ESTIMATION OF CHLOROPHYLL CONTENT IN APPLE LEAVES BASED ON IMAGING SPECTROSCOPY

Ruiyang Yu [1], Xicun Zhu [1,2*], Shujing Cao [1], Jingling Xiong [1], Xin Wen [1], Yuanmao Jiang [3], Gengxing Zhao [1]

[1] College of Resources and Environment, Shandong Agricultural University, Tai'an, 271018, China; e-mail: sdauhsrs@163.com
[2] Key Laboratory of Agricultural Ecology and Environment, Shandong Agricultural University, Tai'an, 271018, China
[3] College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an 271018, China

To promote the use of imaging spectroscopy to assess the nutritional status of apple trees, the models to estimate the chlorophyll content of apple leaves were explored. Spectral data for apple leaves were collected with an imaging spectrometer and then preprocessed with the nine-point moving weighted average method. Correlation analyses were conducted between chlorophyll content and mathematically transformed spectral data. Wavelengths sensitive to chlorophyll content were selected on the basis of the highest correlation coefficients, and partial least squares (PLS), support vector machine (SVM), and random forest (RF) models to estimate chlorophyll content were established and tested. The wavelengths sensitive to chlorophyll content were 414, 424, 429, 439, and 577 nm. The best model was the SVM model with wavelength data subjected to a second order differential of the logarithm transformation (lgR_{414})″, (lgR_{424})″, (lgR_{429})″, (lgR_{439})″, (lgR_{577})″ as the independent variables. For this model, the coefficient of determination V-R^2 was 0.7372, the root mean square error V-RMSE was 0.4477, and the residual predictive deviation V-RPD was 1.8810. Among all the models, this SVM model had the highest V-R^2 and V-RPD values and the lowest V-RMSE value.

Keywords: *imaging spectroscopy, apple leaves, chlorophyll content, estimating models.*

ОЦЕНКА СОДЕРЖАНИЯ ХЛОРОФИЛЛА В ЛИСТЬЯХ ЯБЛОНЬ НА ОСНОВЕ ГИПЕРСПЕКТРАЛЬНЫХ ИЗОБРАЖЕНИЙ

R. Yu [1], X. Zhu [1,2*], S. Cao [1], J. Xiong [1], X. Wen [1], Y. Jiang [3], G. Zhao [1]

УДК 535.317.1;547.979.7

[1] Колледж ресурсов и окружающей среды, Шаньдунский сельскохозяйственный университет, Тайань, 271018, Китай; e-mail: sdauhsrs@163.com
[2] Шаньдунский сельскохозяйственный университет, Тайань, 271018, Китай
[3] Научно-технический колледж садоводства, Шаньдунский сельскохозяйственный университет, Тайань 271018, Китай

(Поступила 6 марта 2018)

Для подтверждения возможности использования гиперспектральных изображений для оценки состояния питания яблонь исследовано содержание хлорофилла в их листьях. Спектральные данные получены с помощью спектрометра и предварительно обрабатывались методом взвешенного скользящего среднего по девяти точкам. С помощью корреляционного анализа установлена связь между содержанием хлорофилла и математически преобразованными спектральными данными. Длины волн, чувствительные к содержанию хлорофилла, выбраны на основе самых высоких коэффициентов корреляции; для оценки содержания хлорофилла использованы методы частных наименьших квадратов (PLS), опорных векторов (SVM) и случайного леса (RF). Длины волн, чувствительные к содержанию хлорофилла: 414, 424, 429, 439 и 577 нм. Лучшей моделью оказалась SVM с независимыми переменными, представляющими собой вторые производные логарифма (lgR_{414})″, (lgR_{424})″,

$(lgR_{429})''$, $(lgR_{439})''$, $(lgR_{577})''$. Для этой модели коэффициент детерминации R^2 = 0.7372, среднеквадратическая ошибка RMSE = 0.4477, остаточное прогнозное отклонение RPD = 1.8810. Модель SVM имела самые высокие R^2 и RPD и самое низкое RMSE по сравнению с другими моделями.

Ключевые слова: *визуализация, листья яблони, содержание хлорофилла, оценочные модели.*

Introduction. Chlorophyll content is an indicator of crop photosynthesis and crop growth [1, 2]. The traditional method to determine chlorophyll content is spectrophotometry. Although this method is very accurate, the experimental procedure is long and complicated, and so it is not wellsuited to meet the needs of rapid, real-time, nondestructive monitoring.

