Postprint of Upland Cotton Leaf Water Content Estimation Based on Feature Band Selection and Machine Learning

Authors: Cui Jintao

Date: 2023-12-06T00:00:00+00:00

Abstract

Timely and accurate monitoring of cotton leaf water content plays a crucial role in evaluating cotton growth status. To precisely estimate cotton leaf water content, this study utilized hyperspectral data and leaf water content data of cotton leaves at the field scale in the Weigan River-Kuqa River Delta Oasis of Xinjiang. Fractional-order differentiation was applied to preprocess the original spectra. Feature bands were selected through correlation coefficient analysis, Competitive Adaptive Reweighted Sampling (CARS), Successive Projections Algorithm (SPA), Genetic Algorithm (GA), Monte Carlo Uninformative Variables Elimination (MC-UVE), and coupled CARS-SPA methods. Leaf water content inversion models for both full bands and feature bands were established using Random Forest Regression (RFR) optimized by the Whale Optimization Algorithm (WOA), and validated using independent samples. The results demonstrated that: (1) Different feature band selection methods yielded varying numbers and positions of bands, with MC-UVE selecting 8 feature bands and CARS selecting 38 feature bands. The band positions from SPA, GA, and CARS-SPA methods were relatively consistent, primarily concentrated in the near-infrared region of 950–1050 nm. (2) The CARS-SPA-WOA-RFR model achieved the best inversion performance, with a coefficient of determination (R^2) = 0.93 and root mean square error (RMSE) = 0.032 for the predicted values. The developed model can provide a decision-making basis for accurately and rapidly monitoring cotton drought conditions and implementing precision irrigation.

Full Text

Abstract

Timely and accurate monitoring of cotton leaf water content (LWC) plays a critical role in evaluating cotton growth status. To precisely estimate cotton LWC, this study utilized hyperspectral data and corresponding leaf water measurements from cotton leaves in the Ugan River-Kuqa River Delta Oasis, Xinjiang. Fractional-order differentiation was applied to preprocess the original spectra. Feature bands were selected using correlation coefficient analysis, competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), genetic algorithm (GA), Monte Carlo uninformative variables elimination (MC-UVE), and a coupled CARS-SPA method. Leaf water content inversion models were constructed using whale optimization algorithm (WOA) improved random forest regression (RFR) for both full bands and selected feature bands, with independent samples employed for validation analysis. The results demonstrate that: (1) Different feature band selection methods yield varying numbers and positions of bands, with MC-UVE extracting the fewest bands (8) and CARS extracting the most (38). The band positions identified by SPA, GA, and CARS-SPA methods show considerable consistency, primarily concentrated in the near-infrared range of 950–1050 nm. (2) The CARS-SPA-WOA-RFR model achieved the best inversion performance, with a determination coefficient (R^2) of 0.93 and root mean square error (RMSE) of 0.032. This model can provide a decision-making basis for accurate and rapid monitoring of cotton drought conditions and precision irrigation.

Keywords: Spectroscopy; Leaf water content; Feature band selection; Machine learning

Introduction

As the world's largest cotton producer, China's primary production regions are concentrated in the Xinjiang cotton region, the Yellow River Basin cotton region, and the Yangtze River Basin cotton region. Among these, Xinjiang ranks first nationally in both cotton planting area and yield. Cotton plants produce energy through photosynthesis in their leaves, and leaf water content plays a vital role in monitoring physiological status, assessing crop growth, and reflecting soil moisture conditions. Therefore, rapid and effective acquisition of leaf water content is of great significance for cotton growth monitoring, yield assessment, and drought evaluation in arid and semi-arid regions.

Hyperspectral remote sensing technology, with its advantages of rapidity, accuracy, and non-destructiveness, overcomes the limitations of traditional laboratory measurements of leaf water content, which are time-consuming, labor-intensive, destructive, and unable to quickly obtain water data across large cotton fields. This technology has been widely applied in crop water inversion research, yielding numerous achievements. Previous studies have primarily employed competitive adaptive reweighted sampling (CARS), successive projections algorithm (SPA), genetic algorithm (GA), random forest (RF), and Monte Carlo uninformative variables elimination (MC-UVE) to select characteristic bands or construct vegetation indices, utilizing machine learning methods such as partial least squares regression (PLSR), support vector machine regression (SVR), and back propagation (BP) to establish inversion models. For instance, some researchers used CARS and SPA to select characteristic bands for tea leaf water content, employing multiple linear regression (MLR) to construct inversion models, with results showing the CARS-MLR method produced optimal model performance. Others combined CARS with SPA to screen characteristic bands for cotton leaf nitrogen content and above-ground biomass, achieving high inversion accuracy using RF models. Numerous scholars have conducted extensive research on quantitative monitoring of crop attributes using different screening methods, achieving significant results. Therefore, selecting appropriate feature band screening methods is crucial for crop water estimation using spectroscopy.

