IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 9, 2019, accepted October 21, 2019, date of publication October 28, 2019, date of current version November 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949866

Estimating Total Leaf Chlorophyll Content of Gannan Navel Orange Leaves Using Hyperspectral Data Based on Partial Least Squares Regression

ZHONGZHENG PENG, LIXIN GUAN, YUBO LIAO, AND SUYUN LIAN
Intelligent Control Engineering and Technology Research Center, Gannan Normal University, Ganzhou 341000, China

Corresponding author: Lixin Guan (lxguan@gnnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61741103, in part by the Guangdong Natural Science Foundation of China under Grant 2017A030313343, in part by the Shenzhen Subject Layout Project of China under Grant 20180172, and in part by the Shenzhen Science and Technology Projection of China under Grant JCYJ20160520173822387.

ABSTRACT The goal of this study was to model the total leaf chlorophyll content (LCC_{tot}) of Gannan navel orange leaves using a field imaging spectroscopy system in the visible and near-infrared domain. The spectral range from 400 to 1000 nm with 176 wavebands (a wavelength interval of 3.41 nm) or 360 wavebands (a wavelength interval of 1.67 nm), labeled as ''Datasets_1.67'' and ''Datasets_3.41'', respectively, were used. Although different spectral data types were used, better prediction results for LCC_{tot} were based on Datasets_1.67 for LCC_{tot} prediction. Several prediction models of LCC_{tot} were built based on partial least squares regression (PLSR), artificial neural networks (ANN), ordinary least squares regression (OLSR), and stepwise linear regression (SLR) using full spectral and effective wavelength (EW) data (raw spectral (RS), first derivative spectral (FDS) and second derivative spectral (SDS) data). The determination coefficient (R^2), the root mean square error (RMSE) and the residual predictive deviation (RPD) were used to evaluate the reliability and accuracy of the predicted LCC_{tot} values. As a result, 14 (7 obtained from Datasets_1.67, 7 obtained from Datasets_3.41), 39 (21 obtained from Datasets_1.67, 18 obtained from Datasets_3.41) and 50 (27 obtained from Datasets_1.67, 23 obtained from Datasets_3.41) wavebands were selected from the RS data, FDS data and SDS data, respectively, as the EWs for LCC_{tot} prediction of navel orange leaves. After that, PLSR and ANN predictive models were established using full spectra, and OLSR and SLR predictive models were built using the selected EWs. The experimental results demonstrated that these various regression methods were useful for estimating LCC_{tot} in the order of PLSR models established using full spectra from RS data (F-RS-PLSR) > PLSR models established using full spectra from SDS data (F-SDS-PLSR) > PLSR models established using full spectra from FDS data (F-FDS-PLSR) > SLR models established using EWs by RS data (EWs-RS-SLR). However, models built with ANN and OLSR, where the RPD values were less than 3, cause the models to be inaccurate. Finally, in comparison, the F-RS-PLSR model exhibited the best performance of LCC_{tot} estimation; with the number of principal components (Pcs) = 5, this model provided high values of the R^2 of calibration (C-R^2) = 0.92 and the R^2 of validation (V-R^2) = 0.96, small values of the RMSE of calibration (C-RMSE)=0.05 mg/g and the RMSE of validation (V-RMSE) = 0.19 mg/g, and sufficient the RPD of calibration (C-RPD)=17.00 and the RPD of validation (V-RPD)=3.63 values. Overall, the best modeling method was PLSR. Hence, the PLSR applicability for assessing chlorophyll content in navel orange leaves was demonstrated.

INDEX TERMS Chlorophyll, hyperspectral data, navel oranges, partial least squares.

I. INTRODUCTION

Chlorophyll is the main photosynthetic pigment present in green plants and plays an important role in controlling carbon exchange and plant productivity [1], [2]. The chlorophyll content increases in young expanding leaves, reaches the highest value at maturity, and then decreases significantly during senescence [3], [4]. Therefore, the chlorophyll content of plant leaves correlated with the nutritional status can theoretically be used as a marker of the growth status of plants.

The associate editor coordinating the review of this manuscript and approving it for publication was Qiangqiang Yuan.

Measurements and estimates of chlorophyll content are regarded as a meaningful indicator of plant health, including nitrogen deficiency, water stress and certain diseases [2], which can provide theoretical guidance for crop nutrient diagnosis and field management.

The traditional wet-chemical method for measuring chlorophyll is precise but costly, time-consuming and inapplicable to large-scale analysis. Hence, scientists have been developing convenient and rapid methods for the measurement of leaf chlorophyll utilizing its unique optical absorption feature. Extracting chlorophyll information from the spectral features of plants has become a major means of estimating chlorophyll contents because of its advantages of being fast, nondestructive and large-scale [1], [5]–[8]. Numerous studies have been conducted using spectral data to retrieve chlorophyll information as a function of time and space in environments such as the ground or airborne and spaceborne environments [9]–[16].

The nearly continuous spectral characteristics of vegetation can be obtained using hyperspectral remote sensing in visible and near-infrared bands, which makes the quantitative analysis and detail extraction of vegetation possible [17], [18]. Several remote sensing techniques have been proposed to estimate chlorophyll content at the leaf or canopy level. Currently, chlorophyll retrieval methods can be summarized into two main categories: the first category is that of the leaf or canopy radiative transfer models such as PROSPECT, LIBERTY, SAIL and PROSAIL, etc. [19]–[23]. These methods have been developed based on physical theory, are highly accurate for predictive modeling and perform well under different model hypotheses. The second category is that of empirical approaches based on the absorption characteristics of pigments, and the relationship between spectral features and biophysical and biochemical parameters are established. By comparing these two chlorophyll retrieval methods, the radiation transfer models of the leaf or canopy are so complex and so many inputs and parameters required that it is difficult to obtain. However, these empirical approaches are popularly used, because these kinds of methods are simple, have a lower computational cost and need fewer parameters.

Several studies have shown that the band positions at 550 and 700 nm are the most effective bands for estimating the chlorophyll content in leaves [7], [14], [24]. However, the near infrared spectra are overwhelmed by the molecular overtone and combination bands. Therefore, it can be difficult to assign specific features to some chemical components of crops. Partial least squares regression (PLRS) appears as a promising method for constructing predictive models for chlorophyll content on the basis of the many and highly correlated reflectance data in the NIR spectrum of leaf samples [25]. PLSR can be applied to reduce collinear spectral variables to a few noncorrelated principal components (Pcs) [26], thereby avoiding overfitting problems. Unlike PLSR, artificial neural networks (ANNs), which are commonly used in remote sensing to predict vegetation parameters and crop yields [27], [28], inevitably suffer from overfitting problems. Moreover, the design parameters and implementation of ANNs require complex and time-consuming processes, and their performance can be weakened when low-dimensional datasets are used [29]. Several studies have been undertaken to estimate the chlorophyll contents of plants using PLSR. Hansen and Schjoerring [30] reported the reflectance measurements of canopy biomass and chlorophyll content in wheat crops using normalized difference vegetation indices and PLSR. Wagner *et al.* [31] reported the use of proximal Vis-NIR spectrometry to retrieve substance concentrations in surface waters utilizing PLSR modeling. Liu *et al.* [32] also obtained nondestructive detection of rape leaf chlorophyll levels based on Vis-NIR spectroscopy using PLSR. In addition, PLSR can be performed on either raw spectral data or on their various derivatives. The calculation of the first and second derivatives of spectral data enhances resolution and removes background effects. Specifically, first derivatives remove additive constant background effects, while second derivatives remove baseline linear slope variations and additive effects [33]. This way, the likely improvement caused by the first and second derivatives on the chlorophyll content prediction ability from the absorbance NIR spectra could also be studied. Moreover, preceding the PLSR analysis, exploratory principal component analyses (PCA) on the raw, first and second derivative datasets were performed to explore the extent of correlation and to obtain guidance on the adequate number of independent uncorrelated variables to be used in the subsequent multivariate calibration [34].

The main goal of this study was therefore 1) to use multivariate statistics, specifically, PLSR, to obtain a reliable model to predict the total leaf chlorophyll content (LCC_{tot}) of Gannan navel orange leaves from their NIR spectra and 2) to then establish an accurate, real-time plant chlorophyll content evaluation system to achieve accurate fertilization and nutritional status monitoring of fruit trees and to provide technical support for navel orange tree yield increase.