In recent years, imaging spectroscopy has been developed as a technique to estimate plant growth status rapidly and nondestructively. It has wide applications in fast, real-time, and nondestructive testing of crop growth and nutrition. For example, Maccioni et al.[3] used single-leaf directional reflectance normalized by the green band and red edge to create a new vegetation index to predict chlorophyll content. Ciganda et al. [4] constructed a red-edge chlorophyll index based on red edge (720–730 nm) and near-infrared (770–800 nm) reflectance to predict the chlorophyll content of the corn canopy. Schlemmer et al. [5, 6] constructed leaf- and canopy-scale vegetation indexes such as EVI2, REP, MTCI, CI green, and CI red edge, which were used to create models for estimating chlorophyll content. Gong et al. [7] constructed models to estimate chlorophyll content by creating "three sides" parameters, a ratio spectral index, and a normalized difference spectral index. Zhu et al. [8] combined spectral parameters, the red edge area, and certain wavelength bands to monitor the chlorophyll content of apple leaves and obtained the best predictive effect with a model incorporating R_{800}/R_{550} and the red edge area. Yue et al. [9] used the original spectrum subjected to various mathematical transformations as independent variables in models for predicting the chlorophyll content of citrus leaves. The models were based on principal component analysis for dimensionality reduction by an support vector machine (SVM) regression algorithm, wavelet de-noising, and partial least squares (PLS) regression. Wang et al. [10] established univariate and multivariable regression models to predict the canopy nitrogen content in maize at the seedling stage with a normalized vegetation index, a normalized spectral vegetation index, a ratio-based spectral index, and a difference vegetation index based on imaging spectroscopy data.

Many studies have focused on the inversion of biochemical components of vegetation using non-imaging hyperspectral techniques. However, relatively few studies have established models based on imaging spectroscopy data. In this study, the spectral data of apple leaves were collected using an imaging spectrometer. The average spectral reflectance of the leaves was determined, and the spectral data were preprocessed with the nine-point moving weighted average method. The correlations between chlorophyll content and spectral data subjected to 11 kinds of mathematical transformation were determined. The wavelengths most sensitive to chlorophyll content (those with the highest correlation coefficients) were selected as the independent variables for the models. Then partial least squares (PLS), support vector machine (SVM), and random forest (RF) models to predict chlorophyll content were established and tested. The aims of this study were to explore models for estimating chlorophyll content based on transformed spectral data and to provide technical support for the application of imaging spectroscopy to monitor crop nutrition.

Experimental. *Sample collection.* Sampling was conducted in September 2016 at apple (cultivar, Red Fuji) orchards in the village of Xia Tan Bu Lin, Gao Du Town, Meng Yin County, Shandong Province, China. Based on the distribution map of the orchards, sampling points were selected and 100 healthy apple trees were chosen for sampling. Three healthy mature leaves without pest or disease damage were collected from vegetative shoots in the middle of each apple tree. The leaves were quickly placed in bags, and then the bags were sealed, labelled, placed in a black bag filled with ice packs, and brought back to the laboratory.

Acquisition of spectral data. Imaging spectral data were obtained from apple leaves using an SOC710VP imaging spectrometer. The spectral range was 400–1000 nm; the spectral resolution was 4.68 nm, and there were 128 bands. Data were acquired under fine weather conditions to meet the signal-to-noise ratio accuracy required for imaging spectrometry. The measurement time was 9:00–15:00 and measurements were conducted without moving objects nearby. ENVI 5.1 was applied to view the images of apple leaves. A suitable area of apple leaves was selected as the region of interest. The spectral average reflectance was extracted from each region of interest without leaf veins, petiole segments, or shaded or wrinkled regions. The average spectral reflectance for all regions of interest represented the spectral reflectance of each entire leaf. This operation was repeated to extract the average spectral reflectance of 100

apple leaf samples. Radiometric calibration of the spectrometer was performed using a reference radiation source (i.e., a white board with reflectance = 1). Reflectance has no unit.

Determination of chlorophyll content. The chlorophyll content of leaves was determined by spectrophotometry [11]. The main source of error in leaf chlorophyll content was losses caused by leaf grinding. The 100 apple leaf samples were ranked in order based on their chlorophyll content. The sample calibration set (75 samples) and verification set (25 samples) were selected by an isometric sampling method (Table 1).

