutation is to improve the local search ability, maintain the diversity of the population, and avoid the phenomenon of premature convergence. The study uses the basic mutation operator, that is, randomly select the start position in a certain individual (string), to do the inverse operation (1 becomes 0, or 0 becomes 1), as shown in Figure 4. The mutation probability is usually small (0.01-0.2).

Figure 4 Sketch map of mutation

5) Termination condition of iteration

This generational process is repeated until a termination condition has been reached. Setting 100 as the maximum generations in this study, that is, when the numbers of generation reach 100, the process terminates.

The operation processes of the methods mentioned above were carried out using Matlab R2014b.

3 Results and discussion

3.1 Wavelet denoising and MSC

All spectral curves of the wheat canopy samples were obtained by using the ASD spectrometer. The original spectral characteristic in the range of 325-1075 nm were shown in Figure 5. It is revealed that the variation tendency of the wheat canopy spectral reflectance was similar with each other with the change of chlorophyll content. However, there were significant fluctuation

noises located in the range of 325-400 nm and 900-1075 nm and slight fluctuation noises in the range of 400-900 nm. Thus it is necessary to eliminate noise influence before further analysis.

Wavelet denoising was employed and implemented in matlab software. The sym6 wavelet basis function and 5 decomposition level were selected by the previous research[14]. The denoised results are shown in Figure 6. Compared to the original signal, it was obvious that the denoised spectral curves were much smoother than the original spectral curve, especially in the range of 325-400 nm and 900-1075 nm. In addition, the slight fluctuations were also filtered within 400-900 nm. The quality of field spectral reflectance data was improved.

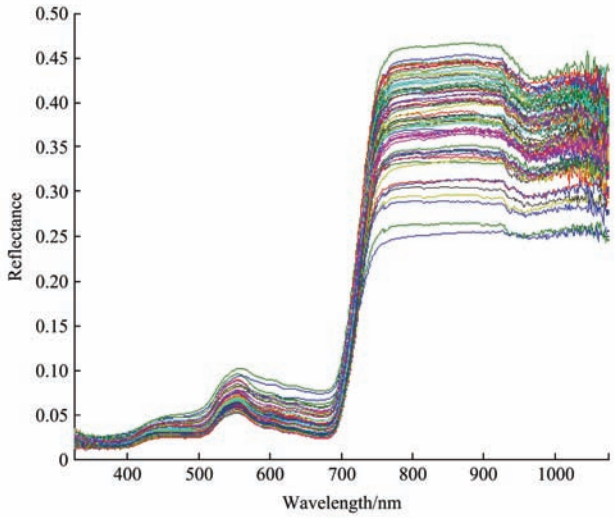

Figure 5 Original spectral curves

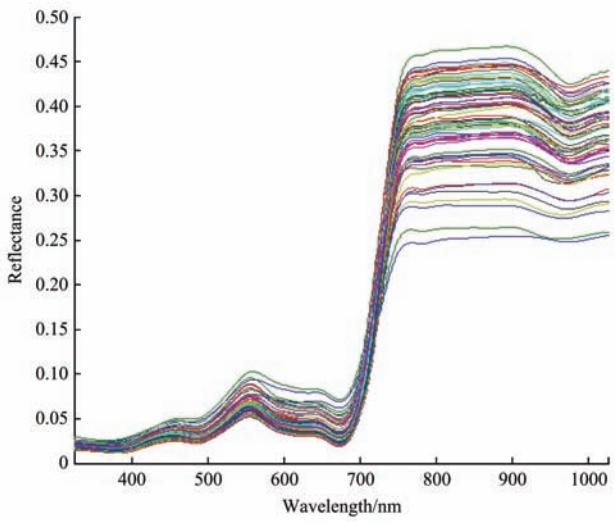

Figure 6 Denoised spectral curves

Scattering problems always existed in the canopy diffuse emission because of the influence from the complex wheat canopy structure[18,19]. Furthermore, there may cause baseline drift due to the long-term use of the instrument. In order to reduce the influence from the baseline shift and scattering, the denoised spectral reflectance was processed by MSC. And the corrected spectral curves are shown in Figure 7.

There are some characteristic regions on the spectral response curve of green plants[20], including a absorption region around 480 nm, a strong reflectance region around the 550 nm, a strong absorption region around the 670 nm, red-edge region around 750 nm and high reflectance platform around 850 nm. And the sensitive wavebands are mainly concentrated in these characteristic regions. Therefore, the sensitive wavebands selection will be conducted only on the spectral region of 400-900 nm, which are helpful in reducing interference from redundant information.