Compared with traditional methods such as PLSR and ridge regression, which suffer from low prediction accuracy and susceptibility to variable and sample size effects, machine learning algorithms effectively compensate for these shortcomings. For example, RF models have been widely applied in studies related to leaf water content, chlorophyll, and soil organic matter. Meanwhile, various optimization algorithms have developed rapidly, including particle swarm optimization (PSO), extreme learning machine (ELM), and whale optimization algorithm (WOA), which are extensively used to optimize traditional machine learning methods. WOA, proposed by Mirjalili in 2016, is a swarm intelligence algorithm that simulates the bubble-net foraging behavior of humpback whales through mathematical modeling of encircling prey, spiral position updating, and random search behavior. WOA offers advantages of simple principles, minimal parameter tuning requirements, high accuracy, rapid convergence, and low tendency to fall into local optima. Some studies have used WOA to optimize SVR parameters for PM2.5 prediction, achieving excellent modeling accuracy compared with standard SVR.

Current research on cotton leaf water content primarily employs linear models and random forest regression, with limited application of WOA in crop water estimation based on spectroscopy. In particular, in-depth research on inversion accuracy and fitting effects remains insufficient. Therefore, this study focuses on cotton leaf water content, using fractional-order differentiation for spectral preprocessing and five feature band selection methods to establish inversion models. WOA-improved RFR models were constructed for both full bands and selected feature bands, with independent samples used for validation to provide technical support for rapid and accurate monitoring of cotton leaf water content.

1.1 Study Area Overview

The Ugan River-Kuqa River Delta Oasis (hereafter referred to as the Ugan-Kuqa Oasis) is located in the Aksu region of Xinjiang, situated in the northern Tarim Basin and southern Tianshan Mountains, forming a typical alluvial fan plain at the mountain front [Figure 1: see original paper]. The region features a typical continental warm temperate arid climate, with an average annual temperature of 10.5–11.4°C, average annual precipitation of 46.5 mm in the plain area and 243.0 mm in mountainous areas, and average annual evaporation of

1374 mm. The climate is characterized by aridity, scarce precipitation, and frequent sandstorms. According to 2021 statistics, the cotton planting area in the Ugan-Kuqa Oasis accounts for 8.56% of Xinjiang's total and 38.2% of the Aksu region's total, while cotton yield accounts for 8.41% of Xinjiang's total and 40.34% of the Aksu region's total, making it one of Xinjiang's primary cotton production areas. The experimental field was located in Wuzhen Town, central-eastern Kuqa City, with geographic coordinates between 41°31 29.65 – 41°49 6.73 N and 83°00 13.7 –83°19 15.95 E. The main economic crops in the study area include cotton and pepper, with the cotton variety being upland cotton planted at row spacing of 70 cm and plant spacing of 10 cm. Fertilization and management followed the requirements of the Xinjiang Agricultural Technology Extension Station. Drip irrigation was applied once during the entire growth period, generally 10 times, with foliar fertilization applied once each during the seedling stage, two-leaf stage, and before the first irrigation, with irrigation and fertilization ending around August 20.

1.2 Data Collection

The research team conducted leaf spectral measurement experiments in Wuzhen Town, Kuqa City, Aksu Region, Xinjiang, on July 20–21, 2021. An ASD FieldSpec HandHeld portable spectrometer was used to measure cotton leaf hyperspectral data across a wavelength range of 325–1075 nm. Measurements were taken under stable solar radiation conditions with no wind or clouds between 11:00–14:00. The probe was positioned 25 cm above the cotton leaf canopy. Spectral scanning time was set at 30 minutes, with whiteboard calibration performed after each sample measurement to eliminate light variation effects. Samples were selected from disease-free, pest-free cotton canopies with uniform growth, with the second or third cotton leaf chosen for spectral measurement. The measured reflectance data were processed using ViewSpec PRO software to calculate the average reflectance spectrum for each sample point. A total of 120 cotton leaf samples were collected, designated as Dataset I. To better validate model stability, 120 cotton leaf samples from the seedling stage were selected as Dataset II for further validation.

