Estimating the leaf water content of *Coffea arabica* L. based on hyperspectral reflectance and dataset construction

Xiaogang Liu[1,2], Kailun Peng[1], Shaomin Chen[1,2*], Yunfei Tuo[3], Shuai Zhang[1], Shuai Tan[1,2], Fangchuan Lou[1], Jinxue Li[4], Zhenjia Chen[5]

(1. *Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China*;
2. *Yunnan Provincial Field Scientific Observation and Research Station on Water-Soil-Crop System in Seasonal Arid Region, Kunming University of Science and Technology, Kunming 650500, China*;
3. *College of Soil and Water Conservation, Southwest Forestry University, Kunming 650224, China*;
4. *College of Biotechnology and Engineering, West Yunnan University, Lincang 677000, Yunnan, China*;
5. *China Coffee Engineering Research Center, Dehong Normal University, Dehong 678400, Yunnan, China*)

Abstract: Currently, irrigation decisions in coffee cultivation primarily rely on empirical knowledge, resulting in inefficient practices. Combining real-time leaf water content (LWC) data can help improve the accuracy of the irrigation planning. Spectral remote sensing is a fast, reliable, and non-invasive method to detect vegetation moisture content. In this study, a model to estimate the LWC of *Coffea arabica* L. was built using hyperspectral reflectance of the canopy under various irrigation levels. For this purpose, common spectral indices, two-band spectral indices [ratio spectral index (RSI); difference spectral index (DSI); and normalized difference spectral index (NDSI)], and three-band spectral indices were constructed. Feature bands were extracted using the successive projections algorithm (SPA). Optimal spectral indices were extracted using the correlation coefficient method, and the feature wavebands and spectral indices were combined into five datasets. These datasets were split into modeling and validation datasets by sample set partitioning based on the joint *x-y* distance (SPXY) algorithm. A linear model [partial least squares regression (PLSR)] and three non-linear models [support vector machine (SVM); extreme learning machine (ELM); back propagation artificial neural network (BPANN)] were built to estimate LWC of *Coffea arabica* L. The results indicated that the non-linear models surpassed the linear model. The accuracy was the highest when the modeling was performed using the dataset combination 5. Among various modeling methods, the predictive performance of ELM was the best (modeling dataset: R^2=0.745, RMSE=2.241%, RRMSE=3.482%; validation dataset: R^2=0.721, RMSE=2.142%, RRMSE=3.364%). ELM outperformed PLSR, SVM, and BPANN in LWC retrieval. The obtained results indicated that the dataset built by the combined use of different methods was superior to those from a single data source in accuracy. This study provides a scientific basis for the quantitative diagnosis of coffee tree water status, with significant implications for optimizing field irrigation management.

Keywords: spectral index, feature band, dataset, *Coffea arabica* L., leaf water content, machine learning
DOI: 10.25165/j.ijabe.20251805.9067

Citation: Liu X G, Peng K L, Chen S M, Tuo Y F, Zhang S, Tan S, et al. Estimating the leaf water content of *Coffea arabica* L. based on hyperspectral reflectance and dataset construction. Int J Agric & Biol Eng, 2025; 18(5): 287–297.

1 Introduction

Coffee is an important cash crop in Yunnan Province. Coffee, tea, and cocoa are the most abundant beverages worldwide, and coffee is the most popular beverage today[1]. Leaves are the vital constituent of the plant canopy and also the locale where major biochemical processes take place. Leaf water content (LWC) is an important indicator reflecting the overall water status of crops and can be used to guide crop irrigation[2,3]. The leaf sampling and drying method is a time- and labor-consuming method to detect LWC. Besides, this method is destructive and unable to determine the LWC of coffee leaves in a fast, non-invasive, and real-time manner. Leaf sampling and drying method is not suitable for dynamic monitoring of plant LWC[4]. There is an urgent need for fast and accurate LWC estimation in coffee leaves to inform irrigation regimens and water resource conservation[5].

Hyperspectral remote sensing offers the benefits of fastness, economic efficiency, and non-invasiveness in target detection and measures the spectral characteristics of plant canopy at wavelengths ranging from 350 to 2500 nm. Spectra in the near-infrared region (780-2500 nm) are more sensitive to water and can be used to estimate the canopy water content. Spectra in the shortwave infrared

Received date: 2024-05-10 **Accepted date:** 2025-07-22
Biographies: Xiaogang Liu, PhD, Professor, research interest: theory and new technology of agricultural water saving and optimization of water and soil resources, Email: liuxiaogangjy@126.com; **Kailun Peng**, MS, research interest: agricultural remote sensing, Email: pengkailun22@163.com; **Yunfei Tuo**, PhD, Associate Professor, research interest: agricultural water saving and optimization of water and soil resources, Email: tyunfei@163.com; **Shuai Zhang**, MS, research interest: agricultural remote sensing, Email: shuaizhang1012@163.com; **Shuai Tan**, PhD, Associate Professor, research interest: efficient utilization of water and fertilizer in crops, Email: tans90@163.com; **Fangchuan Lou**, MS, research interest: agricultural remote sensing, Email: fangchuanlou@163.com; **Jinxue Li**, PhD, Researcher, research interest: cultivation and regulation of fruit trees, Email: ljxue810@163.com; **Zhenjia Chen**, PhD, Professor, research interest: efficient utilization of water and fertilizer in crops, Email: 905088983@qq.com.

***Corresponding author: Shaomin Chen**, PhD, Associate Professor, research interest: efficient utilization of water and fertilizer in crops, agricultural remote sensing, Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming, China. Tel: +86-871-65920379, Email: chenshaomin2021@kust.edu.cn.

region (1300-2500 nm) are insensitive to noises posed by the internal structure of leaves, but there are also bands that are sensitive to moisture. For this reason, this spectral region is suitable for monitoring water content changes of plants[6].

According to previous studies, LWC can be predicted from spectral indices by leveraging the correlation between LWC and hyperspectral reflectance. It is generally believed that wavebands sensitive to crop water contents include 820, 950-970, 1200, 1450, 1600, 1940, and 2500 nm[7]. The water index (WI, R_{900}/R_{950}) constructed from the reflectance at the wavelengths of 900 nm and 950 nm can reflect the water content of plants[8]. LWC of rice, peanuts, soybeans, and wheat is capable of being well prognosticated using the ratio between WI and normalized difference vegetation index (NDVI)[9]. However, due to the influences of crop type, plant age, environmental conditions, and growth factors, the correlations between different spectral indices and plant traits seem to vary. As a result, the wavelengths sensitive to the same parameter to be retrieved also vary for different plants[10-12]. The spectral indices need to be constructed utilizing reflectance at two or even three wavelengths to estimate LWC of coffee leaves more reliably.

Extracting characteristic wavebands can help mitigate data redundancy, reduce data dimensions, and improve the modeling efficiency. In previous studies, competitive adaptive reweighted sampling (CARS) and genetic algorithm (GA), combined or not combined with successive projections algorithm (SPA), have been extensively applied to feature band selection[13-15]. Nguyen et al.[16] employed GA for characteristic band selection and used extreme gradient boosting (XGBoost) to predict soil water content with high accuracy. However, predictive studies on LWC of coffee trees based on hyperspectral imaging have been scarce. Building a dataset of characteristic bands using the feature extraction method is of high importance for predicting LWC of coffee trees.

By leveraging the correlation between LWC and spectral reflectance, researchers may choose from a variety of modeling methods to reveal the linear and non-linear features of such correlation. Statistical and machine learning techniques commonly used for the above purpose include partial least squares regression (PLSR), support vector machine (SVM), and artificial neural network (ANN)[17,18]. Independent variable construction is a critical step in the modeling process. Feature bands and vegetation indices are usually chosen as independent variables, as they carry information on reflectance of vegetation in different wavebands, which in turn is related to water content. To verify the reliability and universality of the models, researchers usually need to perform water content retrieval on different datasets. It is considered highly necessary in predicting LWC of coffee trees to enact the combined use of statistical and machine learning techniques and on different datasets.