II. MATERIALS AND METHODS

A. STUDY AREA

Gannan (Gannan area (see Figure 1 for locations, latitude 114°93′ and longitude 25°83′), Jiangxi, China) is a hilly mountainous area spanning 45.6 million acres. It has mountain resources for planting navel oranges, of which 4.5 million acres are suitable for navel orange cultivation. The average annual precipitation and temperature of the region are 1606 mm and 18.8°C, respectively. In addition, the precipitation shows a distinct seasonal pattern with 80% of rainfall occurring from April through October. It is a typical subtropical humid monsoon climate, with an early spring, a long summer, a short autumn, a warm winter, four distinct seasons, abundant rainfall, sufficient sunshine, a long frost-free period, a large temperature difference between day and night from

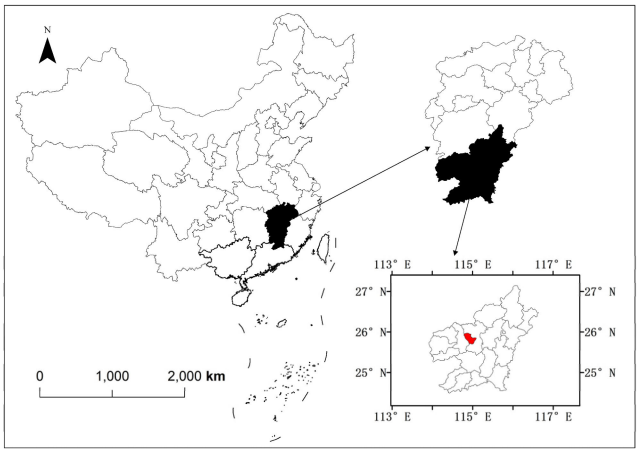

FIGURE 1. Location of the Gannan area used in this study.

September to November, a hot rainy season, and extremely beneficial conditions for navel orange planting.

At present, Gannan has become the largest planting area for Navel oranges, the third annual production origin of navel oranges in the world. Phenology dynamics of Gannan navel orange growth are listed as follows: spring shoots grow from February to April, summer shoots grow from May to June, autumn shoots grow from August to September, budding occurs in March, blooming occurs in April, the first physiological fruit drop occurs from May to early June, the second physiological fruit drop occurs in mid-June, leaf yellowing occurs from July to early August, and fruit ripen in late December. Gannan navel oranges have been listed as one of the nationally superior agricultural products, which is protected by geographical indications and won the honor of being a "famous Chinese fruit".

B. PREPARATION OF LEAF SAMPLES

Leaves of Gannan navel orange plants were sampled for this study. Field sampling was conducted in a navel orange planting site stand located in the National Navel Orange Engineering Research Center, Gannan Normal University, Ganzhou, China. Ten navel orange trees were selected as experimental samples. The average height of the trees was 1.5 m, the crown width was 1.5 m, on average, and the spacing was 4 m×3 m. The soil in the navel orange orchard was loose organic red soil, which ensured good light conditions and suitable growth temperatures throughout the year.

Navel orange leaves were collected on 23 June, 29 June, 3 July, and 4 July 2018, between 9:00 and 11:00 or 14:00 and 16:00. The leaves were stored in sealed plastic bags and transported to the laboratory. A total of 8 leaves of each tree (2 facing east, 2 facing west, 2 facing south and 2 facing north) were collected, and healthy new and old leaves that were uniform in size were collected from the middle of the tree. The 80 samples were washed with clean water to clean the dust and other contaminants on the surface. Then, they were dried with clean, cotton cloth, put into fresh bags and numbered.

C. HYPERSPECTRAL DATA ACQUISITION AND PROCESSING

A field hyperspectral imaging system (GaiaField-V10E, Sichuan Dualix Spectral ImageTechnology Company Ltd., Sichuan, China) was used to acquire reflectance spectra of the leaf samples. The instrument was equipped with Specview graphical operation software and data storage for the spectrometer. Imaging spectrograph (ImSpectorV10E) covered the spectral range from 400 to 1000 nm with 176 wavebands (wavelength interval of 3.41 nm) or 360 wavebands (wavelength interval of 1.67 nm). The distance between the lens and the leaves was 10 cm. Before each measurement, black and white tiles was used for calibration by recording 10 measurements of each tile. Three-point average spectral reflectance was used as the spectral value of each navel orange leaf sample. Spectral data were collected at wavelength intervals of 1.67 nm and 3.41 nm and were labeled as "Datasets_1.67" (60 samples) and "Datasets_3.41" (80 samples), respectively.

For Datasets_1.67 and Datasets_3.41, 45 and 59 samples were randomly selected as calibration samples, and the remainder were implemented as validation samples. The derivative of a spectrum can improve spectral resolution, highlight the specificity of spectral features, enhance the ability to identify overlapping peaks and valleys in the spectrum, and eliminate background noise. Due to the discreteness of crop reflectance spectra, the first derivative of the spectra was approximated by the difference method, as shown in equation (1):

$$R'(\lambda_i) = \frac{dR(\lambda_i)}{d\lambda_i} = \frac{R(\lambda_{i+1}) - R(\lambda_{i-1})}{2\Delta\lambda} \quad (1)$$

The second derivative $R''(\lambda_i)$ of the spectrum can be obtained by calculating the first derivative of $R'(\lambda_i)$.

D. TOTAL LEAF CHLOROPHYLL CONTENT MEASUREMENTS

In this study, the chlorophyll content of the leaf samples was measured by spectrophotometry. Crushed fresh samples weighing 0.1 g to 0.2 g were added to a small amount of quartz sand, calcium carbonate powder and 2-3 ml 95% ethanol. Then, they were ground into a uniform pulp, mixed with 10 ml ethanol and placed in the dark at room temperature (25°C) for 10 minutes. The volume was brought to 25 ml with 95% ethanol, and the absorbance at 665 nm (A665) and 649 nm (A649) was subsequently measured with a V-5100 spectrophotometer. The chlorophyll content was calculated according to the reference (Arnon, 1949), using the correction equations (2) and (3):

$$Chl_a = (13.95A_{665}\text{-}6.80A_{649}) \times \frac{V}{W \times 1000} \quad (2)$$

$$Chl_b = (24.96A_{649}\text{-}7.32A_{665}) \times \frac{V}{W \times 1000} \quad (3)$$

where V is the volume of the extraction solution (ml) and W is the weight of the leaf sample (g). The total leaf chlorophyll content (LCC_{tot}) is equal to Chl a + Chl b (mg/g).

E. MEASUREMENT MODELING METHODS

1) PARITIAL LEAST SQUARES REGRESSION (PLSR)

PLS, based on principal component analysis (PCA), was first introduced by Wold *et al.* [35].. PLSR is the PLS approach in its simplest form and is used to relate two data matrices, namely, the input variables (X) and the response variables (Y). PLSR could determine a linear relationship between X and Y by simultaneously decomposing and screening data matrices with a few independent principal components (Pcs). These data matrices could be noisy, collinear, and incomplete.

$$Y = \alpha_0 + \alpha_1 S_{1\lambda} + \alpha_2 S_{2\lambda} + \ldots + \alpha_n S_{n\lambda} \tag{4}$$

where α_n is the regression coefficient obtained by the linear regression of $S_{n\lambda}$ versus Y_λ in a calibration iteration process and $S_{n\lambda}$ is the score of the nth principal component at spectral band λ.

2) ARTIFICIAL NEURAL NETWORKS (ANN)

In this study, artificial neural networks (ANN) were used. The NN were made up of an input layer, a hidden layer and an output layer. The number of nodes in the input layer was determined by the spectral dimensions obtained by the hyperspectral system. For the generation, training and implementation of the neural network in this experiment, we used the back propagation NN (BP-NN) toolbox of MATLAB software. For the sake of minimizing the mean square error (MSE) between the output of the neural network and LCCtot, the neural network was iteratively trained. In each BP-NN iteration, the network weights (w_i) and biases (b_i) were adjusted along with the gradient decrease of the MSE. To deal with the iteration problem, a transformation T (activation function) was introduced in equation (5), which can provide an output value y for the network:

$$y = T(\sum_{i=1}^{n} w_i X_i + b_i) \tag{5}$$

In this experiment, the minimum MSE, minimum gradient, and maximum number of iterations (epochs) were set to 10^{-3}, 10^{-6} and 2000, respectively. If any of the above conditions were met, the training process of the artificial neural network stopped.

3) ORDINARY LEAST SQUARES REGRESSION (OLSR)

The ordinary least squares (OLS) is a linear approximation that minimizes the sum of the squares of the distances between the observation points and the estimated points. The slope formula of OLS estimation is $\hat{\beta}_1 = S_{XY}/S_{XX}$. OLS is more suitable for the cases in which one of the two variables is a random variable:

$$\sum (y_i - \beta_0 - \beta_1 x_i)^2 \tag{6}$$

4) STEPWISE LINEAR REGRESSION (SLR)

Stepwise linear regression (SLR) is a method of regressing multiple factors prior to linear filtering to obtain retention factors. The relevant variables that are filtered out by SLR each time are the weakest. Finally, we observed the order of the contributions that best explain the distribution needed to get the required variables. Therefore, the linear relationship between spectral indices and photosynthetic pigments was modeled using a stepwise linear regression (SLR) method.