TABLE 1. Statistical Information for Apple Leaf Chlorophyll Content (mg/g)

Sample	Sample number	Maximum	Minimum	Mean	Standard deviation
Total sample	100	4.83	0.87	3.53	0.81
Calibration set	75	4.81	0.88	3.51	0.83
Verification set	25	4.83	0.87	3.56	0.84

Data preprocessing. The spectral curves required smoothing to reduce noise. A previous comparative study found that the nine-point weighted moving average method [12] was effective to smooth spectral curves for noise reduction. This method is summarized by formula:

$$R_i' = 0.04R_{i-4} + 0.08R_{i-3} + 0.12R_{i-2} + 0.16R_{i-1} + 0.20R_i + 0.16R_{i+1} + 0.12R_{i+2} + 0.08R_{i+3} + 0.04R_{i+4}, \quad (1)$$

where R_i is the spectral reflectance of each wavelength from the 5th wavelength to the 124th wavelength before smoothing, i is the number of the spectral channel, and R_i' is the spectral reflectance after smoothing.

Spectral data transformations. The original spectral reflectance data of apple leaves were subjected to smoothing and then 11 different mathematical transformations, as follows: first-order differential R', second-order differential R'', reciprocal $1/R$, logarithm lg$R$, square root $R^{1/2}$, first-order differential of the reciprocal $(1/R)'$, first-order differential of the logarithm $(\lg R)'$, first-order differential of the square root $(R^{1/2})'$, second-order differential $(1/R)''$, second-order differential of the logarithm $(\lg R)''$, and second-order differential of the square root $(R^{1/2})''$. The R' and R'' transformations can eliminate background interference, while $1/R$, lg$R$, and $R^{1/2}$ can highlight spectral features [13].

Models for estimating chlorophyll content. We established and tested PLS, SVM, and RF models for estimating the chlorophyll content of apple leaves. The first method, PLS, is an important method for multivariate statistical analysis [13]. It combines basic functions of multiple linear regression analysis, canonical correlation analysis, and principal component analysis. It can effectively overcome the one-line problem of a general multicomponent linear regression.

The SVM model is currently the fastest-growing method of machine learning. Based on the least structural risk, it solves the practical problems of nonlinearity, over-learning, dimensional disaster, local minimum, and small sample size. It has excellent generalization and promotion abilities [14, 15].

The RF model is a method of machine learning that consists of multiple decision trees. The decision maker combination consists of a bagging (boot-strap aggregating) algorithm and a randomization algorithm. This model is widely used in classification and regression analysis, and it produces good results for both linear and nonlinear data. Only two parameters are set in the RF model: the number of decision trees, and the number of input features when each node of the decision tree splits [16, 17].

Tests of accuracy of models. The coefficient of determination (C-R^2), root mean square error (C-RMSE), slope (C-S), and residual predictive deviation (C-RPD) were calculated and used to evaluate the accuracy of each model in predicting the chlorophyll content. The coefficient of determination (V-R^2), root mean square error (V-RMSE), slope (V-S), and residual predictive deviation (V-RPD) were calculated and used to evaluate the results of verification. The larger the R^2, the smaller the RMSE, and the closer the S to 1, the greater the accuracy of the estimation. The residual predictive deviation (RPD) is the ratio of sample standard deviation (SD) to root mean square error (RMSE). When RPD > 2, the model works very well and can be used for quantitative analysis; when 1.8 < RPD < 2.0, the model is good and can be used for quantitative estimation; when 1.4 < RPD < 1.8, the model can be used to make rough estimates; when RPD <1.4, the model is less effective [7]. The formulas below show the calculations for these indicators:

$$R^2 = \sum_{i=1}^{N}(y_i - x_i)^2 \Big/ \sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\,, \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2}, \quad (3)$$

$$S = (y_{i+1} - y_i)/(x_{i+1} - x_i), \quad (4)$$

$$\text{RPD} = \text{SD}/\text{RMSE}, \quad (5)$$

among which

$$\text{SD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}, \quad (6)$$

where x_i is the measured value of chlorophyll content, y_i is the predictive value of chlorophyll content, $\bar{x}$ is the average measured value, and N is the sample quantity, and I is the number of samples.

Results and discussion. *Apple leaves with different chlorophyll contents have different spectral characteristics.* The apple leaves were divided into three groups on the basis of their chlorophyll content. The average spectral reflectance curves of each group were obtained (Fig. 1). In all groups, the spectral reflectance first increased with increasing wavelength, then decreased as the wavelength increased in the green part of the spectrum (500–600 nm). Because of the strong absorption of blue-violet light and red light by chlorophyll, a reflection peak formed by reflection of the green light band. The spectral reflectance gradually decreased with increasing chlorophyll content. Because of the strong absorption of red light by chlorophyll and the strong reflection of the near infrared, the spectral reflectance increased sharply at 700–800 nm (from red to near infrared). In the near infrared region of 800–900 nm, the spectral reflectance changed little with increasing wavelength. At the wavelengths of 900–1000 nm, the spectral reflectance decreased as the wavelength increased.