The present study was conducted to solve the above problems and mainly consisted of the following contents: 1) Datasets from multiple data sources were built, consisting of common spectral indices, spectral indices utilizing reflectance at two and three wavelengths constructed in this study, and feature bands; 2) The datasets were combined in different ways, and the combination with a higher accuracy was chosen for modeling; 3) Models were constructed using different techniques, including PLSR, SVM, ELM, and back-propagation artificial neural network (BPANN) and run on different datasets. The model performance was evaluated. This study may guide the dataset combination and the optimal model determination for LWC retrieval of coffee trees.

2 Materials and method

2.1 Experimental design

Experiments were conducted in the greenhouse of the College of Modern Agricultural Engineering, Kunming University of Science and Technology in Yunnan, China, from April 2022 to June 2023. The test base (102°51′49.5″E, 24°84′40.6″N) has an average altitude of 1778.9 m and belongs to the subtropical plateau monsoon climate. The annual average temperature is 21°C in the greenhouse where the experiments were conducted, with a relative humidity of 45% to 70%. The typical red soil in the test base has a pH value of 6.5-7.5. The organic matter content is 15.05 g/kg; the total nitrogen, phosphorus, and potassium contents are 0.87 g/kg, 0.68 g/kg, and 13.9 g/kg, respectively; the initial contents of nitrate nitrogen, rapidly available phosphorus, and rapidly available potassium were 57.48 mg/kg, 12.61 mg/kg, and 85.53 mg/kg, respectively. The planting area of coffee trees was 120 m^2 (12 m×10 m), and four-year-old Arabica coffee trees with consistent growth status were chosen as research subjects (canopy diameter about 0.8 m and plant height about 1.0 m). The trees were spaced apart by 0.8 m and the rows by 1.2 m. Each tree occupied an area of 0.96 m^2. Four irrigation levels were set up by reference to previous studies[19]. That was, 1.0 L/tree, 1.5 L/tree, 2.0 L/tree, and 2.5 L/tree, respectively, and the irrigation cycle was approximately once every seven days. A water-soluble fertilizer ($N\text{-}P_2O_5\text{-}K_2O$, 20%-20%-20%) was applied at 20 g/tree through a fertigation system. Fertilization was conducted five times during the experimental period (1 July 2022, 28 September 2022, 31 January 2023, 5 April 2023, 20 May 2023). As shown in Figure 1, the soil water content monitored during the experiment indicated that the treatment had significant differences. Five replicates were set up for each irrigation level, and a total of 20 trees were included for the experiments. The choice of fertilization regimen at each growth stage of coffee trees was based on existing literature[20].

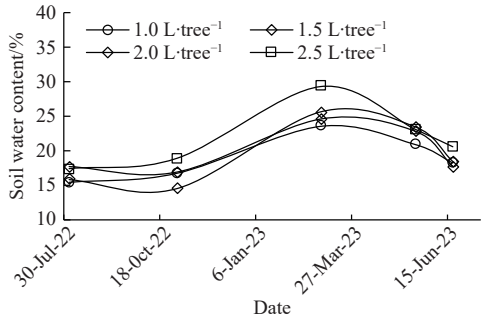


Figure 1 The average soil water content in 0-40 cm soil layer during the experimental period

2.2 Measurement indicators and methods

2.2.1 Spectral data acquisition from the canopy layer of coffee trees

The canopy spectra were acquired using the SR-2500 portable ground-object spectroradiometer (Spectral Evolution, Inc. 1 Canal St. Unit B-1 Lawrence, MA 01840 USA). The spectroradiometer spans a wavelength range from 350 to 2500 nm and has 2151 channels. The spectral resolution in the wavelength range from 350 to 1000 nm is 3.5 nm, with a sampling interval of 1.4 nm; the spectral resolution is 22 nm in the wavelength scope from 1000 to 2500 nm, with a sampling interval of 6 nm. The spectroradiometer's data output interval is 1 nm. A 1.5 m long-length optical fiber was used for spectral acquisition, the field of view being 8°. Canopy

spectra were measured at 10:00-14:00 local time on sunny and cloudless days to ensure data integrity and an appropriate solar elevation angle for the acquisition. Twenty trees were measured each time. The spectroradiometer was preheated for 40 min before the measurement to ensure stability. For spectral data acquisition, the fiber optic probe was directed straight down at the target from a height of 1 m above the canopy. The range covered by the field of view did not exceed the canopy. Ten spectral curves were acquired for each tree and averaged to obtain the spectral reflectance of the canopy. Before each measurement, the spectroradiometer was calibrated using the standard whiteboard. The time for spectral measurement and sampling number are listed in Table 1.

Table 1 Time points for spectral data acquisition

Spectral measurement and sampling time	Number of samples
August 3, 2022	20
November 1, 2022	20
March 1, 2023	20
May 19, 2023	20
June 20, 2023	20
Subtotal	100

2.2.2 LWC determination

With spectral reflectance of the coffee tree canopy acquired, LWC was determined simultaneously using the drying method, which has proved accurate in LWC measurement. From each tree, 20 mature leaves with consistent size and shape and free from pests, diseases, and mechanical damage were collected from the above and below parts and in the periphery of the canopy. The leaves were weighed using an electronic balance scale with a precision of 0.01, and the fresh leaf mass F_w was obtained. Then the leaves were placed into an oven at 105°C for 30 min. Finally, the leaves were dried at 80°C until reaching constant mass, and the dry leaf mass D_w was determined. LWC is calculated by Equation (1):

$$\mathrm{LWC}(\%) = \frac{F_w - D_w}{F_w} \times 100 \quad (1)$$

where, LWC is the leaf water content, %; F_w is the leaf fresh weight, g; and D_w is the leaf dry weight, g.

2.2.3 Leaf area index determination

The leaf area index (LAI) was measured using a plant canopy analyzer (Win Scanopy, Regent Instruments Inc., Canada) immediately after the acquisition of hyperspectral data. This device is composed of a high-definition digital camera, a hemispherical fisheye lens, and a self-balancing frame. During the measurement, the fisheye lens was placed 10 cm away from the tree trunk. After horizontal correction, two photos were taken on the east and west sides of the tree, respectively, making a total of four images. Subsequently, the images were processed through a dedicated analysis software to calculate the LAI of the corresponding images. The average value of the four images was taken as the LAI data measurement of the coffee tree.

2.3 Dataset construction

2.3.1 Spectral data preprocessing and dataset splitting

Considering the influences of atmospheric moisture and noises on spectral reflectance, spectral reflectance data in the wavelength ranges from 350 to 1349 nm, 1541 to 1799 nm, and 1951 to 2449 nm (totaling 1848 wavelengths) were chosen for further study. Disturbances to spectral measurement are usually caused by stray lights and baseline drift, which further interferes with spectral data analysis. In this study, the spectral data were first preprocessed by multiplicative scatter correction to eliminate spectral variabilities caused by different scattering levels, thereby enhancing the correlations between the spectrum and the data[21].

For the sake of representativeness and independence of dataset splitting, sample set partitioning based on the joint *x*-*y* distance (SPXY) algorithm[22] was employed to split the dataset into the modeling dataset and the validation dataset at a ratio of 2:1.

2.3.2 Common spectral indices

Spectral indices are commonly used parameters for predicting plant water status through spectral reflectance. Based on published studies, 18 spectral indices related to plant water status were chosen for the research, as listed in Table 2.