F. EVALUATION OF MODELS

The performance of a predictive model is usually evaluated by several indices, including the root mean square error of calibration (C-RMSE) and validation (V-RMSE), the determination coefficient of calibration (C-R^2) and validation (V-R^2) and the residual predictive deviation of calibration (C-RPD) and validation (V-RPD). A good model should provide high values of V-R^2 and RPD (>3), small values of C-RMSE and V-RMSE, and a small variation between C-RMSE and V-RMSE [3], [36]. R^2, RMSE and RPD were calculated as follows:

$$R^2 = \frac{\sum_{i=1}^{n} (y_i - y_i')^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i}^{n} (y_i' - y_i)^2}{n}} \tag{8}$$

where n is the number of samples in the training or testing set, y_i is the reference value, y_i' is the predicted value, and $\bar{y}$ is the average value of all reference values.

$$RPD = \frac{SD}{RMSE} \tag{9}$$

where SD is the standard deviation of the reference values of the samples.

III. RESULTS AND DISCUSSION

A. SPECTRAL REFLECTANCE

The intensity of radiant flux for a given surface and a given wavelength can be measured by a hyperspectral sensor, i.e., a physical quantity in units of W/srm^2. To be exact, the light emitted and reflected by each surface unit of the object exists in the form of a spectrum of hundreds of channels, and it is captured by a sensor for obtaining a spectral response curve (Figure 2(a)). The complex combination of the scattering process and the overlapping absorption of moisture and biochemical components in plant leaves forms the corresponding spectrum [37], [38]. The peak at approximately 550 nm is related to respiratory pigments [39]. The red edge usually refers to the 660-770 nm region [40], which is caused by the combined effects of strong chlorophyll absorption in the red wavelengths and high reflectance in the NIR wavelengths due to internal leaf scattering [41]. There is a small peak at 970 nm, which is related to the water content of the leaves and corresponds to the O-H second overtone stretch [42]–[44]. To carry out an in-depth analysis of these characteristics,

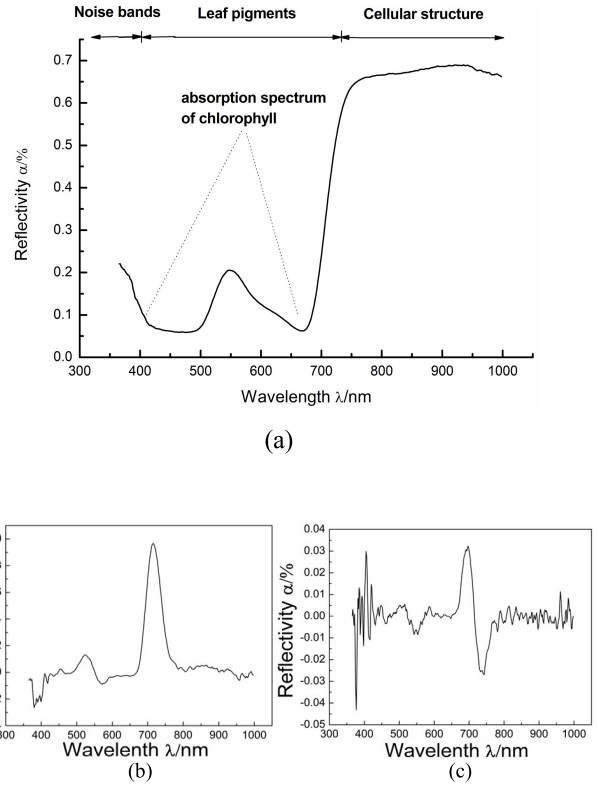

FIGURE 2. Spectral reflectance spectra with a wavelength interval of 1.67 nm: (a) Raw spectral data, (b) first derivative, and (c) second derivative.

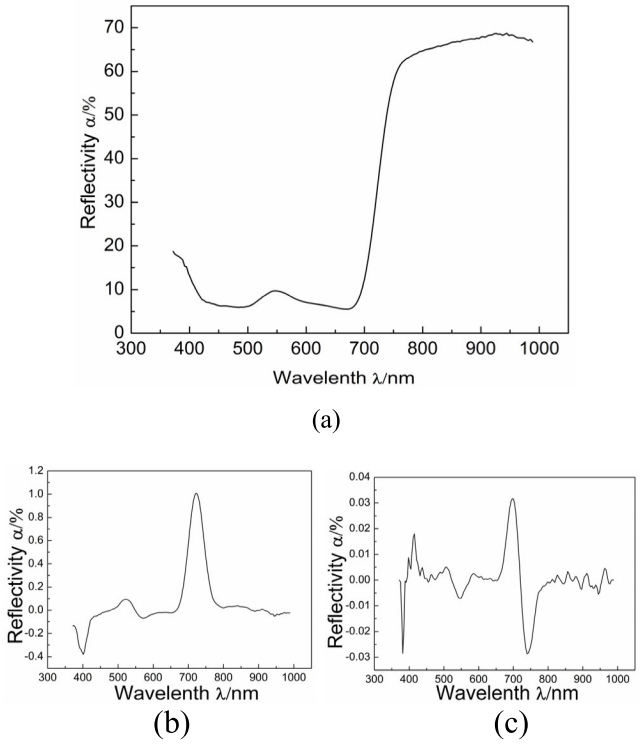

FIGURE 3. Spectral reflectance spectra with a wavelength interval of 3.41 nm: (a) Raw spectral data, (b) first derivative, and (c) second derivative.

quantitative analysis of various parameters of crops using spectral information is the goal of crop hyperspectral remote sensing monitoring.

Regarding Datasets_1.67 and Datasets_3.41, each dataset randomly selects one sample to represent the reflectance spectra of each dataset. The spectral reflectance of samples with wavelength intervals of 1.67 nm and 3.41 nm are shown in Figures 2 and 3.

B. MEASURED LCC_{TOT} OF GANNAN NAVEL ORANGE LEAVES

Tables 1 and 2 show a summary of the statistical analysis of the LCC_{tot} of all samples in the calibration and validation sets. The LCC_{tot} exhibited evident variation in the calibration set, ranging from 0.57 to 3.37 mg/g for Datasets_1.67 and from 0.57 to 4.92 mg/g for Datasets_3.41, which contributed to the generation of robust calibration models.

TABLE 1. CHL A, CHL B, and LCCtot values of leaves for Datasets_1.67.

Chemical Constituent	N	Range	Mean±SD
Calibration set			
Chla (mg/g)	45	0.48-2.73	1.71 ± 0.66
Chlb (mg/g)	45	0.08-0.72	0.42 ± 0.19
LCCtot (mg/g)	45	0.57-3.37	2.12 ± 0.85
Validation Set			
Chla (mg/g)	15	0.96-2.53	2.00 ± 0.54
Chlb (mg/g)	15	0.23-0.75	0.52 ± 0.16
LCCtot (mg/g)	15	1.19-3.28	2.57 ± 0.69

N = number of samples, and SD = standard deviation.

TABLE 2. CHL A, CHL B, and LCCtot values of leaves for Datasets_3.41.

Chemical Constituent	N	Range	Mean±SD
Calibration set			
Chla (mg/g)	59	0.48-3.70	1.95 ± 0.81
Chlb (mg/g)	59	0.08-1.67	0.50 ± 0.27
LCCtot (mg/g)	59	0.57-4.92	2.46 ± 1.06
Validation Set			
Chla (mg/g)	21	0.51-3.59	1.83 ± 0.90
Chlb (mg/g)	21	0.14-1.04	0.51 ± 0.27
LCCtot (mg/g)	21	0.68-4.63	2.34 ±1.16

C. ANALYSIS OF PLSR MODELING RESULTS

In this section, PLSR was used to generate a robust relationship equation between the spectral reflectance and the LCC_{tot} of samples. Several prediction models of LCC_{tot} were built using PLSR for three spectral datasets (raw spectral data, first derivative spectral data and second derivative spectral data).

1) MODELING WITH RAW SPECTRA

Table 3 shows the performance indices of the LCC_{tot} predictive model established based on PLSR by using raw spectral data (RS-PLSR). As shown in Table 3, with the same Pcs, Datasets_1.67 exhibits overall better performance of LCC_{tot}

TABLE 3. The performance indices of the RS-PLSR models.