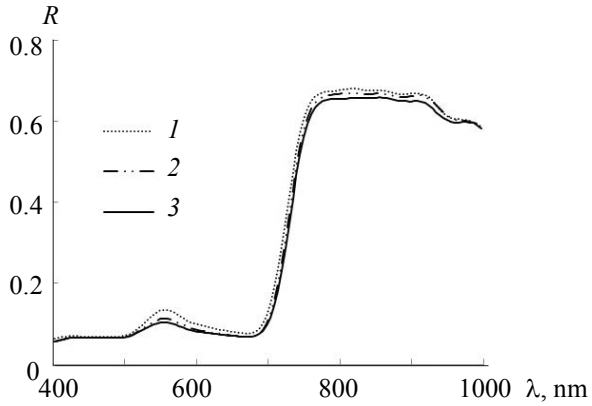

Fig. 1. Hyperspectral characteristics of apple leaves with chlorophyll content 2–3 (1), 3–4 (2), and 4–5 mg/g (3).

Correlation analyses between spectral data and chlorophyll content. The spectral data were subjected to preprocessing and various mathematical transformations, and then correlation analyses between chlorophyll content and the transformed data were conducted. As shown in Table 2, in each transformed dataset, the five wavelengths with the largest absolute correlation coefficient values were selected as the wavelengths most sensitive to chlorophyll content. In the table, these wavelengths are listed in the order from the highest to the lowest correlation coefficient with chlorophyll content.

Establishment of PLS model. We constructed PLS models with the five wavelengths with the largest correlation coefficients as the independent variables and chlorophyll content as the dependent variable. The results are shown in Table 2.

As shown in Table 3, the PLS models based on data transformed with R', R'', $(1/R)'$, $(\lg R)'$, $(R^{1/2})'$, and $(\lg R)''$ as the independent variables had C-R^2 values of 0.6971, 0.6984, 0.6466, 0.6944, 0.7144, and 0.7003, respectively, significantly higher than the C-R^2 values of models based on data transformed with $1/R$, $\lg R$, $R^{1/2}$, $(1/R)''$, and $(R^{1/2})''$ as the independent variables. The C-RMSE values of the PLS models based on data transformed with R', R'', $(1/R)'$, $(\lg R)'$, $(R^{1/2})'$, and $(\lg R)''$ were 0.4529, 0.7043, 0.5200, 0.4594, 0.4397, and 0.4564, respectively, significantly lower than those of the models with data transformed with $1/R$, $\lg R$, $R^{1/2}$, $(1/R)''$, and $(R^{1/2})''$ as the independent variables. The C-S values of the PLS models based on data

transformed with R', R'', $(1/R)'$, $(\lg R)'$, $(R^{1/2})'$, and $(\lg R)''$ were 0.6947, 0.4518, 0.4758, 0.6263, 0.7012, and 0.6240, respectively, significantly higher than those of PLS models based on data transformed with $1/R$, $\lg R$, $R^{1/2}$, $(1/R)''$, and $(R^{1/2})''$. All the PLS models based on data transformed with R', R'', $(R^{1/2})'$, and $(\lg R)''$ as independent variables had C-RPD values >1.80.

In summary, among all the PLS models, the one based on data transformed with $(R^{1/2})'$ as the independent variables had the highest C-R^2 and C-RPD values of 0.7144 and 1.8808, respectively, and the lowest C-RMSE value of 0.4397.

TABLE 2. Wavelengths Sensitive to Chlorophyll Content in Transformed Spectral Datasets

Mathematical transformation	λ, nm	Mathematical Transformation	λ, nm
R'	541, 536, 530, 525, 587	$(\lg R)'$	525, 561, 551, 520, 459
R''	454, 424, 429, 434, 546	$(R^{1/2})'$	444, 525, 618, 566, 530
$1/R$	713, 707, 718, 702, 723	$(1/R)''$	419, 697, 639, 424, 629
$\lg R$	713, 707, 718, 556, 702	$(\lg R)''$	577, 424, 414, 429, 439
$R^{1/2}$	707, 713, 556, 718, 551	$(R^{1/2})''$	449, 439, 490, 454, 434
$(1/R)'$	744, 676, 505, 541, 572		

TABLE 3. PLS Model and its Accuracy Parameters (n = 75)