Table 2 Conventional spectral indices to estimate LWC

Index number	Moisture spectral index	Expression	References
1	Normalized Difference Moisture Index (NDWI)	$(R_{820}-R_{1240})/(R_{820}+R_{1240})$	[23]
2	Water Stress Index (MSI)	R_{1600}/R_{820}	[24]
3	Moisture Index (WI)	R_{900}/R_{970}	[25]
4	Moisture Band Index (WBI)	R_{950}/R_{900}	[26]
5	Normalized Difference Vegetation Index (NDVI)	$(R_{\mathrm{nir}}-R_{\mathrm{red}})/(R_{\mathrm{nir}}+R_{\mathrm{red}})$	[27]
6	Simple Ratio Water Index (SRWI)	R_{820}/R_{1200}	[28]
7	Normalized Difference Infrared Index (NDII)	$(R_{850}-R_{1650})/(R_{850}+R_{1650})$	[29]
8	Enhanced Vegetation Index (EVI)	$[2.5(R_{\mathrm{nir}}-R_{\mathrm{red}})]/(1+R_{\mathrm{nir}}+6R_{\mathrm{red}}-7.5R_{460})]$	[30]
9	Ratio Index (RI)	R_{1650}/R_{2220}	[31]
10	Hyperspectral Normalized Difference Vegetation Index (hNDVI)	$(R_{900}-R_{680})/(R_{900}+R_{680})$	[32]
11	Vegetation Dryness Index (VDI)	$(R_{970}-R_{900})/(R_{970}+R_{900})$	[25]
12	Normalized Difference Moisture Index (NDMI)	$(R_{1649}-R_{1722})/(R_{1649}+R_{1722})$	[33]
13	Normalized Heading Index (NHI)	$[(R_{1100}-R_{1200})/(R_{1100}+R_{1200})]/[(R_{850}-R_{670})/(R_{850}+R_{670})]$	[34]
14	Ratio Vegetation Index (SR)	R_{900}/R_{680}	[35]
15	Photochemical Reflectance Index (PRI)	$(R_{570}-R_{531})/(R_{570}+R_{531})$	[36]
16	Shortwave Infrared Water Stress Index (SIWSI)	$(R_{858}-R_{1640})/(R_{858}+R_{1640})$	[37]
17	Normalized Difference Water Index (NDWI-h)	$(R_{1070}-R_{1200})/(R_{1070}+R_{1200})$	[38]
18	Moisture Index/ Normalized Index (WI/NDVI)	$[R_{900}(R_{800}-R_{680})]/[R_{970}(R_{800}-R_{680})]$	[39]

Note: The average reflectance in the wavelength range of 670±10 nm was represented by R_{red} and that in the wavelength range of 850±10 nm by R_{nir}.

2.3.3 Construction of two-band spectral indices

Three forms of two-band spectral index were selected (Table 3). The reflectance was measured at 1 nm wavelength intervals in each of the three wavelength ranges, namely, 350-1349 nm, 1401-

1799 nm, and 1951-2449 nm. Two reflectance values were combined using the formula shown in Table 3. Then, a correlation analysis was executed among the spectral indices generated using this method and LWC.

Table 3 Spectral indices utilizing reflectance at two wavelengths

Spectral index	Abbreviation	Expression	Reference
Ratio spectral index	RSI	R_1/R_2	[33]
Normalized difference spectral index	NDSI	$(R_1-R_2)/(R_1+R_2)$	[33]
Difference spectral index	DSI	R_1-R_2	[33]

Note: R_1 and R_2 were reflectance values at any two wavelengths within the ranges of 350-1349 nm, 1401-1799 nm, and 1951-2449 nm.

2.3.4 Construction of three-band spectral indices

Three-band spectral indices were established based on those two-band spectral indices plus a third wavelength (R_3). The purpose was to improve the modeling accuracy and resistance to disturbance while reducing or eliminating spectral saturation that usually occurs in two-band spectral indices[40,41]. The two wavelengths corresponding to the optimal spectral index (R_1 and R_2) were chosen as the basis for constructing the three-band spectral indices. The latter was constructed using formulae 2 to 8, respectively. Variations of the correlation coefficient between the three-band spectral indices and LWC with the changing of R_3 were analyzed. The optimal R_3 was determined, along with the optimal form of three-band spectral index. Seven classic forms of three-band spectral indices[42] were calculated as follows:

$$(R_1-R_2)/(R_1+R_2-R_3) \quad (2)$$

$$(R_1-R_2-R_3)/(R_1+R_2+R_3) \quad (3)$$

$$(R_1-R_2+2R_3)/(R_1+R_2-2R_3) \quad (4)$$

$$R_1/(R_2+R_3) \quad (5)$$

$$(R_1+R_3)/R_2 \quad (6)$$

$$(R_1-R_3)/R_2 \quad (7)$$

$$(R_1-R_3)/(R_2-R_3) \quad (8)$$

where, R_3 is the reflectance in any of the three wavelength ranges, namely, 350-1349 nm, 1401-1799 nm, and 1951-2449 nm.

2.3.5 Screening for feature bands

SPA was employed for feature extraction from the preprocessed hyperspectral data of the canopy to identify representative wavelengths for modeling. This method was expected to promote the efficiency and precision of the LWC retrieval model.

SPA involves a vector projection analysis, which compares the magnitude of projection vectors between different wavelengths[14]. The wavelength with the largest projection vector was chosen as the candidate wavelength, and the characteristic wavelengths were finally identified based on the correction model.

Let the initial vector for iteration be $x_{k(0)}$, the variable to be extracted be N, and the spectral matrix have J columns. The algorithm consists of the following steps:

Use the hyperspectral reflectance dataset as the modeling set. Choose one column (*j-th* column) of the spectral matrix randomly from the modeling set. Assign the *j-th* column of the modeling set to variable x_j. Denote the value as $x_{k(0)}$.

Denote the set comprising of the positions of non-selected column vector as S:

$$S = \{j, 1 \le j \le J, j \notin \{k(0), k(1), \ldots, k(n-1)\}\} \quad (9)$$

Calculate the projection of x_j with respect to each of the remaining column vectors:

$$P_{xj} = x_j - (x_j^{\mathrm{T}} \cdot x_{k(n-1)}) \cdot x_{k(n-1)} \cdot (x_{k(n-1)}^{\mathrm{T}} \cdot x_{k(n-1)})^{-1}, (j \in s) \quad (10)$$

Extract the wavelength $k(n)$ with the largest projection vector:

$$k(n) = \arg(\max(\|p_{xj}\|)), \quad j \in s \quad (11)$$

Let $x_j = p_{xj}$, $j \in s$, and calculate the cumulative sum of n. If $n<N$, perform a cyclic calculation using Equation (9).

Finally, establish a multiple linear regression model for $k(0)$ and N in each cycle. Implement leave-one-out cross-validation (LOOCV) to screen for characteristic wavelengths. Calculate the root mean squared error of cross-validation (RMSECV) for the modeling set, and identify the candidate subsets. $k(0)$ and N corresponding to the minimum RMSECV would be the optimal values.

2.4 Model construction

Linear and non-linear models were constructed based on the dataset combination consisting of common spectral indices, two-band spectral indices, three-band spectral indices, and feature bands. Linear model was built using partial least squares regression (PLSR). Non-linear models were built using support vector machine (SVM), extreme learning machine (ELM), and back propagation artificial neural network (BPANN), respectively.

2.4.1 Construction of a linear model

PLSR is an integration of principal component analysis, multiple linear regression analysis, and least squares regression method[15]. Input variables for PLSR are compressed into several latent variables (LVs). Determining the number of LVs is an essential step in PLSR. Here, an appropriate number of LVs was chosen by cross-validation combined with the minimum RMSECV principle.

2.4.2 Construction of non-linear models

SVM is a learning system that utilizes a linear function hypothesis in a high-dimensional feature space[41]. This study combined cross-validation with grid search to find the optimal values of the penalty factor c and kernel parameter g in the SVM model. An SVM model was built for each c and g, and the RMSECV was calculated. The optimal c and g values were determined to construct a retrieval model based on SVM.

ELM is based on a single-hidden layer feedforward neural network and has the benefits of high learning speed and strong generalization performance[11]. ELM randomly generates connection weights and thresholds between the input and hidden layers. The activation function residing within neurons in the hidden layer is Sigmoid function by default. The initial number of neurons in the hidden layer was 5 for ELM training and increased stepwise to 50 at an increment of 1. Each model was run for 10 000 iterations. The optimal number of neurons was determined, and the optimal ELM parameters were obtained.

BPANN has the foundational algorithm of gradient descent, and gradient search is implemented with the purpose of minimizing average value squared error between the actual and forecasted outputs of the network[14]. In this study, BPANN had three layers, namely, input layer, hidden layer, and output layer, and the model parameters were optimized. The transfer, training, and learning

functions of BPANN were customization options. The Sigmoid function, a continuously differentiable non-linear function, was used as the activation function, with a maximum step size of 10 000, a learning rate of 0.01, a learning goal of 0.01 and learning rate and momentum coefficient of 0.01. The steps to determine the number of neurons in the hidden layer of BPANN were consistent with those in ELM.