Pcs	Datasets_1.67						Datasets_3.41					
	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD
1	0.39	0.30	0.43	0.66	1.98	1.05	0.35	0.41	0.71	0.90	1.49	1.29
2	0.88	0.90	0.09	0.27	9.44	2.56	0.71	0.86	0.07	0.42	15.14	2.76
3	0.89	0.92	0.08	0.26	10.63	2.65	0.80	0.86	0.22	0.44	4.82	2.64
4	0.90	0.92	0.07	0.27	12.14	2.56	0.83	0.86	0.19	0.45	5.58	2.58
5	0.92	0.96	0.05	0.19	17.00	3.63	0.86	0.86	0.15	0.43	7.07	2.70
6	0.93	0.94	0.05	0.21	17.00	3.29	0.89	0.88	0.12	0.40	8.83	2.90
7	0.95	0.94	0.03	0.23	28.33	3.00	0.92	0.86	0.09	0.44	11.78	2.64
8	0.96	0.92	0.03	0.26	28.33	2.65	0.93	0.86	0.08	0.43	13.25	2.70
9	0.97	0.92	0.02	0.31	42.50	2.23	0.93	0.86	0.07	0.42	15.14	2.76
10	0.98	0.90	0.01	0.30	85.00	2.30	0.95	0.84	0.05	0.47	21.20	2.47

Pcs = number of principal components

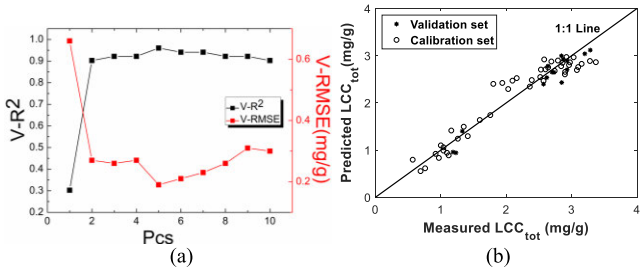

FIGURE 4. (a) Relations between V-R^2, V-RMSE and Pcs of RS-PLSR model; (b) linear regression plot of the RS-PLSR model.

estimation than Datasets_3.41. Therefore, in the next part of this paper, the retrieval performance of the LCC_{tot} for Datasets_1.67, with a wavelength interval of 1.67 nm, will be focused on.

With the same datasets, when the number of Pcs is in the range of 1 to 5, the values of V-R^2 and V-RPD increase with the larger Pcs, and when the number of Pcs is in the range of 5 to 10, the values of V-R^2 and V-RPD decrease with increasing Pc. It should be noted that the trend of V-RMSE is opposite to that of V-R^2 and V-RPD. The relations between V-R^2, V-RMSE and Pcs are presented in Figure 4(a). When Pcs = 5, the predictive model of LCC_{tot} for Datasets_1.67 exhibits the best performance, C-R^2 =0.92, V-R^2 = 0.96, C-RMSE = 0.05 mg/g, V-RMSE = 0.19 mg/g C-RPD = 17.00 and V-RPD = 3.63. However, there is evident variation between C-RMSE (0.05 mg/g) and V-RMSE (0.19 mg/g) and between C-RPD (17.00) and V-RPD (3.63).

Linear regression plots of the RS-PLSR model for Datasets_1.67 are presented in Figure 4(b) (Pcs = 5). As shown in Figure 4(b), the plots of the calibration and validation sets are scattered around the ideal 1:1 line, and the results indicate that the LCCtot in Gannan navel orange leaves can be accurately estimated by the RS-PLSR model.

TABLE 4. The performance indices of the FDS-PLSR AND SDS-PLSR models.

Pcs	FDS-PLSR						SDS-PLSR					
	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD
1	0.89	0.92	0.28	0.26	3.04	2.65	0.90	0.92	0.27	0.26	3.15	2.65
2	0.93	0.94	0.22	0.21	3.86	3.29	0.95	0.94	0.19	0.22	4.47	3.14
3	0.96	0.94	0.16	0.23	5.31	3.00	0.98	0.92	0.12	0.26	7.08	2.65
4	0.98	0.92	0.12	0.31	7.08	2.23	0.99	0.92	0.07	0.31	12.14	2.23
5	0.99	0.88	0.10	0.38	8.50	1.82	1.00	0.92	0.04	0.30	21.25	2.30
6	0.99	0.85	0.08	0.42	10.63	1.64	1.00	0.92	0.03	0.30	28.33	2.30
7	1.00	0.81	0.06	0.45	14.17	1.53	1.00	0.92	0.02	0.29	42.50	2.38
8	1.00	0.81	0.04	0.46	21.25	1.50	1.00	0.92	0.01	0.29	85.00	2.38
9	1.00	0.81	0.03	0.49	28.33	1.41	1.00	0.92	0.01	0.30	85.00	2.30
10	1.00	0.79	0.03	0.50	28.33	1.38	1.00	0.92	0.00	0.30	171.00	2.30

2) MODELING WITH FIRST DERIVATIVE AND SECOND DERIVATIVE SPECTRA

For Datasets_1.67, Table 4 shows the performance indices of the LCC_{tot} predictive model established based on PLSR by using first derivative spectra (FDS-PLSR) and second derivative spectra (SDS-PLSR).

As shown in Table 4, for the FDS-PLSR model, the value of V-R^2 and V-RPD with a small number of Pcs (1-4) are higher than those with a large number of Pcs (5-10). For example, when the number of Pcs = 1, the value of V-R^2 is 0.92, which is much higher than that obtained from the RS-PLSR model (0.30); whereas, when the number of Pcs = 10, the value of V-R^2 is only 0.79, which is lower than that of the RS-PLSR model (0.90). The relations between V-R^2, V-RMSE and the number of Pcs of the FDS-PLSR model are presented in Figure 5(a). In comparison with the RS-PLSR model, these results indicate that the FDS-PLSR model exhibits better performances of LCC_{tot} estimation when the number of Pcs is small. It should be noted that when the number of Pcs = 2, the FDS-PLSR model exhibits the best performance, C-R^2 = 0.93, V-R^2 = 0.94, C-EMSE = 0.22 mg/g, V-RMSE = 0.21 mg/g, C-RPD = 3.86 and V-RPD = 3.29; moreover, there is small variation between C-RMSE (0.22 mg/g) and V-RMSE (0.21 mg/g) and between C-RPD (3.86) and V-RPD (3.29).

For the SDS-PLSR model, the values of V-R^2 and V-RPD are higher than 0.92 and 2.23 for all numbers of Pcs. The relations between V-R^2, V-RMSE and the number of Pcs of the SDS-PLSR model are shown in Figure 5(b). The results show that the performances of the SDS-PLSR model are better for all numbers of Pcs, in general. With Pcs = 2, the SDS-PLSR model shows the best performance, C-R^2 =0.95, V-R^2 = 0.94, C-RMSE=0.19 mg/g, V-RMSE = 0.22 mg/g, C-RPD = 4.47 and V-RPD = 3.14. there is small variation between C-RMSE (0.19 mg/g) and V-RMSE (0.22 mg/g); however, the variation between C-RPD (4.47) and V-RPD (3.14) is evident.

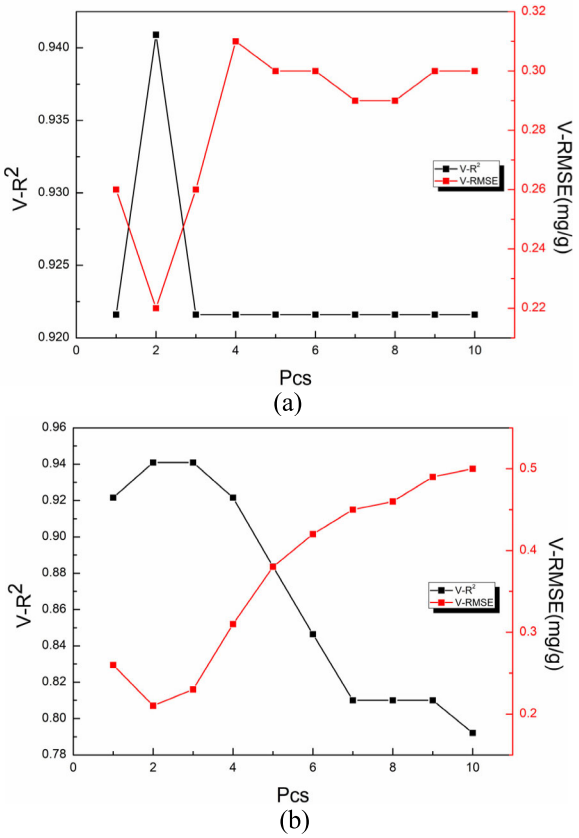

FIGURE 5. Relations between V-R^2, V-RMSE and Pcs: (a) FDS-PLSR model, (b) SDS-PLSR model.