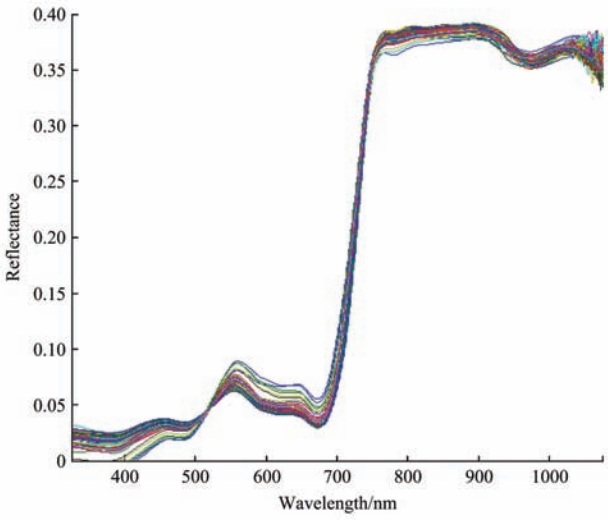

Figure 7 The spectral curves after corrected

3.2 Abnormal data removal

The both datasets including crop canopy spectral reflectance data and chlorophyll content value data were standardized. According to the Pauta criterion, the data whose z-score are higher than 3 or lower than –3 would be regarded as abnormal data. After applying the criterion, abnormal samples had been removed and 68 samples were remained for the following analysis.

The chlorophyll content of the remaining samples is within the limits 41.10-59.60 mg/L, the average is 52.14 mg/L, and the standard deviation is 3.59. Chlorophyll content mainly concentrated in the range of 46-52 mg/L. Among the remaining samples, three pairs have had the same chlorophyll content and five pairs had the very close chlorophyll content, only a difference of 0.1 mg/L. In order to avoid the interference of similar chlorophyll content on the accuracy of prediction model, samples with similar or same values on chlorophyll contents were averaged. After that operation, there were 60 samples remaining.

3.3 Partition of sample set

Then the remaining samples were divided into modeling dataset and verification dataset by SPXY. The modeling dataset contains 45 samples while the verification dataset contains 15 samples. The statistics of chlorophyll content are shown in Table 1. The average chlorophyll content values of model dataset and validation dataset were 51.91 mg/L and 52.55 mg/L, respectively. And both the standard deviations were 3.94 and 3.33, respectively. The result that the two dataset had the similar distribution proves the SPXY method is valid.

Table 1 Chlorophyll content of each data set (mg/L)

	Num	Max	Min	Mean	Std
All data	60	59.60	41.10	52.15	3.60
Modeling	45	59.60	41.10	51.91	3.94
Verification	15	56.10	48.40	52.55	3.33

3.4 Spectral regions selection by MWPLS

Applying MWPLS to the 400-900 nm spectral region with the window width of 25, a total of 476 windows were obtained. The PLS models for predicting chlorophyll content were established for all the spectral windows. The RMSECV for each model was evaluated and plotted in Figure 8. The spectral regions whose RMSECV was less than 2.5 (under the dotted line in Figure 8) were

selected and consecutive regions were merged. The finally obtained spectral regions were: 473-506 nm, 550-575 nm, 655-700 nm and 825-870 nm, totally containing 151 wavebands.

The corresponding positions of the selected spectral regions on the typical green plant characteristic spectral curve are shown in Figure 9. As can be seen, all the selected spectral regions located at the characteristic regions for chlorophyll content detecting. The 473-506 nm region corresponds to the absorption valley of chlorophyll and carotenoid. The 550-575 nm region corresponds to the strong reflectance peak of chlorophyll. The 655-700 nm region corresponds to the strong absorption valley of chlorophyll. The 825-870 nm region corresponds to high reflectance platform mainly because of leaf tissue structure.