2.5 Model evaluation indicators

The model's accuracy was assessed based on determination coefficient (R^2), root mean square error (RMSE), and relative root mean square error (RRMSE), as given below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{13}$$

$$\mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\bar{y}} \times 100\% \tag{14}$$

where, y_i is the measured value; $\hat{y}$ is the predicted value; $\bar{y}$ is the mean of the predicted value; n is the number of samples, i=1, 2, …, n. The higher the correlation coefficient and the lower the RMSE, the higher the predictive accuracy. The predictive accuracy was considered very high when RRMSE was below 10%; the accuracy was fairly high when RRMSE was above 10% and below 30%; the accuracy was poor when RRMSE was above 30%.

2.6 Analytical software

Spectral preprocessing was conducted using the Unscrambler X 10.4. Statistical analysis and plotting were carried out using Microsoft Excel, SPSS, and Origin 2022. Spectral matrix construction, correlation coefficient calculation, plotting of isolines of correlation coefficient, and retrieval model construction were realized in Matlab 2020b.

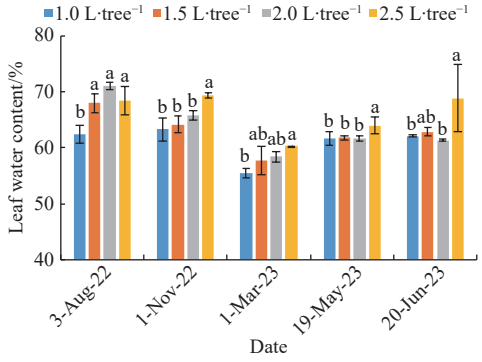
3 Results and analysis

3.1 Descriptive statistics of LWC and sample splitting

Multiple comparisons were done for data acquired at each time point (p<0.05). Generally speaking, the LWC increased significantly as the irrigation level increased. LWC of coffee trees varied significantly across the irrigation levels (Figure 2), indicating that the design of irrigation regimens was reasonable.

Figure 3 shows the spectral curves under each irrigation level. It can be seen that the hyperspectral reflectance was lower in the range from 400 to 720 nm, and the reflectance was not significantly correlated with the irrigation level, either positively or negatively. At around 720 nm, as the irrigation level increased, the spectral curve showed a redshift, indicating a better growth status; the determination of leaf area index could also confirm this phenomenon (Figure 4). In the near-infrared region (780-1350 nm), the reflectance of the hyperspectral curve varies in direct proportion to the amount of irrigation per plant. The differences in the hyperspectral curves under different irrigation levels in the near-infrared region (1450-2450 nm) might be attributed to the influences of atmospheric moisture and soil background.

A total of 100 groups of leaf samples were collected in the course of the experiments. The data were split into modeling and validation datasets, whose sample sizes were 67 and 33, respectively. Descriptive statistics of each LWC sample set can be found in Table 4. The statistical results on the modeling and validation datasets obtained by SPXY were comparable to those on the entire LWC sample set. The numerical range of the modeling dataset encompassed that of the validation dataset, thereby ensuring the representativeness of the sample sets.



Note: The different letters indicate significant differences among treatments at 0.05 level under the method of Duncan. Same as below.

Figure 2 LWC at different irrigation levels during the experimental period

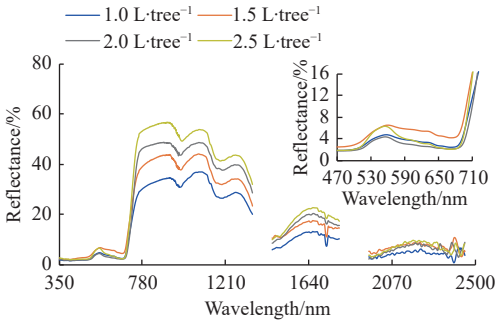


Figure 3 Hyperspectral reflectance curves of canopies at different irrigation levels

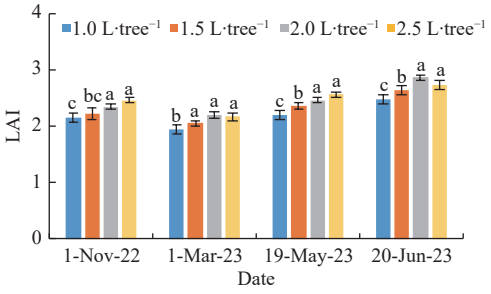


Figure 4 LAI at different irrigation levels during the experimental period

Table 4 Descriptive statistics of LWC of coffee trees

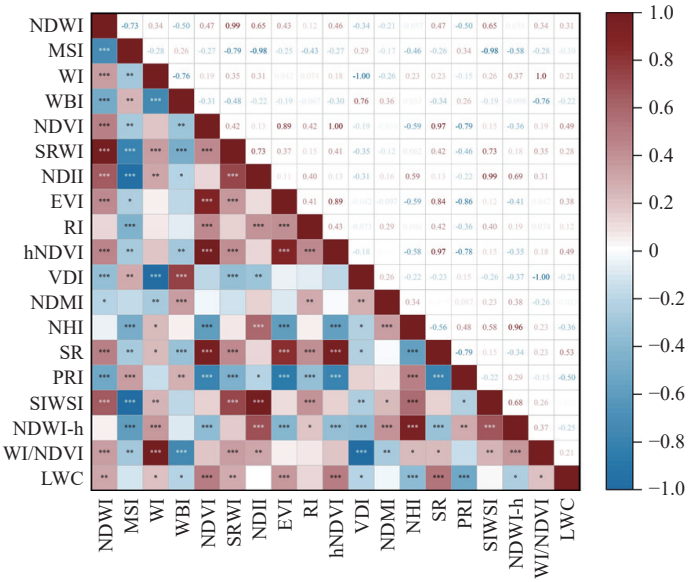
Dataset	Sample size	Maximum value/%	Minimum value/%	Mean value/%	Standard deviation/%
Complete dataset	100	73.93	54.23	64.13	4.19
Modeling dataset	67	73.93	54.23	64.64	4.56
Validation dataset	33	71.12	58.66	63.69	3.12

3.2 Dataset construction based on spectral indices and feature bands

3.2.1 Screening of common spectral indices

Eighteen common spectral indices were chosen and estimated, and the Pearson correlation coefficients between LWC of coffee

trees and the spectral index were computed. The correlation coefficient heatmap of the common spectral indices vs. LWC of coffee trees was shown in Figure 5. The spectral index most significantly correlated to LWC of coffee trees was selected.



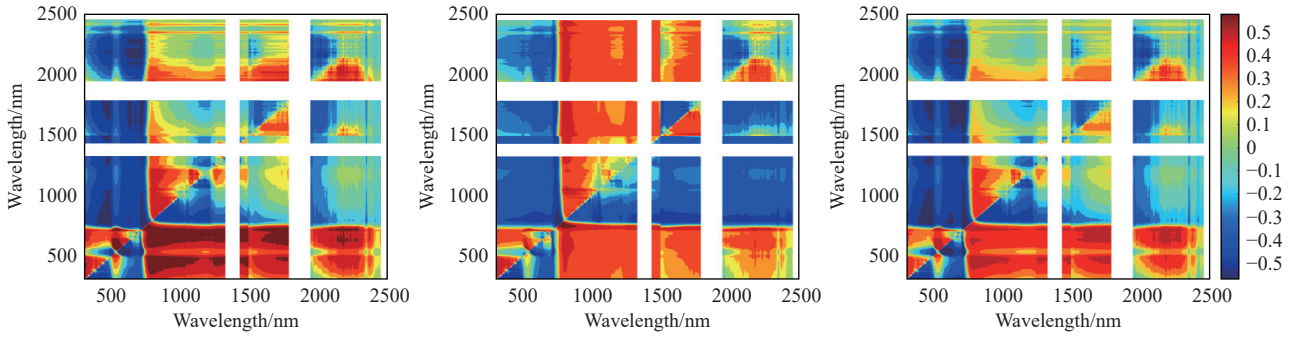
Note: *, **, *** indicate significant correlation at the p=0.05, p=0.01, and p=0.001 levels, respectively.

Figure 5 Correlation coefficient heatmap of the common spectral indices vs. LWC of coffee trees

As shown in Figure 5, except for MSI, NDII, RI, NDMI, and SIWSI, the correlation coefficients between other spectral indices and LWC all passed the significance test of 0.05 or 0.01. Among the 18 spectral indices, four spectral indices with the highest correlation coefficients were selected as part of the model input variables. These spectral indices were NDVI (0.492), hNDVI (0.486), SR (0.532), and PRI (–0.504).