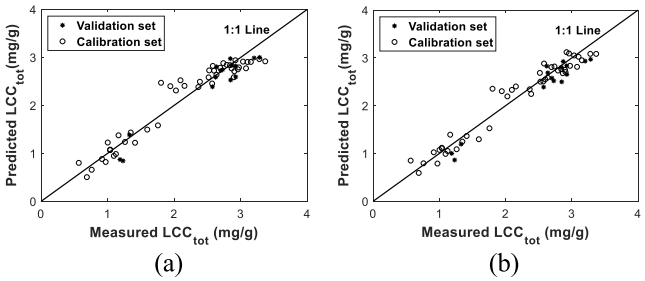

FIGURE 6. Linear regression plots: (a) FDS-PLSR model, (b) SDS-PLSR model.

Linear regression plots of the FDS-PLSR and SDS-PLSR model are presented in Figure 6 (Pcs = 2). As shown in Figure 6, the plots of the calibration and validation sets are closely scattered around the ideal 1:1 line.

D. ANALYSIS OF ANN MODELING RESULTS

The ANN model was built using the calibration set and then used to invert the quantitative relationships between spectral reflectance and LCC_{tot} in navel orange leaves. Table 5 shows the summarized validation criteria computed for the calibration and validation datasets used in the ANN models. The results of Table 5 show that ANNs were able to learn and describe the nonlinear relationships between the full spectral data (RS data, FDS data and SDS data) and LCC_{tot}.

TABLE 5. The performance indices of the ANN models.

RM	SPD	Datasets_1.67						Datasets_3.41					
		C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD
	RS	0.96	0.85	0.17	0.71	5.00	0.97	0.92	0.89	0.30	0.83	3.53	1.40
ANN	FDS	0.92	0.84	0.24	0.71	3.54	0.97	0.91	0.82	0.28	0.85	3.79	1.36
	SDS	0.00	-0.37	0.84	0.55	1.01	1.25	0.00	-0.30	0.95	0.49	1.12	2.37

RM=regression method, SPD = spectral dataset

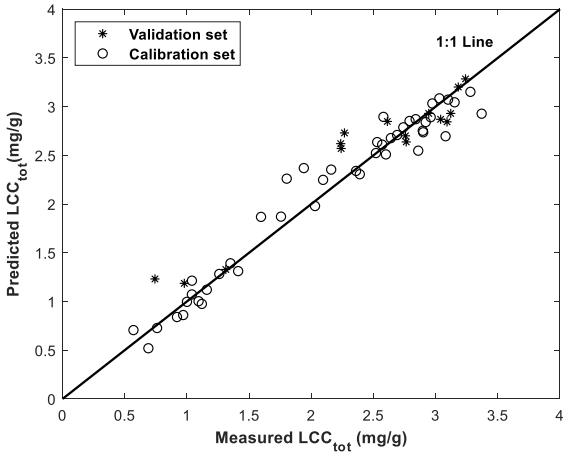

FIGURE 7. Linear relationship between the predicted LCC_{tot} content (RS-ANN) and measured LCC_{tot} content.

When SDS data were treated as the inputs, the validated accuracy of the ANN model is as poor as the calibration accuracy shown by V-R^2 = −0.37, which indicates that SDS data is not good for ANN modeling in this case. However, excellent correspondence was observed between the validation datasets using RS data and FDS data in the ANN model, which both have V-R^2 values greater than 0.80, whether in Dataset-1.67 or Dataset-3.41. In addition, it can be seen from the table that the effect when modeling with RS data is still better than when modeling with FDS data. In particular, the RMSEs of the ANN model with RS data in Datasets_1.67 were 0.17 mg/g (C-RMSE) and 0.71 mg/g (V-RMSE), which were below the 0.30 mg/g (C-RMSE) and 0.83 mg/g (V-RMSE) in Datasets_3.41. In comparison, it can be concluded that RS data (Datasets_1.67) is the most valid and effective of the spectral data tested among several ANN models. Figure 7 shows the regression relation between the predicted LCC_{tot} content based on the RS-ANN (Datasets_1.67) model and the measured LCC_{tot} concentration.

E. SELECTION OF EFFECTIVE WAVELENGTHS FOR SLR AND OLSR MODELING

The choice of effective wavelengths (EWs) is critical for reducing redundant information in hyperspectral data, and

TABLE 6. EWs based on RC curves.

SPD	Datasets_1.67	Datasets_3.41
RS	394 nm, 444 nm, 448 nm, 556 nm, 673 nm, 723 nm, 784 nm	388 nm, 418 nm, 488 nm, 556 nm, 671 nm, 721 nm, 779 nm
FDS	371 nm, 403 nm,446 nm, 456 nm, 464 nm,503 nm, 527 nm, 566 nm,666 nm, 678 nm, 705 nm,768 nm, 817 nm, 845 nm,861 nm, 869 nm, 883 nm,898 nm, 942 nm, 951 nm,985 nm	378 nm, 401 nm, 458 nm, 474 nm, 498 nm, 542 nm, 566 nm, 661 nm, 675 nm, 707 nm, 768 nm, 815 nm, 841 nm, 867 nm, 878 nm, 911 nm, 955 nm, 981 nm,
SDS	375 nm, 391 nm, 398 nm, 408 nm, 416 nm, 441 nm, 453 nm, 458 nm, 473 nm, 508 nm, 556 nm, 563 nm, 589 nm, 626 nm, 638 nm, 645 nm, 689 nm, 719 nm, 755 nm, 772 nm, 779 nm, 821 nm, 852 nm, 861 nm, 885 nm, 950 nm, 955 nm,	378 nm, 408 nm, 421 nm, 438 nm, 454 nm, 461 nm, 468 nm, 508 nm, 560 nm, 584 nm, 629 nm, 650 nm, 685 nm, 718 nm, 764 nm, 823 nm, 830 nm, 837 nm, 852 nm, 867 nm, 900 nm, 940 nm, 974 nm

TABLE 7. The performance indices of the OLSR and SLR models.

		Datasets_1.67						Datasets_3.41					
RM	SPD	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD
OLSR	RS	0.75	0.82	0.39	0.38	2.18	1.82	0.76	0.79	0.52	0.51	2.04	2.27
	FDS	0.90	0.89	0.25	0.30	3.40	2.30	0.79	0.76	0.47	0.57	2.26	2.04
	SDS	0.86	0.76	0.29	0.44	2.93	1.57	0.80	0.76	0.46	0.57	2.30	2.04
SLR	RS	0.90	0.92	0.26	0.19	3.27	3.63	0.78	0.86	0.50	0.43	2.12	2.70
	FDS	0.93	0.86	0.22	0.25	3.86	2.76	0.86	0.75	0.38	0.62	2.79	1.87
	SDS	0.93	0.82	0.22	0.28	3.86	2.46	0.76	0.83	0.46	0.51	2.30	2.27

it also helps to optimize the design of multispectral imaging systems [45]. Generally, EWs are typically located at the peaks of regression coefficient (RC) curves [3]. Based on the RC curves shown in Figure 8 and the EWs obtained from the RC curve are shown in Table 6. Fourteen (7 obtained from Datasets_1.67, 7 obtained from Datasets_3.41), 38 (20 obtained from Datasets_1.67, 18 obtained from Datasets_3.41) and 50 (27 obtained from Datasets_1.67, 23 obtained from Datasets_3.41) wavebands were selected from the RS data, FDS data and SDS data, respectively, as the EWs for LCC_{tot} prediction of navel orange leaves.

As shown in Table 7, several prediction models of LCC_{tot} are built on OLSR and SLR using EWs (Datasets_1.67 and Datasets_3.41). For LCC_{tot} prediction, the results of Datasets_1.67 were better than those of the Datasets_3.41.

Modeling with OLSR, FDS data (Dataset_1.67) produces the best results (C-R^2 = 0.90, V-R^2 = 0.89,C-RMSE = 0.25,V-RMSE = 0.30, C-RPD = 3.40 and V-RPD = 2.30). FDS data is better than RS data and SDS data in terms of R2, RMSE and RPD.

The SLR models built by the RS data provided relatively robust results (C-R^2 = 0.90, V-R^2 = 0.92, C-RMSE = 0.26,V-RMSE = 0.19, C-RPD = 3.27 and V-RPD = 3.63) for predicting LCC_{tot} compared with those provided by the other models established by using FDS and SDS data, which might have been affected by redundant information in the FDS and SDS data. RPD values were over 3, indicating reliable validation for analytical purposes. Therefore, in this case, EWs-RS-SLR is the recommended model for determination of LCC_{tot} in navel orange leaves.

Linear regression plots of the FDS-OLSR and RS-PLSR model are presented in Figure 9. As shown in Figure 9, the plots of the calibration and validation sets are closely scattered around the ideal 1:1 line. However, the scatter plot of the FDS-OLSR model is more dispersed than that of the RS-SLR model, illustrating that the modeling effect of the RS-SLR model is better than that of the FDS-OLSR model.