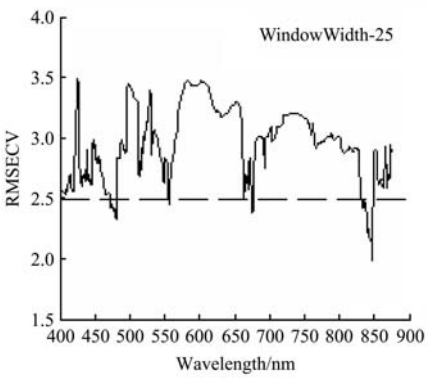

Figure 8 RMSECV for each model

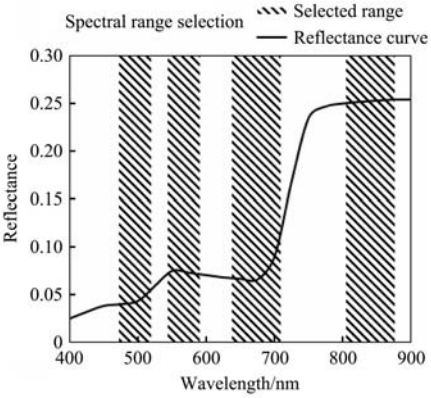

Figure 9 Position of selected spectral regions on typical green plant characteristic spectral curve

3.5 Wavebands selection by GA

In order to further remove the invalid wavebands mixed in the continuous waveband regions, the GA optimization algorithm was applied to the waveband regions selected by MWPLS. And the parameters settings were as the followings: initial population 200, maximum selection variable number 151, cross probability 0.8, variation probability 0.1, genetic generations 100. In order to overcome the random errors caused by group initialization, the genetic iterative operation was carried out for 200 rounds.

One of the useful outputs of the GA procedure is called the frequency of selection plot, which gives the compiled results for the frequency of a certain variable being involved in the models through all replicate runs. The higher selection frequency of one waveband variable, the more it should be retained for model establishment. The selected frequencies of all candidate wavebands are shown in Figure 10. There are 36 wavebands whose selected frequency reaches 0.4. There are 15 wavebands whose selected frequency reaches 0.45. The performance of these wavebands in chlorophyll content prediction will be tested by establishing PLS models.

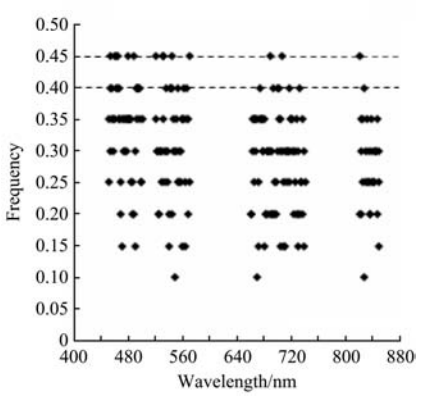

Figure 10 Selected frequency of candidate wavebands

3.6 Comparison between PLS models

Using all the wavebands in the range of 400-900 nm as input variables, a PLS model (Model 1) for chlorophyll content prediction was established, as shown in Figure 11. Using all the 151 wavebands in the MWPLS selected regions as input variables, a PLS model (Model 2) for was established, as shown in Figure 12. Using the 36 wavebands selected by GA as input variables, a PLS model (Model 3) was established, as shown in Figure 13. Using the 15 wavebands selected by GA as input variables, a PLS model (Model 4) was established, as shown in Figure 14. And the comparison among these models are shown in Table 2.

In terms of the R_c^2 and R_v^2, the descending orders are model 3, model 4, model 2 and model 1, while in terms of the *RMSEM* and *RMSEP*, the ascending orders are model 3, model 4, model 2, and model 1. The model 1 has worst performance than other models, with R_c^2 and R_v^2 being 0.4468 and 0.3821, respectively, *RMSEM* and *RMSEP* being 2.9057 and 1.7589, respectively. The results showed that selecting sensitive wavebands and reducing irrelevant or redundant wavebands can improve model predictive power.