3.2.2 Construction and screening of two-band spectral indices

To build two-band spectral indices, this study combined wavelengths pairwise and analyzed the correlation coefficients between LWC and the spectral indices utilizing reflectance at a pair of wavelengths (DSI, RSI, and NDSI). Isolines of the correlation coefficient were plotted (Figure 6). Based on Figure 6, for DSI, RSI, and NDSI, the top five spectral indices with the highest correlation coefficient were chosen as the optimal indices, respectively. As shown in Table 5, the three forms of spectral indices were sensitive to similar wavelength ranges, that is, 550-1200, 1300-1700, and 2050-2300 nm. The maximum correlation coefficient was observed with DSI (540 nm, 700 nm), the value being 0.58. It was indicated that the accuracy of retrieving LWC of coffee trees using the dataset of two-band spectral indices was higher and the correlation with LWC was stronger. The two-band spectral indices with the top five correlation coefficients with LWC were selected to participate in the construction of the dataset (Table 5).



a. RSI b. DSI c. NDSI

Figure 6 Correlation coefficient between the three forms of two-band spectral indices and LWC

Table 5 Waveband combinations of spectral indices with the top 5 correlation coefficients

Spectral index	Selected wavelength combinations/nm
DSI	(540, 700), (537, 578), (831, 834), (760, 486), (761, 467)
RSI	(758, 755), (749, 735), (763, 732), (744, 739), (758, 746)
NDSI	(831, 834), (540, 573), (556, 701), (723, 717), (2176, 702)

3.2.3 Construction and screening of three-band spectral indices

R_1 and R_2 were reflectance at 540 nm and 700 nm, the two wavelengths corresponding to the optimal DSI identified in Section 2.2.2, respectively. Reflectance was stepping successively in three waveband ranges, namely, 350-1349 nm, 1401-1799 nm, and 1951-2449 nm, as the value of R_3 in the three-band spectral index. Different forms of three-band spectral indices were calculated using Equations (2) to (8). Correlation coefficient curves (absolute value) between the spectral indices and LWC are listed in Figure 7. $(R_1+R_3)/R_2$, (R_{540}, R_{700}, R_{2176}) and $(R_1–R_3)/(R_2–R_3)$ (R_{540}, R_{700}, R_{466}) had the highest performance among all three-band spectral indices. Their correlation coefficients with LWC were 0.586 and 0.602, respectively, and both were higher compared with those for two-band spectral indices. Therefore, these two indices were included in the dataset for retrieving LWC of coffee trees.

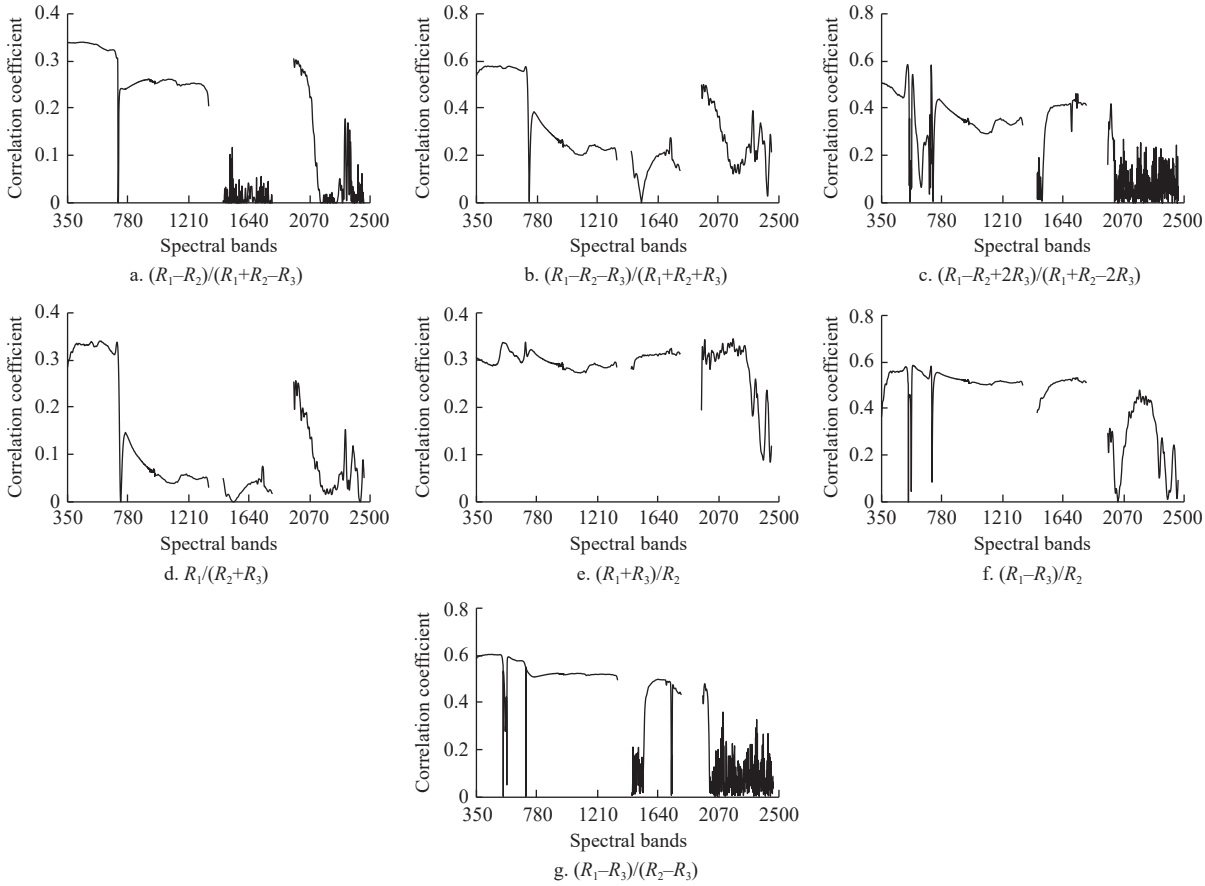


Figure 7 Correlation coefficient curves between the three-band spectral indices and LWC of coffee trees

3.2.4 Screening for wavelengths sensitive to LWC based on feature extraction method

Feature bands were identified by using SPA, and their positions are shown in Figure 8. The eight feature bands were 721 nm, 981 nm, 1252 nm, 1283 nm, 1733 nm, 2257 nm, 2320 nm, and 2414 nm. As shown in Figure 9, RMSECV first decreased and then increased as the number of variables increased. When the number of variables exceeded eight (indicated by the red block), the RMSECV did not reduce as the number of variables increased. SPA maximally compressed the spectral data, and the feature bands accounted for 0.43% of the full spectrum.

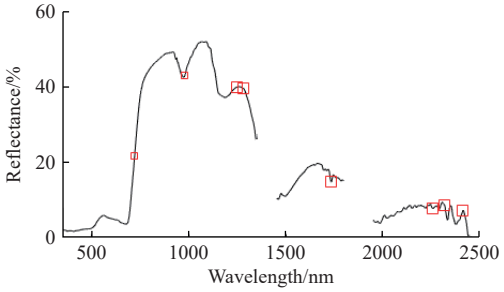


Figure 8 Positions of feature bands

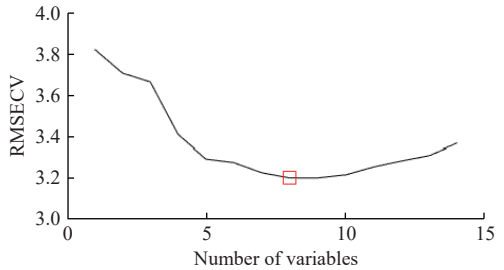


Figure 9 RMSECV under different numbers of variables

3.2.5 Results of dataset combinations

Different datasets were generated using different methods: common spectral indices, two-band spectral indices, three-band spectral indices, and feature bands identified by SPA. These datasets were combined in different ways to obtain different combinations, as listed in Table 6.