F. COMPARATIVE ANALYSIS OF THE OPTIMAL MODELS

Comparison of the prediction results of LCC_{tot} by using different regression methods (PLSR, ANN, OLSR and SLR) in both Datasets_1.67 and Datasets_3.41 using full and EW data (RS data, FDS data, and SDS data) are shown in Table 8. In this study, although different spectral data types were used, better prediction results for LCC_{tot} are based on Datasets_1.67. In particular, the reflectance data showed that the information extracted from Datasets_1.67 was more informative than that from Datasets_3.41. When modeling with full spectra or selected EWs, the results of the F-PLSR model were better than those of the EWs-OLSR and EWs-SLR models, which might have been affected by there being more information in the full spectra. In addition, the modeling effect of the F-ANN model was worse than that of the EWs-OLSR and EWs-SLR models. On the other hand, this also reflects the superiority of the PLSR technique for modeling hyperspectral data. In comparison, the results of the PLSR models were better than those of the ANN, OLSR and SLR models in predicting LCC_{tot} in the validation set. At the same time, in order to generate a reliable model, the value of RPD should be greater than 3. The evaluation results showed that the spectral data were compared with their accuracy in predicting LCC_{tot} in navel orange leaves by using RS-PLSR, FDS-PLSR, SDS-PLSR RS-ANN, EWs-FDS-OLSR and EWs-FDS-SLR models. The best prediction model for LCC_{tot} is the RS-PLSR model, with C-R^2 = 0.92,V-R^2 = 0.96, C-RMSE = 0.05, V-RMSE = 0.19, C-RPD = 17.00, and V-RPD =3.63.

IV. DISCUSSION

In this article, the LCC_{tot} of navel orange leaves was accurately estimated based on PLSR, ANN, OLSR, and SLR

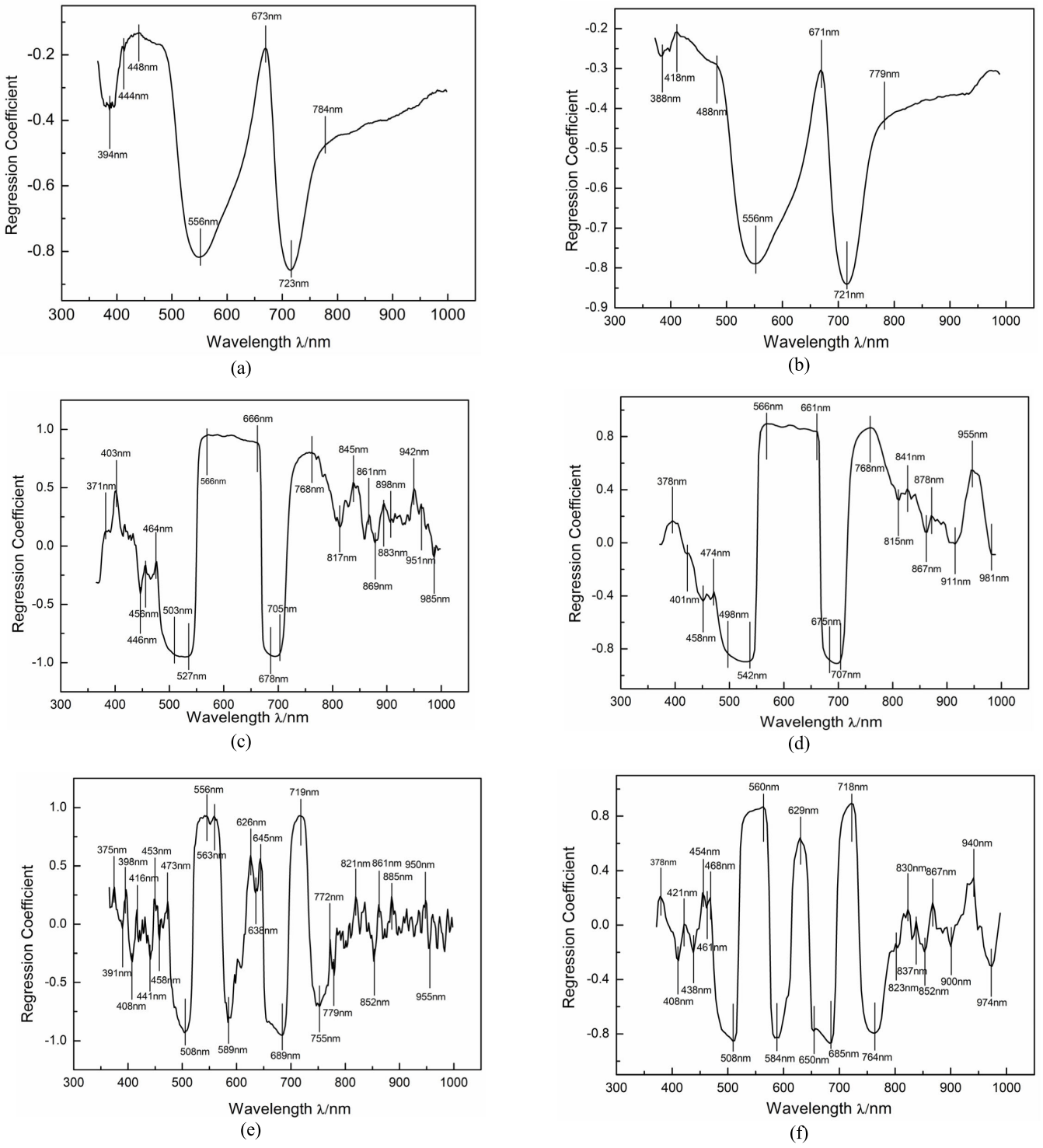

FIGURE 8. RC curves use RS data (Datasets_1.67 produced (a), Datasets_3.41 produced (b)), FDS data (Datasets_1.67 produced (c), Datasets_3.41 produced (d)) and SDS data (Datasets_1.67 produced (e), Datasets_3.41 produced (f)), respectively.

models using full spectral data, EWs from RS data, FDS data and SDS data. The R^2, RMSE and RPD were used to evaluate the reliability and accuracy of the predicted LCC_{tot} values. Among the four regression methods, PLSR exhibited the lowest RMSE values and relatively higher R^2 and RPD values compared with those of the other regression methods. The results indicate (Table 8) that the F-RS-PLSR model can be considered as an optimal, stable and reliable indicator for monitoring LCC_{tot} in navel orange leaves. The results from this study demonstrate that the statistical

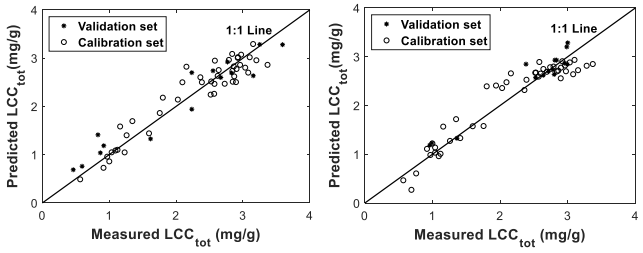

FIGURE 9. Linear regression plots: (a) FDS-OLSR model, (b) RS-PLSR model.

TABLE 8. Optimal modeling for prediction of lcc_{tot} in leaf.

Model	Pcs	Dataset	C-R^2	V-R^2	C-RMSE	V-RMSE	C-RPD	V-RPD
F-RS-PLSR	5	Datasets_1.67	0.92	0.96	0.05	0.19	17.00	3.63
F-FDS-PLSR	2	Datasets_1.67	0.93	0.94	0.22	0.21	3.86	3.29
F-SDS-PLSR	2	Datasets_1.67	0.95	0.94	0.19	0.22	4.47	3.14
F-RS-ANN	-	Datasets_1.67	0.96	0.85	0.17	0.71	5.00	0.97
EWs-FDS-OLSR	-	Datasets_1.67	0.90	0.89	0.25	0.30	3.40	2.30
EWs-RS-SLR	-	Datasets_1.67	0.90	0.92	0.26	0.19	3.27	3.63

F-RS-PLSR=PLSR models established using full spectral RS data, and EWs-FDS-OLSR = OLSR models established using EWs from FDS data.

approach (PLSR) has strong potential for monitoring the navel orange LCC_{tot} status. This is consistent with the results published by Cogliati *et al.* [15] and de Paz *et al.* [46]. However, the methods used in this study are based on specific datasets, and the findings may have limitations in applications to other datasets or other crops. To develop an accurate and robust model with high applicability and reliability, we should confirm the results for a wide range of varieties, growth stages and eco-sites in future work. A simulation technique could be used to improve the results based on composite spectral data, the spectra generated by the RTMs (e.g., PROSPECT or SAIL model) [21], [47]. Physical models can provide a clear link between biophysical variables and canopy spectral reflectance [48]. However, the disadvantage in using a physical method is the unsuitable nature of model inversion [49], showing that the inverse solution is not always unique because different combinations of canopy indices may generate almost identical spectra [50]. If the physical model requires a successful inversion, it usually needs to specify additional input variables, so using a physical model for inversion becomes increasingly complicated [49]. Therefore, from an application perspective, a statistical approach based on remotely sensed data may be a rapid and easy method to monitor chlorophyll status.