Independent variables	Estimating model	C-R^2	C-RMSE	C-S	C-RPD
R_{541}', R_{536}', R_{530}', R_{525}', R_{587}'	$y = 5.006 - 368.263R_{541} - 280.892R_{536}' - 234.810R_{530}' - 213.441R_{525}' + 468.075R_{587}'$	0.6971	0.4529	0.6947	1.8258
R_{454}'', R_{424}'', R_{429}'', R_{434}'', R_{546}''	$y = 4.809 + 2952.426R_{454}'' + 1628.961R_{424}'' + 1454.955R_{429}'' - 1525.094R_{434}'' + 543.849R_{546}''$	0.6984	0.4518	0.7043	1.8300
$1/R_{713}$, $1/R_{707}$, $1/R_{718}$, $1/R_{702}$, $1/R_{723}$	$y = 1.679 + 0.096/R_{713} + 0.069/R_{707} + 0.125/R_{718} + 0.046/R_{702} + 0.151/R_{733}$	0.2401	0.7190	0.2053	1.1501
$\lg R_{713}$, $\lg R_{707}$, $\lg R_{718}$, $\lg R_{702}$, $\lg R_{723}$	$y = 0.326 - 1.039\lg R_{713} - 0.954\lg R_{707} - 1.091\lg R_{718} - 0.737\lg R_{702} - 0.840\lg R_{723}$	0.2946	0.6917	0.2705	1.1955
$\sqrt{R_{707}}$, $\sqrt{R_{713}}$, $\sqrt{R_{556}}$, $\sqrt{R_{718}}$, $\sqrt{R_{551}}$	$y = 7.790 - 1.968\sqrt{R_{707}} - 1.908\sqrt{R_{713}} - 2.056\sqrt{R_{556}} - 1.808\sqrt{R_{718}} - 2.047\sqrt{R_{551}}$	0.3223	0.6775	0.3062	1.2205
$(1/R_{744})'$, $(1/R_{676})'$, $(1/R_{505})'$, $(1/R_{541})'$, $(1/R_{572})'$	$y = 4.090 - 1.240(1/R_{744})' - 16.767(1/R_{676})' - 0.366(1/R_{505})' + 0.356(1/R_{541})' + 0.338(1/R_{572})'$	0.6466	0.5200	0.4758	1.5902
$(\lg R_{525})'$, $(\lg R_{561})'$, $(\lg R_{551})'$, $(\lg R_{520})'$, $(\lg R_{459})'$	$y = 4.907 + 16.892(\lg R_{525})' + 18.232(\lg R_{561})' - 17.724(\lg R_{551})' + 16.607(\lg R_{520})' + 229.849(\lg R_{459})'$	0.6944	0.4594	0.6263	1.7997
$(\sqrt{R_{444}})'$, $(\sqrt{R_{525}})'$, $(\sqrt{R_{618}})'$, $(\sqrt{R_{566}})'$, $(\sqrt{R_{530}})'$	$y = 4.872 + 758.450(\sqrt{R_{444}})' + 56.904(\sqrt{R_{525}})' + 79.663(\sqrt{R_{618}})' + 65.006(\sqrt{R_{566}})' + 59.492(\sqrt{R_{530}})'$	0.7144	0.4397	0.7012	1.8808
$(1/R_{419})'$, $(1/R_{697})'$, $(1/R_{639})'$, $(1/R_{424})'$, $(1/R_{629})'$	$y = 4.076 + 2.666(1/R_{419})' + 6.863(1/R_{697}) - 5.204(1/R_{639})' + 1.323(1/R_{424})' + 4.652(1/R_{629})'$	0.1150	0.7842	0.1598	1.0544
$(\lg R_{577})''$, $(\lg R_{424})''$, $(\lg R_{414})''$, $(\lg R_{429})''$, $(\lg R_{439})''$	$y = 4.683 - 257.491(\lg R_{577})'' + 168.002(\lg R_{424})'' - 86.439(\lg R_{414})'' - 86.420(\lg R_{429})'' + 163.494(\lg R_{439})''$	0.7003	0.4564	0.6240	1.8117
$(\sqrt{R_{449}})''$, $(\sqrt{R_{439}})''$, $(\sqrt{R_{490}})''$, $(\sqrt{R_{454}})''$, $(\sqrt{R_{434}})''$	$y = 3.792 + 254.015(\sqrt{R_{449}})'' - 251.768(\sqrt{R_{439}})'' - 113.554(\sqrt{R_{490}})'' + 195.224(\sqrt{R_{454}})'' - 116.968(\sqrt{R_{434}})''$	0.1239	0.7613	0.1085	1.0035

Establishment of SVM model. To construct the SVM model for leaf chlorophyll content, the transformed data for the five wavelengths most sensitive to chlorophyll content were the independent variables, and chlorophyll content was the dependent variable. The parameters were optimized and regression models were made. After multiple verification tests and comparisons, v-SVR was determined as the SVM type and radial basis function (RBF) was selected as the kernel function type. The parameters of the model are follows: 3°, $\gamma = 0.5$, Coef0 = 0.001, $\nu = 0.5$, $\varepsilon = 0.001$, cache size 100, cost 1, shrinking 1, prob 1, $P = 0.01$.