Table 6 Dataset combinations based on different methods

Index number	Different methods dataset combination	Number of variables
1	Common spectral indices, two-band spectral indices, three-band spectral indices	21
2	Common spectral indices, three-band spectral indices, feature bands	14
3	Two-band spectral indices, three-band spectral indices	17
4	Two-band spectral indices, three-band spectral indices, feature bands	25
5	Common spectral indices, two-band spectral indices, three-band spectral indices, feature bands	29

3.3 Construction and evaluation of models for LWC retrieval

The datasets in Table 6 were used as independent variables for modeling, and a linear model (PLSR) and non-linear models (SVM, ELM, and BPANN) were constructed for regression analysis. The model's accuracy and reliability were assessed. The calculation results using different dataset combinations are listed in Table 7. It is easy to see that the linear model was inadequate for estimating the LWC of coffee trees, with an R^2 ranging from 0.469 to 0.618. There was a certain gap in accuracy compared with non-linear models. Among the four non-linear models, ELM displayed a higher accuracy on both the modeling and validation datasets. As for the modeling results using different datasets, the accuracy was the highest when the prediction was conducted using the ELM

model on dataset combination 5 (a dataset combination consisting of the common spectral indices, two-band spectral indices, three-band spectral indices, and feature bands). The R^2, RMSE, and RRMSE of validation dataset using the ELM model on dataset combination 5 were 0.721, 2.142%, and 3.364%, respectively. There was a dramatic improvement of accuracy compared with the linear model PLSR and other non-linear models. The scatter plots of predicted and measured values using the ELM model on dataset combination 5 are shown in Figure 10. The slopes of the fitted lines were 0.745 and 0.721, respectively. Besides, the data points were distributed mainly below and above the 1:1 line in a uniform and compact manner for both the modeling and the validation datasets. Thus, the model had a high goodness-of-fit and hence a high predictive accuracy for LWC of coffee tree.

Table 7 Prediction results using different methods on different datasets for LWC retrieval

Dataset	Modeling methods	Modeling dataset			Validation dataset			Parameters
		R^2	RMSE/%	RRMSE/%	R^2	RMSE/%	RRMSE/%	
Combination 1	PLSR	0.544	3.094	4.805	0.555	2.276	3.578	LVs=14
	SVM	0.486	3.350	5.203	0.575	2.055	3.231	c=119.42 g=0.001
	ELM	0.659	2.679	4.160	0.624	2.365	3.718	NHLs=28
	BPANN	0.664	3.429	5.326	0.663	3.258	5.121	NHLs=24
Combination 2	PLSR	0.458	3.280	5.111	0.567	2.319	3.621	LVs=13
	SVM	0.448	3.338	5.202	0.728	1.862	2.907	c=0.6156 g=0.0825
	ELM	0.615	2.765	4.321	0.610	2.283	3.989	NHLs=22
	BPANN	0.650	2.651	4.131	0.637	2.659	4.152	NHLs=8
Combination 3	PLSR	0.470	3.199	4.972	0.469	2.661	4.179	LVs=8
	SVM	0.495	3.145	4.887	0.512	2.546	3.998	c=3.7321 g=0.0103
	ELM	0.674	2.509	3.898	0.602	2.889	4.537	NHLs=23
	BPANN	0.666	2.560	3.978	0.565	2.602	4.086	NHLs=6
Combination 4	PLSR	0.514	3.224	5.018	0.511	2.431	3.805	LVs=10
	SVM	0.578	3.054	4.753	0.630	2.161	3.382	c=3.0314 g=0.0385
	ELM	0.613	2.876	4.477	0.612	2.446	3.827	NHLs=11
	BPANN	0.685	2.894	4.505	0.623	2.954	4.622	NHLs=19
Combination 5	PLSR	0.630	2.697	4.190	0.618	2.764	4.342	LVs=20
	SVM	0.552	3.028	4.704	0.414	3.152	4.951	c=0.8706 g=0.0412
	ELM	**0.745**	**2.241**	**3.482**	**0.721**	**2.142**	**3.364**	**NHLs=26**
	BPANN	0.715	2.440	3.790	0.669	2.167	3.403	NHLs=5

Note: LVs represents the number of latent variables; c and g are the penalty factor and the kernel parameter of SVM, respectively; NHLs represents the number of hidden layer neurons.

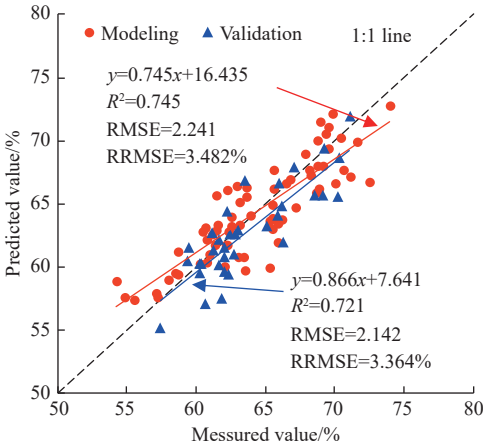


Figure 10 Relationship between the measured and predicted values of LWC

4 Discussion

Water is an essential component of the coffee tree canopy, involved in coffee photosynthesis and respiration, and has a momentous impact on the formation of eventual products. LWC of coffee trees was estimated from hyperspectral reflectance in this study. Reflectance of canopy in the visible region (VIS, 400-780 nm) and the near-infrared region (NIR, 780-1300 nm) was more greatly affected by the leaf structure and water content of canopy. In the VIS region, the reflectance of canopy decreased as the irrigation level of a single tree increased. The reason may be that the lack of irrigation induced drought stress in coffee trees, which further reduced the free water content of leaves but increased the bound water content of leaves. As a result, the leaf area and chlorophyll content decreased, accompanied by variations in the spectral reflectance of canopy[43]. In the NIR region, the reflectance of canopy increased as the irrigation level of a single tree increased. This is because as the irrigation level increased, the plant height, chlorophyll content, and net photosynthetic rate increased, leading to a significant increase in canopy reflectance. Under each irrigation level, the reflectance varied in the green region (G, 490-560 nm) and the red region (R, 620-680 nm). The fluctuation was even more significant in the crest of green light (at about 555 nm) and in the trough of red light (at about 672 nm), which can be attributed to the absorption characteristics of chlorophyll[44].

Existing studies have shown that a single spectral reflectance is inadequate for accurately estimating plant water status due to the strong reflection of radiant energy from fresh leaf surface and lamellar hair and the unique structures of leaf surface and leaf cuticles[45]. Constructing spectral indices is a better way to measure the spectral reflectance of plants, as spectral indices can reduce the scattering effect at a single wavelength and hence increase the prediction accuracy[46]. This study built new spectral indices by combining reflectance at two wavelengths. For RSI, NDSI, and DSI, the top five spectral indices with the highest correlation coefficient with LWC were selected to form the dataset of two-band spectral indices, respectively. However, spectral saturation may

occur in these spectral indices. By contrast, three-band spectral indices contained fewer noises[47] and had a stronger resistance to spectral saturation[48]. The optimal DSI (at 540 and 700 nm) was determined and served as the basis for establishing three-band spectral indices, which were $(R_1+R_3)/R_2$ (R_{540}, R_{700}, R_{2176}) and $(R_1-R_3)/(R_2-R_3)$ (R_{540}, R_{700}, R_{466}) according to this analysis. The latter had higher correlation coefficients with LWC than the former, therefore carrying fewer noises and displaying more robust predictive performance for LWC of coffee trees. To ensure the completeness of information contained in independent variables, this study combined the datasets of common spectral indices, two-band spectral indices, three-band spectral indices, and feature bands identified by SPA. While most studies only used a single source of data to predict LWC of plants[49,50], this study built a dataset from multiple sources, which involved the combined use of several methods. Compared with datasets built by using a single method, the method in this paper dramatically improved the model's predictive accuracy.