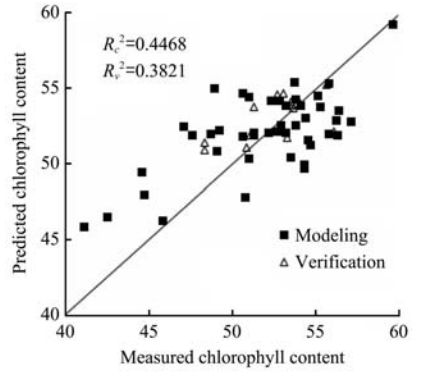

Figure 11 Model 1 by all the wavebands in range of 400-900 nm

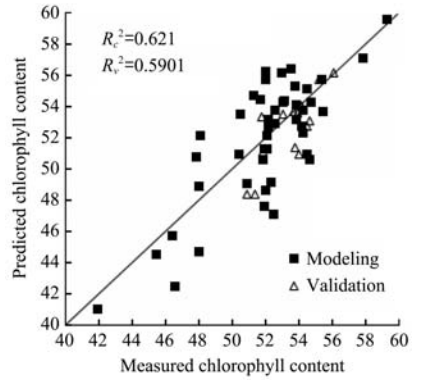

Figure 12 Model 2 by all the wavebands in the selected regions

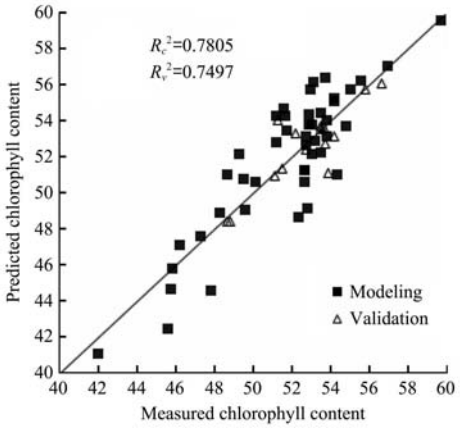

Figure 13 Model 3 with the 36 wavebands selected by GA

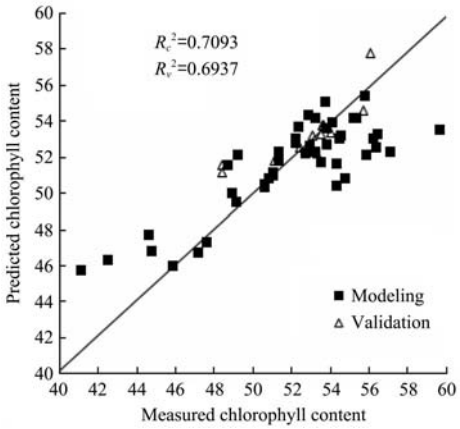

Figure 14 Model 4 with the 15 wavebands selected by GA

Table 2 Comparison between the above models

	Model 1	Model 2	Model 3	Model 4
R_c^2	0.4468	0.6210	0.7805	0.7093
R_v^2	0.3821	0.5901	0.7497	0.6937
RMSEM	2.9057	2.4007	1.8504	2.2679
RMSEP	1.7589	1.6408	1.1315	1.2655
Num	501	151	36	15

The model 3 has the best performance, with R_c^2 and R_v^2 being 0.7805 and 0.7497, respectively, *RMSEM* and *RMSEP* being 1.8504 and 1.1315, respectively. And the results that both model 3 and model 4 performance better than model 2 illustrate that GA can further reduce irrelevant variables based on the regions selected by MWPLS and help improve optimize predictive performance of the models. It also proves that the MWPLS+GA strategy in the paper is effective in waveband selection for chlorophyll content prediction.

However, the prediction performance declines from model 3 to model 4. This reason maybe that some effective spectral wavebands, such as 560 nm and 680 nm, were also be filtered simultaneously. So Model 3 based on the 36 wavebands is recommended. And how to retain necessary valid variables while removing irrelevant variables is a problem worthy of further exploration.

4 Conclusions

The research proposed a MWPLS+GA strategy to select sensitive wavebands from wheat canopy spectrum for chlorophyll content prediction. The spectral data were firstly preprocessed by wavelet denoising, multiple scatter correction and abnormal data removal. Then, the dataset was divided into modeling dataset and validation dataset by SPXY. Finally, the new wavebands selection strategy, MWPLS+GA, was applied and 36 wavebands were selected for establishing the PLS chlorophyll content prediction model, of which, R_c^2=0.7805, *RMSEM*=1.8504; R_v^2= 0.7497, *RMSEV*=1.1315.

Acknowledgements

The project was supported by the National Key Research and Development Program (2016YFD0200600-2016YFD0200602), National Natural Science Fund (Grant No. 31501219), and the graduate training project of China agricultural university (ZYXW037, HJ2019029, YW2019018).

[References]

[1] Xie C Q, Yang C, Hummel Jr A, Johnson G A, Izuno F T. Spectral reflectance response to nitrogen fertilization in field grown corn. Int J Agric & Biol Eng, 2018; 11(4): 118–126.