Table 4 shows the verification results for SVM models based on data transformed with $(\lg R)''$ and $(R^{1/2})''$ as the independent variables: the C-R^2 values were 0.7411 and 0.7308, respectively, significantly higher than those of models based on data transformed with R', R'', $(1/R)'$, $(\lg R)'$, $(R^{1/2})'$, $1/R$, $\lg R$, $R^{1/2}$, and $(1/R)''$ as the independent variables. The C-RMSE values for the SVM models based on data transformed with $(\lg R)''$ and $(R^{1/2})''$ as the independent variables were 0.4196 and 0.4286, respectively, significantly lower than those of the models based on data transformed with R', R'', $(1/R)'$, $(\lg R)'$, $(R^{1/2})'$, $1/R$, $\lg R$, $R^{1/2}$, and $(1/R)''$ as independent variables. The C-*S* values of the SVM models based on data transformed with $(\lg R)''$ and $(R^{1/2})''$ as the independent variables were 0.7577 and 0.6975, respectively, significantly higher than those of the models based on data transformed with R', R'', $(1/R)'$, $(\lg R)'$, $(R^{1/2})'$, $1/R$, $\lg R$, $R^{1/2}$, and $(1/R)''$ as the independent variables. The SVM models based on data transformed with $(\lg R)''$ and $(R^{1/2})''$ as the independent variables had C-RPD values >1.90.

In summary, among all the SVM models, the one based on data transformed with $(\lg R)''$ as the independent variables had the highest C-R^2 and C-RPD values of 0.7411 and 1.9705 and the lowest C-RMSE value of 0.4196.

TABLE 4. SVM Model and its Accuracy Parameters (n = 75)

Independent variable	C-R^2	C-RMSE	C-S	C-RPD
R_{541}', R_{536}', R_{530}', R_{525}', R_{587}'	0.2365	0.7325	0.1578	1.1288
R_{454}'', R_{424}'', R_{429}'', R_{434}'', R_{546}''	0.1512	0.7761	0.1152	1.0654
$1/R_{713}$, $1/R_{707}$, $1/R_{718}$, $1/R_{702}$, $1/R_{723}$	0.07914	0.7955	0.0589	1.0394
$\lg R_{713}$, $\lg R_{707}$, $\lg R_{718}$, $\lg R_{702}$, $\lg R_{723}$	0.04851	0.8093	0.0559	1.0217
$\sqrt{R_{707}}$, $\sqrt{R_{713}}$, $\sqrt{R_{556}}$, $\sqrt{R_{718}}$, $\sqrt{R_{551}}$	0.0619	0.8038	0.0705	1.0287
$(1/R_{744})'$, $(1/R_{676})'$, $(1/R_{505})'$, $(1/R_{541})'$, $(1/R_{572})'$	0.1485	0.7704	0.1123	1.0733
$(\lg R_{525})'$, $(\lg R_{561})'$, $(\lg R_{551})'$, $(\lg R_{520})'$, $(\lg R_{459})'$	0.1441	0.7657	0.1371	1.0799
$(\sqrt{R_{444}})'$, $(\sqrt{R_{525}})'$, $(\sqrt{R_{618}})'$, $(\sqrt{R_{566}})'$, $(\sqrt{R_{530}})'$	0.2245	0.7373	0.1800	1.1215
$(1/R_{419})''$, $(1/R_{697})''$, $(1/R_{639})''$, $(1/R_{424})''$, $(1/R_{629})''$	0.5757	0.54871	0.4828	1.5070
$(\lg R_{577})''$, $(\lg R_{424})''$, $(\lg R_{414})''$, $(\lg R_{429})''$, $(\lg R_{439})''$	0.7411	0.4196	0.7577	1.9705
$(\sqrt{R_{449}})''$, $(\sqrt{R_{439}})''$, $(\sqrt{R_{490}})''$, $(\sqrt{R_{454}})''$, $(\sqrt{R_{434}})''$	0.7308	0.4286	0.6975	1.9292

TABLE 5. RF Models and their Accuracy Parameters (n = 75)