To estimate LWC based on spectral reflectance, this study built PLSR, SVM, ELM, and BPANN models, which were run on the five datasets, respectively. The modeling performance was highly differential on different datasets (Table 7), given the varying sizes of the datasets built by different methods. A larger dataset helped reduce overfitting and improved the model's generalization capacity. Both linear and non-linear models had a better performance when the datasets of common spectral indices, two-band spectral indices, three-band spectral indices, and feature bands were combined together. The predictive accuracy was compared across different datasets. Underfitting occurred when SVM was used for prediction on dataset combination 2. SVM performed poorly compared to other models on dataset combination 5, probably due to unreasonable optimal parameters found by grid search in SVM. Of all four models, the ELM model had the best predictive performance for LWC. During the modeling and validation processes, R^2 of the ELM model was 0.745 and 0.721, RMSE was 2.241% and 2.142%, and RRMSE was 3.482% and 3.364%, respectively. All of the performance evaluation indicators of the ELM model were better than those of PLSR, SVM, and BPANN. This is because the ELM model has the advantages of fewer training parameters, higher learning speed, and more robust generalization capacity than other models. Studies have shown that ELM is a reliable modeling method[51-53], and it is also found that ELM outperformed other models in LWC retrieval of coffee trees. The non-linear correlation between the spectral reflectance and LWC might cause the inferiority of the linear model PLSR. Non-linear models, such as SVM and BPANN, place higher requirements on sample size and feature dimensions, and hence there is a greater need for more adjustments during parameter selection. This may be one reason for the lower accuracy of SVM and BPANN compared with ELM[54,55].

This study focuses on LWC as the research target, without considering the influence of other canopy parameters on spectral information. Sensitivity analysis using the canopy radiative transfer model PROSAIL demonstrates that parameters including canopy leaf area index, leaf pigments (chlorophyll, carotenoids), and leaf dry matter content significantly affect the canopy reflectance[56]. Therefore, to weaken the impact of other canopy parameters, this research employed conventional spectral indices, two-band spectral indices, three-band spectral indices, and characteristic wavelengths to construct the dataset. The best-of-breed LWC prediction model was established based on dataset and machine learning. However, some limitations are still left unresolved. The present study was only conducted in Kunming, China and in only a single coffee variety. Whether the LWC retrieval model constructed in this study is equally applicable to LWC estimation of other coffee varieties in other regions remains to be further verified. Future research should continue to analyze and attempt to eliminate interference factors present in the spectral remote sensing process. Subsequently, the proposed method should be optimized for different coffee varieties across various regions, so as to obtain an LWC retrieval model for coffee trees with higher universality. Such models provide essential information for coffee growth, drought monitoring, and technical support for improving coffee yield.

5 Conclusions

1) The NIR region (780-1300 nm) was sensitive to LWC of coffee trees, and the reflectance in this region varied significantly under different irrigation levels. As the irrigation level per coffee tree increased, the spectral curve showed a redshift, indicating a better growth status.

2) Datasets generated using different methods were combined in different manners. It was found that the combination of optimized common spectral indices, two-band spectral indices, three-band spectral indices, and feature bands resulted in higher accuracy of LWC retrieval compared with datasets from a single data source, and the model's predictive performance was also better.

3) Linear and non-linear models were run on different dataset combinations, and the predictive accuracy was compared. The most satisfactory performance was observed with ELM, among all models (modeling dataset: R^2=0.745, RMSE=2.241%, RRMSE=3.482%; validation set: R^2=0.721, RMSE=2.142%, RRMSE=3.364%).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 52379040), Yunnan Fundamental Research Projects (Grant No. 202301AS070030 and 202101AU070039), Yunnan Major Science and Technology Special Plan (Grant No. 202302AE090024), Yunnan Science and Technology Talent and Platform Program (Grant No. 202305AM070006), and the Yunnan Intelligent Water-Fertilizer-Pesticide Integration Technology and Equipment Innovation Team (No. 202505AS350025).

[References]

[1] Zhang S, Liu X G, Li R M, Wang X L, Cheng J H, Yang Q L, et al. AHP-GIS and MaxEnt for delineation of potential distribution of Arabica coffee plantation under future climate in Yunnan, China. Ecological Indicators, 2021; 132: 108339.

[2] Zhang J J, Zhang W, Xiong S P, Song Z X, Tian W Z, Shi L, et al. Comparison of new hyperspectral index and machine learning models for prediction of winter wheat leaf water content. Plant Methods, 2021; 17: 34.

[3] Kong W P, Huang W J, Zhou X F, Mortimer H, Ma L L, Tang L L, et al. Estimating leaf water content at the leaf scale in soybean inoculated with arbuscular mycorrhizal fungi from in situ spectral measurements. Int J Agric & Biol Eng, 2019; 12(6): 149–155.

[4] Elsayed S, El-Hendawy S, Khadr M, Elsherbiny O, Al-Suhaibani N, Dewir Y H, et al. Integration of spectral reflectance indices and adaptive neuro-fuzzy inference system for assessing the growth performance and yield of potato under different drip irrigation regimes. Chemosensors, 2021; 9(3): 55.

[5] Magyar T, Fehér Z, Buday-Bódi E, Tamás J, NagyA. Modeling of soil moisture and water fluxes in a maize field for the optimization of irrigation. Computers and Electronics in Agriculture, 2023; 213: 108159.