The question regarding whether there is full spectral or selected EWs to use in order to estimate Gannan navel orange leaf chlorophyll content has been answered in this paper. A number of robust and consistent spectral data and regression methods for LCC_{tot} are proposed and could be seen as priority indices to be tested in any follow up work. For instance, the results of the F-PLSR model were better than those of the EWs-OLSR and EWs-SLR models. Further research is recommended regarding whether the chlorophyll content can be retrieval using the image information in the hyperspectral data of the leaf. At the same time, we should also antinoise hyperspectral images, which is very important for extracting information from hyperspectral images.

V. CONCLUSION

In this study, four prediction methods (PLSR, ANN, OLSR and SLR) of LCC_{tot} for Gannan navel orange leaves using hyperspectral data (RS data, FDS data and SDS data) in the region of 400-1000 nm were evaluated. After obtaining the hyperspectral data, the RCs curves were used to select EWs. Full spectral data were used as model input vectors for PLSR and ANN, and EWs were used as model input vectors for OLSR and SLR. There were two kinds of spectral data, Datasets_1.67 and Datasets_3.41. The main conclusions are the following:

Although different spectral data types were used, better prediction results for LCC_{tot} were based on Datasets_1.67 for LCC_{tot} prediction. When modeling with full spectra or EWs, the results of the F-PLSR model were better than those of the EWs-OLSR and EWs-SLR models. In addition, the modeling effect of F-ANN was worse than that of the EWs-OLSR and EWs-SLR models. On the other hand, this also reflects the superiority of PLSR for modeling hyperspectral data. Moreover, since the RPD value of the ANN model is less than 3, the model established by the ANN is unreliable. In comparison, the F-RS-PLSR model exhibited the best performances of LCC_{tot} estimation; with the number of Pcs=5, it provides high values for C-R^2 (0.92) and V-R^2(0.96), small values for C-RMSE(0.05 mg/g) and V-RMSE (0.19 mg/g), and the C-RPD(17.00) and V-RPD (3.63) values exceed 3. In conclusion, using PLSR obtains a reliable method to predict LCC_{tot} in Gannan navel orange leaves from their NIR spectra.

REFERENCES

[1] M. Kalacska, M. Lalonde, and T. R. Moore, "Estimation of foliar chlorophyll and nitrogen content in an ombrotrophic bog from hyperspectral data: Scaling from leaf to image," *Remote Sens. Environ.*, vol. 169, pp. 270–279, Nov. 2015.

[2] P. Liu, R. Shi, and W. Gao, "Estimating leaf chlorophyll contents by combining multiple spectral indices with an artificial neural network," *Earth Sci. Informat.*, vol. 11, no. 1, pp. 147–156, Mar. 2018.

[3] K.-Q. Yu, Y.-R. Zhao, F.-L. Zhu, X.-L. Li, and Y. He, "Mapping of chlorophyll and SPAD distribution in pepper leaves during leaf senescence using visible and near-infrared hyperspectral imaging," *Trans. ASABE*, vol. 59, no. 1, pp. 13–24, 2016.

[4] R. Vicente, R. Morcuende, and J. Babiano, "Differences in Rubisco and chlorophyll content among tissues and growth stages in two tomato (Lycopersicon esculentum Mill.) varieties," *Agronomy Res.*, vol. 9, pp. 501–507, 2011.
[5] W. Kong, W. Huang, R. Casa, X. Zhou, H. Ye, Y. Dong, and R. Casa, "Off-nadir hyperspectral sensing for estimation of vertical profile of leaf chlorophyll content within wheat canopies," *Sensors*, vol. 17, no. 12, P. 2711, Dec. 2017.
[6] Y. Inoue, M. Guérif, F. Baret, A. Skidmore, A. Gitelson, and M. Schlerf, "Simple and robust methods for remote sensing of canopy chlorophyll content: A comparative analysis of hyperspectral data for different types of vegetation," *Plant Cell Environ.*, vol. 39, no. 12, pp. 2609–2623, Dec. 2016.
[7] X. Zou, R. Hernández-Clemente, P. Tammeorg, C. L. Torres, F. L. Stoddard, P. Mäkelä, P. Pellikka, and M. Mõttus, "Retrieval of leaf chlorophyll content in field crops using narrow-band indices: Effects of leaf area index and leaf mean tilt angle," *Int. J. Remote Sens.*, vol. 36, no. 24, pp. 6031–6055, Dec. 2015.
[8] K. Wang, W. Li, L. Deng, Q. Lyu, Y. Zheng, S. Yi, R. Xie, Y. Ma, and S. He, "Rapid detection of chlorophyll content and distribution in citrus orchards based on low-altitude remote sensing and bio-sensors," *Int. J. Agricult. Biol. Eng.*, vol. 11, no. 2, pp. 164–169, Mar. 2018.
[9] H. Croft, J. M. Chen, Y. Zhang, and A. Simic, "Modelling leaf chlorophyll content in broadleaf and needle leaf canopies from ground, CASI, Landsat TM 5 and MERIS reflectance data," *Remote Sens. Environ.*, vol. 133, pp. 128–140, Jun. 2013.
[10] R. Main, M. A. Cho, R. Mathieu, M. M. O'Kennedy, A. Ramoelo, and S. Koch, "An investigation into robust spectral indices for leaf chlorophyll estimation," *Isprs J. Photogram. Remote Sens.*, vol. 66, no. 6, pp. 751–761, Nov. 2011.
[11] J. G. P. W. Clevers and L. Kooistra, "Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 5, no. 2, pp. 574–583, Apr. 2012.
[12] E. R. Hunt, Jr., P. C. Doraiswamy, J. E. McMurtrey, C. S. T. Daughtry, E. M. Perry, and B. Akhmedov, "A visible band index for remote sensing leaf chlorophyll content at the canopy scale," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 21, pp. 103–112, Apr. 2013.
[13] Z. Malenovský, L. Homolová, R. Zurita-Milla, P. Lukeš, J.-P. Gastellu-Etchegorry, and M. E. Schaepman, "Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer," *Remote Sens. Environ.*, vol. 131, pp. 85–102, Apr. 2013.
[14] S. Jay, N. Gorretta, J. Morel, F. Maupas, R. Bendoula, G. Rabatel, D. Dutartre, A. Comar, and F. Baret, "Estimating leaf chlorophyll content in sugar beet canopies using millimeter- to centimeter-scale reflectance imagery," *Remote Sens. Environ.*, vol. 198, pp. 173–186, Sep. 2017.
[15] S. Cogliati, M. Rossini, T. Julitta, M. Meroni, A. Schickling, A. Burkart, F. Pinto, U. Rascher, and R. Colombo, "Continuous and long-term measurements of reflectance and sun-induced chlorophyll fluorescence by using novel automated field spectroscopy systems," *Remote Sens. Environ.*, vol. 164, pp. 270–281, Jul. 2015.
[16] B. Liu, W. Shen, Y.-M. Yue, R. Li, Q. Tong, and B. Zhang, "Combining spatial and spectral information to estimate chlorophyll contents of crop leaves with a field imaging spectroscopy system," *Precis. Agricult.*, vol. 18, no. 14, pp. 491–506, Aug. 2017.
[17] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[18] A. Sedaghat, M. Mokhtarzade, and H. Ebadi, "Uniform robust scale-invariant feature matching for optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4516–4527, Nov. 2011.
[19] S. Jacquemoud and F. Baret, "PROSPECT: A model of leaf optical properties spectra," *Remote Sens. Environ.*, vol. 34, no. 2, pp. 75–91, Nov. 1990.
[20] W. Verhoef, "Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model," *Remote Sens. Environ.*, vol. 16, no. 12, pp. 125–141, Oct. 1984.
[21] S. Jacquemoud, W. Verhoef, B. Baret, C. Bacour, P. J. Zarco-Tejada, G. P. Asner, C. François, and S. L. Ustin, "PROSPECT + SAIL models: A review of use for vegetation characterization," *Remote Sens. Environ.*, vol. 113, pp. S56–S66, Sep. 2009.
[22] A. Kuusk, "The angular distribution of reflectance and vegetation indices in barley and clover canopies," *Remote Sens. Environ.*, vol. 37, no. 2, pp. 143–151, Aug. 1991.
[23] J.-B. Feret, C. François, G. P. Asner, A. A. Gitelson, R. E. Martin, L. P. R. Bidel, S. L. Ustin, G. l. Maire, "PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments," *Remote Sens. Environ.*, vol. 112, pp. 3030–3043, Jun. 2008.
[24] Y. Dian, Y. Le, S. Fang, Y. Xu, C. Yao, and G. Liu, "Influence of spectral bandwidth and position on chlorophyll content retrieval at leaf and canopy levels," *J. Indian Soc. Remote Sens.*, vol. 44, pp. 583–593, Aug. 2016.
[25] R. D. Tobias, "An introduction to partial least squares regression," in *Proc. 20th Annu. SAS Users Group Int. Conf.*, Apr. 1995, pp. 1250–1257.
[26] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemometrics Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001.
[27] K. Yu, F. Li, M. L. Gnyp, Y. Miao, G. Bareth, and X. Chen, "Remotely detecting canopy nitrogen concentration and uptake of paddy rice in the Northeast China Plain," *ISPRS J. Photogram. Remote Sens.*, vol. 78, pp. 102–115, Apr. 2013.
[28] J. Farifteh, F. Van der Meer, C. Atzberger, and E. J. M. Carranza, "Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN)," *Remote Sens. Environ.*, vol. 110, no. 1, pp. 59–78, 2007.
[29] S. Song, W. Gong, B. Zhu, and X. Huang, "Wavelength selection and spectral discrimination for paddy rice, with laboratory measurements of hyperspectral leaf reflectance," *ISPRS J. Photogram. Remote Sens.*, vol. 66, pp. 672–682, Sep. 2011.
[30] P. M. Hansen and J. K. Schjoerring, "Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression," *Remote Sens. Environ.*, vol. 86, no. 4, pp. 542–553, 2003.
[31] A. Wagner, S. Hilgert, T. Kattenborn, and S. Fuchs, "Proximal VIS-NIR spectrometry to retrieve substance concentrations in surface waters using partial least squares modelling," *Water Sci. Technol.-Water Supply*, vol. 19, no. 4, pp. 1204–1211, Jun. 2019.
[32] J. Liu, J. Han, X. Chen, L. Shi, and L. Zhang, "Nondestructive detection of rape leaf chlorophyll level based on Vis-NIR spectroscopy," *Spectrochim. Acta A, Mol. Biomol. Spectrosc.*, vol. 222, May 2019, Art. no. 117202.
[33] R. A. Viscarra Rossel, "ParLeS: Software for chemometric analysis of spectroscopic data," *Chemometrics Intell. Lab. Syst.*, vol. 90, pp. 72–83, Jan. 2008.
[34] I. Jolliffe, *Principal Component Analysis*. vol. 98. 2nd ed, 2005. [Online]. Available: http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=017716/(100)
[35] S. Wold, H. Martens, and H. Wold, "The multivariate calibration problem in chemistry solved by the PLS method," in *Matrix Pencils* (Lecture Notes in Mathematics), vol. 973, 1983, pp. 286–293.
[36] T. Fearn, "Assessing calibrations: SEP, RPD, RER and R^2," *NIR News*, vol. 13, pp. 12–13, Dec. 2017.
[37] R. F. Kokaly, "Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration," *Remote Sens. Environ.*, vol. 75, pp. 153–161, Feb. 2001.
[38] K.-Q. Yu, Y.-R. Zhao, X.-L. Li, Y.-N. Shao, F. Liu, and Y. He, "Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant," *Plos One*, vol. 9, Dec. 2014, Art. no. e116205.
[39] H. Schulz, M. Baranska, and R. Baranski, "Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis," *Biopolymers*, vol. 77, no. 4, pp. 21–212, Mar. 2005.
[40] J. Moran, A. K. Mitchell, G. Goodmanson, and K. A. Stockburger, "Differentiation among effects of nitrogen fertilization treatments on conifer seedlings by foliar reflectance: A comparison of methods," *Tree Physiol.*, vol. 20, no. 16, pp. 1113–1120, Oct. 2000.
[41] M. A. Cho and A. K. Skidmore, "A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method," *Remote Sens. Environ.*, vol. 101, no. 2, pp. 181–193, Mar. 2006.
[42] D. Samuel, B. Park, M. Sohn, and L. Wicker, "Visible-near-infrared spectroscopy to predict water-holding capacity in normal and pale broiler breast meat," *Poultry Sci.*, vol. 90, pp. 914–921, Apr. 2011.
[43] A. H. Sivertsen, K. Heia, K. Hindberg, and F. Godtliebsen, "Automatic nematode detection in cod fillets (Gadus morhua L.) by hyperspectral imaging," *J. Food Eng.*, vol. 111, pp. 675–681, Aug. 2012.
[44] H.-J. He, D. Wu, and D.-W. Sun, "Non-destructive and rapid analysis of moisture distribution in farmed atlantic salmon (Salmo salar) fillets using visible and near-infrared hyperspectral imaging," *Innov. Food Sci. Emerg. Technol.*, vol. 18, pp. 237–245, Apr. 2013.