Independent variable	C-R^2	C-RMSE	C-S	C-RPD
R_{541}', R_{536}', R_{530}', R_{525}', R_{587}'	0.9153	0.2560	0.8087	3.2303
R_{454}'', R_{424}'', R_{429}'', R_{434}'', R_{546}''	0.9094	0.2653	0.7981	3.1172
$1/R_{713}$, $1/R_{707}$, $1/R_{718}$, $1/R_{702}$, $1/R_{723}$	0.8523	0.3533	0.6754	2.3407
$\lg R_{713}$, $\lg R_{707}$, $\lg R_{718}$, $\lg R_{702}$, $\lg R_{723}$	0.8521	0.3541	0.6726	2.3351
$\sqrt{R_{707}}$, $\sqrt{R_{713}}$, $\sqrt{R_{556}}$, $\sqrt{R_{718}}$, $\sqrt{R_{551}}$	0.8519	0.3523	0.6798	2.3474
$(1/R_{744})'$, $(1/R_{676})'$, $(1/R_{505})'$, $(1/R_{541})'$, $(1/R_{572})'$	0.9279	0.2549	0.778	3.2438
$(\lg R_{525})'$, $(\lg R_{561})'$, $(\lg R_{551})'$, $(\lg R_{520})'$, $(\lg R_{459})'$	0.9187	0.2526	0.8104	3.2735
$(\sqrt{R_{444}})'$, $(\sqrt{R_{525}})'$, $(\sqrt{R_{618}})'$, $(\sqrt{R_{566}})'$, $(\sqrt{R_{530}})'$	0.9248	0.2388	0.8318	3.4623
$(1/R_{419})'$, $(1/R_{697})'$, $(1/R_{639})'$, $(1/R_{424})'$, $(1/R_{629})'$	0.8654	0.3492	0.6655	2.3680
$(\lg R_{577})''$, $(\lg R_{424})''$, $(\lg R_{414})''$, $(\lg R_{429})''$, $(\lg R_{439})''$	0.9274	0.2379	0.8254	3.4763
$(\sqrt{R_{449}})''$, $(\sqrt{R_{439}})''$, $(\sqrt{R_{490}})''$, $(\sqrt{R_{454}})''$, $(\sqrt{R_{434}})''$	0.9157	0.2534	0.8160	3.2630

Construction of RF model. To construct the RF model to estimate chlorophyll content, the transformed spectral data for the five wavelengths with larger absolute correlation coefficient values were the independent variables, and chlorophyll content was the dependent variable. The proportion of training samples was 50%; the number of trees in the random forest was 300, and the number of variable trees randomly sampled at nodes was 2. The results are shown in Table 5.

Several RF models were constructed with various spectral reflectance transformations as the independent variables. For all the RF models, the C-R^2 values were >0.800, the C-RMSE values were <0.400, the C-S values were >0.600, and the C-RPD values were >2.0 (Table 5). The RF model with (lg$R$)″ as the independent variable had the highest C-RPD value of 3.4763 and the lowest C-RMSE value of 0.2379.

Model verification. To test the reliability of the models, the chlorophyll content and spectral reflectance data of 25 samples were used to test the PLS model with data transformed with ($R^{1/2}$)′ as the independent variables, and the SVM and RF models with data transformed with (lg$R$)″ as the independent variables. The results are shown in Fig. 2. As illustrated in Fig. 2, among the tested PLS, SVM and RF models, the SVM model with data transformed with (lg$R$)″ as the independent variables had the highest V-R^2 and V-RPD values of 0.7372 and 1.8810, respectively, and the lowest V-RMSE value (0.4477). Therefore, the SVM model with (lgR_{577})″, (lgR_{424})″, (lgR_{414})″, (lgR_{429})″, (lgR_{439})″ as the independent variables was the most accurate for predicting the chlorophyll content of apple leaves.