[6] Romero A P, Alarcón A, Valbuena R I, Galeano C H. Physiological assessment of water stress in potato using spectral information. Frontiers in Plant Science, 2017; 8: 1608.
[7] Crusiol L G T, Nanni M R, Furlanetto R H, Sibaldelli R N R, Sun L, Gonçalves S L, et al. Assessing the sensitive spectral bands for soybean water status monitoring and soil moisture prediction using leaf-based hyperspectral reflectance. Agricultural Water Management, 2023; 277: 108089.
[8] Peñuelas J, Inoue Y. Reflectance indices indicative of changes in water and pigment contents of peanut and wheat leaves. Photosynthetica, 1999; 36(3): 355–360.
[9] Inoue Y, Morinaga S, Shibayama M. Non-destructive estimation of water status of intact crop leaves based on spectral reflectance measurements. Japanese Journal of Crop Science, 1993; 62(3): 462–469.
[10] El-Hendawy S E, Al-Suhaibani N A, Elsayed S, Hassan W M, Dewir Y H, Refay Y, et al. Potential of the existing and novel spectral reflectance indices for estimating the leaf water status and grain yield of spring wheat exposed to different irrigation rates. Agricultural Water Management, 2019; 217: 356–373.
[11] Chen S M, Hu T T, Luo L H, He Q, Zhang S W, Li M Y, et al. Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods. Infrared Physics & Technology, 2020; 111: 103542.
[12] Yang F F, Liu T, Wang Q Y, Du M Z, Yang T L, Liu D Z, et al. Rapid determination of leaf water content for monitoring waterlogging in winter wheat based on hyperspectral parameters. Journal of Integrative Agriculture, 2021; 20(10): 2613–2626.
[13] Tian H, Wang S, Xu H R, Ying Y B. Modeling method for SSC prediction in pomelo using Vis-NIRS with wavelength selection and latent variable updating. Int J Agric & Biol Eng, 2024; 17(1): 251–260.
[14] Zhang D Y, Xu L, Liang D, Xu C, Jin X L, Weng S X. Fast prediction of sugar content in dangshan pear (*Pyrus* spp) using hyperspectral imagery data. Food Analytical Methods, 2018; 11: 2336–2345.
[15] Tian P, Meng Q H, Wu Z F, Lin J J, Huang X, Zhu H, et al. Detection of mango soluble solid content using hyperspectral imaging technology. Infrared Physics & Technology, 2023; 129: 104576.
[16] Nguyen T T, Ngo H H, Guo W S, Chang S W, Nguyen D D, Nguyen C T, et al. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. Science of the Total Environment, 2022; 833: 155066.
[17] Chemura A, Mutanga O, Dube T. Remote sensing leaf water stress in coffee (*Coffea arabica*) using secondary effects of water absorption and random forests. Physics and Chemistry of the Earth, Parts A/B/C, 2017; 100: 317–324.
[18] Mirzaie M, Darvishzadeh R, Shakiba A, Matkan A A, Atzberger C, Skidmore A. Comparative analysis of different uni-and multi-variate methods for estimation of vegetation water content using hyper-spectral measurements. International Journal of Applied Earth Observation and Geoinformation, 2014; 26: 1–11.
[19] Li R M, Cheng J H, Liu X G, Wang Z H, Li H Y, Guo J J, et al. Optimizing drip fertigation at different periods to improve yield, volatile compounds and cup quality of Arabica coffee. Frontiers in Plant Science, 2023; 14: 1148616.
[20] Khemira H, Medebesh A, Mehrez K H, Hamadi N. Effect of fertilization on yield and quality of Arabica coffee grown on mountain terraces in southwestern Saudi Arabia. Scientia Horticulturae, 2023; 321: 112370.
[21] Gao L L, Zhu X C, Li C, Cheng L Z, Wang L, Zhao G X, et al. Improve the prediction accuracy of apple tree canopy nitrogen content through multiple scattering correction using spectroscopy. Agricultural Sciences, 2016; 7(10): 651–659.
[22] Galvao 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.
[23] Gao B C. NDWI–A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 1996; 58(3): 257–266.
[24] Ceccato P, Flasse S, Tarantola S, Jacquemoud S, Grégoire J M. Detecting vegetation leaf water content using reflectance in the optical domain. Remote Sensing of Environment, 2001; 77(1): 22–33.
[25] Peñuelas J, Filella I, Biel C, Serrano L, Savé R. The reflectance at the 950–970 nm region as an indicator of plant water status. International Journal of Remote Sensing, 1993; 14(10): 1887–1905.
[26] Riedell W E, Blackmer T M. Leaf reflectance spectra of cereal aphid–damaged wheat. Crop Science, 1999; 39(6): 1835–1840.
[27] Tucker, C J. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 1979; 8(2): 127–150.
[28] Zarco-Tejada P J, Rueda C A, Ustin S L. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. Remote Sensing of Environment, 2003; 85(1): 109–124.
[29] Sriwongsitanon N, Gao H K, Savenije H H G, Maekan E, Saengsawang S, Thianpopirug S. Comparing the normalized difference infrared index (NDII) with root zone storage in a lumped conceptual model. Hydrology and Earth System Sciences, 2016; 20(8): 3361–3377.
[30] Chen D Y, Huang J F, Jackson T J. Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near-and short-wave infrared bands. Remote Sensing of Environment, 2005; 98(2-3): 225–236.
[31] Elvidge C D, Lyon R J P. Estimation of the vegetation contribution to the 1.65/2.22 μm ratio in airborne thematic-mapper imagery of the Virginia Range, Nevada. International Journal of Remote Sensing, 1985; 6(1): 75–88.
[32] Rouse J W J, Haas R H, Schell J A, Deering D W. Monitoring vegetation systems in the great plains with ERTS. NASA Special Publication, 1974; 351(1): 309–317.
[33] Wang L L, Hunt J E R, Qu J J, Hao X J, Daughtry C S T. Towards estimation of canopy foliar biomass with spectral reflectance measurements. Remote Sensing of Environment, 2011; 115(3): 836–840.
[34] Pimstein A, Eitel J U H, Long D S, Mufradi I, Karnieli A, Bonfil D J. A spectral index to monitor the head-emergence of wheat in semi-arid conditions. Field Crops Research, 2009; 111(3): 218–225.
[35] Gitelson A A, Merzlyak M N. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. Journal of Plant Physiology, 1996; 148(3-4): 494–500.
[36] Gamon J A, Peñuelas J, Field C B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sensing of Environment, 1992; 41(1): 35–44.
[37] Fensholt R, Sandholt I. Derivation of a shortwave infrared water stress index from MODIS near-and shortwave infrared data in a semiarid environment. Remote Sensing of Environment, 2003; 87(1): 111–121.
[38] Ustin S L, Roberts D A, Gardner M, Dennison P. Evaluation of the potential of Hyperion data to estimate wildfire hazard in the Santa Ynez front range, Santa Barbara, California. International Geoscience and Remote Sensing Symposium, 2002; 2: 796–798.
[39] Peñuelas J, Pinol J, Ogaya R, Filella I. Estimation of plant water concentration by the reflectance water index WI (R900/R970). International Journal of Remote Sensing, 1997; 18(13): 2869–2875.
[40] Schneider P, Roberts D A, Kyriakidis P C. A VARI-based relative greenness from MODIS data for computing the Fire Potential Index. Remote Sensing of Environment, 2008; 112(3): 1151–1167.
[41] Huang W J, Lu J J, Ye H C, Kong W P, Mortimer A H, Shi Y. Quantitative identification of crop disease and nitrogen-water stress in winter wheat using continuous wavelet analysis. Int J Agric & Biol Eng, 2018; 11(2): 145–152.
[42] Yao X, Jia W Q, Si H Y, Guo Z Q, Tian Y C, Liu X J, et al. Exploring novel bands and key index for evaluating leaf equivalent water thickness in wheat using hyperspectra influenced by nitrogen. PLoS One, 2014; 9(6): e96352.
[43] Poblete T, Watt M S, Buddenbaum H, Zarco-Tejada P J. Chlorophyll content estimation in radiata pine using hyperspectral imagery: A comparison between empirical models, scaling-up algorithms, and radiative transfer inversions. Agricultural and Forest Meteorology, 2025; 362: 110402.
[44] Yang H B, Li F, Wang W, Yu K. Estimating above-ground biomass of potato using random forest and optimized hyperspectral indices. Remote Sensing, 2021; 13(12): 2339.
[45] Yang C, Tan Y L, Bruzzone L, Lu L J, Guan R C. Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images. Remote Sensing, 2017; 9(8): 782.
[46] Huang X, Lin D, Mao X M, Zhao Y. Multi-source data fusion for estimating maize leaf area index over the whole growing season under different mulching and irrigation conditions. Field Crops Research. 2023; 303: 109111. doi: 10.1016/j.fcr.2023.109111.

[47] Ma Y M, Duan B, Xu B C, Zou R, Shi Y C, Yu Q Y, et al. Rapeseed yield prediction based on fractional-order differentiation and UAV hyperspectral index optimization. Transactions of the CSAE, 2025; 41(10): 166–175. (in Chinese)

[48] Wang W, Yao X, Yao X F, Tian Y C, Liu X J, Ni J, et al. Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. Field Crops Research, 2012; 129: 90–98.

[49] Li D, Yu W G, Zheng H B, Guo C L, Yao X, Zhu Y, et al. Towards practical semi-empirical models for the estimation of leaf and canopy water contents from hyperspectral reflectance. Computers and Electronics in Agriculture, 2023; 214: 108309.

[50] Sun J J, Yang W D, Zhang M J, Feng M C, Xiao L J, Ding G W. Estimation of water content in corn leaves using hyperspectral data based on fractional order Savitzky-Golay derivation coupled with wavelength selection. Computers and Electronics in Agriculture, 2021; 182: 105989.

[51] Liu X G, Zhang S, Chen S M, Tuo Y F, Peng K L, Tan S, et al. Predicting leaf nitrogen content of coffee trees using the canopy hyperspectral reflectance feature bands, vegetation index and machine learning. International Journal of Remote Sensing, 2024; 45(22): 8471–8498.

[52] Guo Y, Guo J X, Shi Y, Li X L, Huang H, Liu Y C. Estimation of leaf moisture content in cantaloupe canopy based on SiPLS-CARS and GA-ELM. Spectroscopy and Spectral Analysis, 2022; 42(8): 2565–2571. (in Chinese)

[53] Tan C X, Zhang Z T, Xu C H, Ma Y, Yao Z H, Wei G F, et al. Soil water content inversion model in field maize root zone based on UAV multispectral remote sensing. Transactions of the CSAE, 2020; 36(10): 63–74. (in Chinese)

[54] Lv M Q, Song Y J, Weng H Y, Sun D W, Dong X Y, Fang H, et al. Effect of near infrared hyperspectral imaging scanning speed on prediction of water content in arabidopsis. Spectroscopy and Spectral Analysis, 2020; 40(11): 3508–3514. (in Chinese)

[55] Sun H, Liu N, Wu L, Zheng T, Li M Z, Wu J Z. Visualization of water content distribution in potato leaves based on hyperspectral image. Spectroscopy and Spectral Analysis, 2019; 39(3): 910–916. (in Chinese)

[56] Wan L, Zhang J F, Dong X Y, Du X Y, Zhu J P, Sun D W, et al. Unmanned aerial vehicle-based field phenotyping of crop biomass using growth traits retrieved from PROSAIL model. Computers and Electronics in Agriculture, 2021; 187: 106304.