[45] D. Wu, D.-W. Sun, and Y. He, "Application of long-wave near infrared hyperspectral imaging for measurement of color distribution in salmon fillet," *Innov. Food Sci. Emerg. Technol.*, vol. 16, pp. 361–372, Oct. 2012.
[46] J. M. de Paz, F. Visconti, M. Chiaravalle, and A. Quiñones, "Determination of persimmon leaf chloride contents using near-infrared spectroscopy (NIRS)," *Anal. Bioanal. Chem.*, vol. 408, no. 13, pp. 3537–3545, May 2016.
[47] X. Yao, Y. Huang, G. Shang, C. Zhou, T. Cheng, Y. Tian, W. Cao, and Y. Zhu, "Evaluation of six algorithms to monitor wheat leaf nitrogen concentration," *Remote Sens.*, vol. 7, no. 11, pp. 14939–14966, 2015.
[48] H. Croft, J. M. Chen, Y. Zhang, A. Simic, T. L. Noland, N. Nesbitt, and J. Arabian, "Evaluating leaf chlorophyll content prediction from multispectral remote sensing data within a physically-based modelling framework," *ISPRS J. Photogram. Remote Sens.*, vol. 102, pp. 85–95, Apr. 2015.
[49] R. Darvishzadeh, C. Atzberger, A. Skidmore, and M. Schlerf, "Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models," *ISPRS J. Photogram. Remote Sens.*, vol. 66, no. 6, pp. 894–906, 2011.
[50] M. Weiss, B. Frederic, R. Myneni, A. Pragnère, and Y. Knyazikhin, "Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data," *Agronomie*, vol. 20, no. 1, Jan./Feb. 2000, Art. no. 2000105, doi: 10.1051/agro:2000105.

ZHONGZHENG PENG was born in Ganzhou, Jiangxi, China, in 1992. He received the B.S. degree from the School of Physics and Electronic Information, Gannan Normal University, Ganzhou, in 2015, where he is currently pursuing the M.S. degree in agricultural engineering. His current research interests include intelligent information processing, machine learning, and the application of remote sensing in agricultural engineering.

LIXIN GUAN was born in Ganzhou, Jiangxi, China, in 1974. He received the B.S. degree from the School of Information engineering, Nanchang University, Nanchang, China, in 1998, the M.S. degree in communication and information from Tianjin University, Tianjin, China, in 2006, and the Ph.D. degree in information and communication engineering from the ATR Key Laboratory of National Defense Technology, Shenzhen University, Shenzhen, China, in 2018.

Since 2009, he has been an Associate Professor with the Intelligent Control Engineering and Technology Research Center, Gannan Normal University, Ganzhou. His current research interests include intelligent information processing, machine learning, and the application of remote sensing in agricultural engineering.

YUBO LIAO was born in Ganzhou, Jiangxi, China, in 1980. He received the B.S. degree from the School of Physics and Electronic Information, Gannan Normal University, Ganzhou, in 2004, the M.S. degree in optics from South China Normal University, Guangzhou, China, in 2007, and the Ph.D. degree in optical engineering from Shenzhen University, Shenzhen, China, in 2018.

He is currently a Lecturer with the Intelligent Control Engineering and Technology Research Center, Gannan Normal University. His current research interests include intelligent information processing, machine learning, and the application of remote sensing in agricultural engineering.

SUYUN LIAN was born in Ganzhou, Jiangxi, China, in 1996. She received the B.S. degree from the School of Physics and Electronic Information, Gannan Normal University, Ganzhou, in 2019, where she is currently pursuing the M.S. degree in agricultural engineering. Her current research interests include intelligent information processing, machine learning, and the application of remote sensing in agricultural engineering.