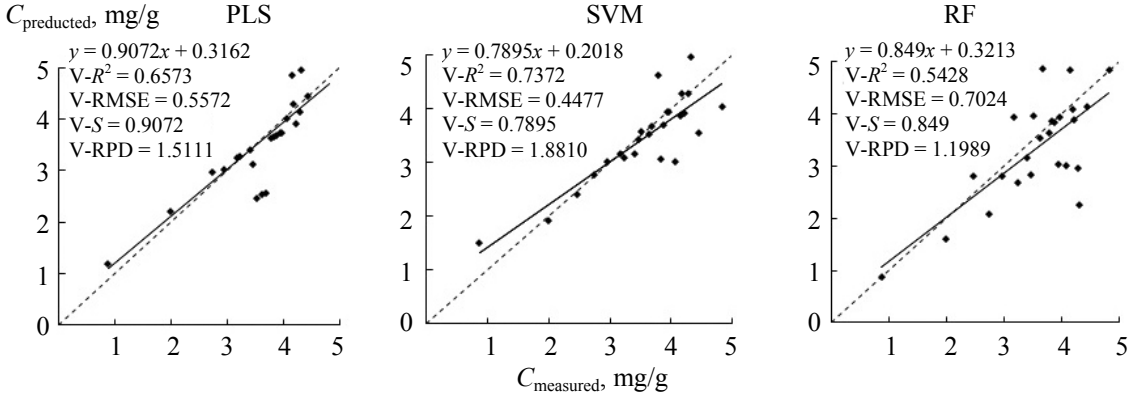

Fig. 2. Scatter plot of measured and predicted chlorophyll contents based on PLS, SVM, and RF models (n = 25).

Conclusion. The wavelengths sensitive to chlorophyll content were identified as 577, 424, 414, 429, and 439 nm by the maximum correlation coefficient method. Transformed data for these sensitive wavelengths were used as independent variables in the models. An SVM model was constructed with (lgR_{577})″, (lgR_{424})″, (lgR_{414})″, (lgR_{429})″, (lgR_{439})″ as independent variables, and its V-R^2, V-RMSE, and V-RPD values were 0.7372, 0.4477, and 1.8810, respectively. Among all the models, this SVM model had the highest V-R^2 and V-RPD values and the lowest V-RMSE value. Therefore, this SVM model was suitable for the quantitative estimation of chlorophyll content in apple leaves.

Acknowledgment. This paper was supported by the National Nature Science Foundation of China (41671346) and the Funds of Shandong "Double Tops" Program (SYL2017XTTD02).

REFERENCES

1. X. G. Xu, C. J. Zhao, J. H. Wang, W. J. Huang, C. J. Li, H. J. Liu, *Spectrosc. Spect. Anal.*, **31**, 188–191 (2011).
2. D. N. H. Horler, J. Barber, J. P. Darch, D. C. Ferns, A. R. Barringer, *Adv. Space Res.*, **3**, 175–179 (1983).
3. A. Maccioni, G. Agati, P. Mazzinghi, *J. Photochem. Photobiol. B*, **61**, 52–61 (2001).
4. V. Ciganda, A. Gitelson, J. Schepers, *J. Plant Physiol.*, **166**, 157–167 (2009).
5. M. Schlemmer, A. Gitelson, J. Schepers, R. Ferguson, Y. Peng, J. Shanahan, D. Rundquist, *Int. J. Appl. Earth Observ.*, **25**, 47–54 (2013).

6. E. R. Hunt Jr , P. C. Doraiswamy, J. E. McMurtrey, C. S. T. Daughtry, E. M. Perry, B. Akhmedov, *Int. J. Appl. Earth Observ.*, **21**, 103–112 (2013).
7. Z. N. Gong, Y. L. Zhao, W. J. Zhao, C. Lin, T. X. Cui, *Acta Ecologica Sinica*, **34**, 5736–5745 (2014).
8. X. C. Zhu, G. X. Zhao, R. Y. Wang, F. Dong, L. Wang, T. Lei, *Sci. Agricult. Sinica*, **43**, 1189–1197 (2010).
9. X. J. Yue, D. P. Quan, T. S. Hong, J. Wang, X. M. Qu, H. M. Gan, *Trans. CSAE*, **31**, 294–302 (2015).
10. S. W. Wang, S. Zhao, C. L. Zhang, Z. B. Su, L. F. Wang, Y. Zhao, *Trans. CSAE*, **32**, 149–154 (2016).
11. H. S. Li, *Plant Physiology and Biochemistry Experimental Principles and Techniques*, Beijing, Higher Education Press, 45 (2000).
12. T. He, *Study on Land Quality Hyperspectral Remote Sensing Monitoring Method*, Wuhan University, Wuhan, 44–54 (2003).
13. Y. He, F. Liu, X. L. Li, Y. N. Shao, *Spectroscopy and Imaging Technology in Agriculture*, Science Press, Beijing, 93, 109, 145 (2016).
14. N. Cristianini, J. S. Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Method*, Publishing House of Electronics Industry, Beijing, 24–26 (2004).
15. H. Yan, X. G. Zhang, Y. Q. Ma, Y. D. Li, *Acta Automat. Sinica*, **28**, 450–455 (2002).
16. L. Breiman, *Machine Learning*, **45**, 5–32 (2001).
17. E. Gaston, J. M. Frias, P. J. Cullen, C. P. O'Donnell, A. A. Gowen, *J. Agr. Food Chem.*, **58**, 6226–6233 (2010).