[2] Huang S Y, Miao Y X, Yuan F, Cao Q, Ye H C, Victoria I S, et al. In-season diagnosis of rice nitrogen status using proximal fluorescence canopy sensor at different growth stages. Remote Sens. 2019, 11(16), 1847; https://doi.org/10.3390/rs11161847

[3] Ye X J. Abe S, Zhang S H. Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging. Precision Agric., Sringer, 2019; https://doi.org/10.1007/s11119-019-09661-x.

[4] Padilla F M, Souza R D, Peña-FleitasT M, Grasso R, Gallardo M, Thompson R B. Influence of time of day on measurement with chlorophyll meters and canopy reflectance sensors of different crop N status. Precision Agric., 2019; https://doi.org/10.1007/s11119-019-09641-1.

[5] Yang J, Cheng Y J, Du L, Gong W, Shi S, Sun J, Chen B W. Selection of the optimal bands of first-derivative fluorescence characteristics for leaf nitrogen concentration estimation. Applied Optics, 2019; 58(21): 5720–5727.

[6] ZhangY, Li M Z, Zheng L H, Qin Q M, Lee W S. Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm. Geoderma, 2019; 333(1): 23-34

[7] Wang L, Lin Y, Wang X, Xiao N, Xu Y, Li H, Xu Q. A selective review and comparison for interval variable selection in spectroscopic modeling. Chemometrics and Intelligent Laboratory Systems, 2018; 172(15): 229–240.

[8] Khorshidi N, Niazi A. Moving window partial least squares after orthogonal signal correction as a coupling method for determination ofuranium and thorium by ultrasound assisted emulsification microextraction. Journal of Chemometrics, 2019; 33: e3083.

[9] Li Z, Guan A H, Ge H Y, Lian F Y. Wavelength selection of amino acid THz absorption spectra for quantitative analysis by a self-adaptive genetic algorithm and comparison with MWPLS. Microchemical Journal, 2017; 132: 185–189

[10] Jiang H, Xu W D, Chen Q S. Comparison of algorithms for wavelength variables selection from near-infrared (NIR) spectra for quantitative monitoring of yeast (Saccharomyces cerevisiae) cultivations. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019; 214(5): 366–371

[11] Du X L, Li X Y, Liu Y, Zhou W H, Li J L. Genetic algorithm optimized non-destructive prediction on property of mechanically injured peaches during postharvest storage by portable visible/shortwave near-infrared spectroscopy. Scientia Horticulturae, 2019; 249(30): 240–249.

[12] Li Z. Wavelength selection for quantitative analysis in terahertz spectroscopy using a genetic algorithm. IEEE Transactions on Terahertz Science and Technology, 2016; 6(5): 658–663

[13] Li H, Kou J S, Li M Q. Influence of solution space scale on stability of GA. Journal of Systems Engineering, 2007; 22(2): 162–169.

[14] Liu H J, Li M Z, Zhang J Y, Gao D H, Sun H, Yang L W. Estimation of chlorophyll content in maize canopy using wavelet denoising and SVR method. Int J Agric & Biol Eng, 2018; 11(6): 132–137.

[15] Xie H, Zhao A, Huang S Y, Han J, Liu S C, Xu X, et al. Unsupervised hyperspectral remote sensing image clustering based on adaptive density. IEEE Geoscience and Remote Sensing Letters, 2018; 15(4): 632–636.

[16] Yang Z F, Xiao H, Zhang L, Feng D J, Zhang F Y, Jiang M S, et al. Fast

determination of oxide content in cement raw meal using NIR spectroscopy with the SPXY algorithm. Analytical Methods, 2019; 11: 3936–3942.

[17] Galvão R K H, Araujo M C U, José G E, Pontes M J C, Silva E C, Saldanha T C B. A method for calibration and validation subset partitioning. Talanta, 2005; 67(4): 736–740.

[18] Rabatel G, MakdessiN A, Ecarnot M, Roumet P. A spectral correction method for multi-scattering effects in close range hyperspectral imagery of vegetation scenes: application to nitrogen content assessment in wheat. Advances in Animal Biosciences: Precision Agriculture (ECPA), 2017: 8(2): 353–358

[19] Kamlesh G, Balasundram S K, Ganesan V, Biswajeet P. Estimating chlorophyll content at leaf scale in viroid-inoculated oil palm seedlings (Elaeisguineensis Jacq.) using reflectance spectra (400 nm–1050 nm). International Journal of Remote Sensing, 2019; 40(19): 7647-7662.

[20] Li M Z. Spectral analysis technique and its application. Beijing: Science Press, 2006; pp.176–180. (in Chinese)