1.3 Leaf Water Content Measurement

Immediately after spectral measurement, collected leaves were weighed using an electronic balance with 0.001 g precision to obtain fresh weight (FW). Leaves were then placed in sealed bags, and drying was performed in the laboratory after fieldwork completion. Samples were oven-dried at 105°C for 30 minutes, then at 80°C until constant weight was achieved before measuring dry weight (DW). Leaf water content (LWC) was calculated as:

$$\text{LWC} = (\text{FW} - \text{DW}) / \text{FW} \times 100\%$$

1.4 Division of Calibration and Validation Sets

Random sampling was used to divide the collected samples, with 80 samples designated as the calibration set and the remaining 40 as the validation set. Dataset II was divided using the same proportion, with 80 samples as the calibration set and 40 as the validation set. Statistical analysis of each subset is presented in . The statistical indicators for the sample set, calibration set, and validation set in Dataset I are relatively consistent, indicating that the sample division meets the requirements of randomness and representativeness for spectroscopic modeling.

1.5 Spectral Feature Variable Selection Methods

To evaluate the effectiveness of various screening methods for leaf water inversion, five feature variable selection methods were employed (TABLE:2). To fully leverage the role of fractional-order differentiation in refining spectral information and improve modeling effectiveness, fractional-order differentiation was applied to preprocess the original spectral data for correlation coefficient (CC) analysis. CARS, SPA, GA, and MC-UVE methods used raw data for feature band selection.

TABLE:2 Spectral characteristic variable screening methods

Method	Principle	Advantages	Disadvantages
CC	Calculates correlation coefficients between spectral variables and target values	Simple and efficient operation	Multicollinearity among variables
CARS	Uses adaptive reweighted sampling and exponential function to remove low-weight variables	Effectively removes highly autocorrelated bands, suitable for high-dimensional data	Multicollinearity among variables, low stability in band selection

Method	Principle	Advantages	Disadvantages
SPA	Selects variables with minimal redundancy through forward variable selection	Minimal variable redundancy and multicollinearity, reduces modeling time	Does not consider multicollinearity among all characteristic wavelengths; tends to select variables with low collinearity rather than effective variables
GA	Uses genetic operators (selection, crossover, mutation) for global optimization	Global optimization capability	Requires multiple runs to determine optimal variable subset
MC-UVE	Uses Monte Carlo sampling and leave-one-out cross-validation to calculate variable stability	High stability	Requires threshold definition, leading to variable number changes
CARS-SPA	Combines CARS for coarse screening and SPA for refined selection	Further removes redundant information, extracts effective bands with low multicollinearity, high computational efficiency	Fewer selected variables may lose key information; modeling effectiveness of characteristic wavelength set is heavily influenced by coarse screening results

1.6 Model Construction and Validation

The machine learning method employed in this study was WOA-improved RFR, compared with standard RFR. The whale optimization algorithm, proposed by Mirjalili in 2016, simulates the bubble-net foraging behavior of humpback whales through mathematical modeling of encircling prey, spiral position updating, and random search behavior.

1.6.1 Encircling Prey During hunting, the whale closest to the target prey is considered the optimal position, with other whales moving toward this position to encircle the prey. The mathematical expressions are:

$D = |C \cdot X(t) - X(t)/ X(t+1) = X(t) - A \cdot D$

where D is the distance between the optimal individual position and the current individual position; X*(t) is the position vector of the current optimal solution; X(t) is the position vector of the current solution; X(t+1) is the iteration position vector; t is the iteration number; A and C are parameter vectors calculated as:

$A = 2a \cdot r_1 - a$ $C = 2 \cdot r_2$ $a = 2 - 2t/T_{max}$

where r_1 and r_2 are random vectors in [0,1]; a linearly decreases from 2 to 0; and T_{max} is the maximum iteration number.

1.6.2 Spiral Bubble-Net Attacking Whales attack through two mechanisms: shrinking encirclement and spiral position updating. The mathematical model is established by calculating the distance to the prey:

$D' = |X(t) - X(t)/ X(t+1) = D' \cdot e^{(bl)} \cdot \cos(2\pi l) + X(t)$

where D' is the distance between the whale and the prey; b is the logarithmic spiral coefficient; and l is a random number in (-1,1).

1.6.3 Searching for Prey In addition to spiral bubble-net searching, whales also perform random movements based on their position relative to the prey. This behavior is selected according to vector A, which takes random values outside (-1,1) to search for more suitable targets and enhance global optimization capability:

$X(t+1) = X_{rand} - A \cdot D$ $D = |C \cdot X_{rand} - X(t)|$

where X_{rand} is the position vector of a randomly selected individual from the whale population.

Based on experimental results, the whale evolution parameters were set as follows: maximum iterations $T_{max} = 30$, population size $N = 20$. These parameters were used to optimize the number of regression trees, maximum tree depth, minimum samples required for node splitting, and minimum samples for leaf nodes.

Model prediction accuracy was evaluated using determination coefficient (R^2) and root mean square error (RMSE) for both calibration and validation sets. R^2 closer to 1 indicates higher model accuracy and better fitting effect. RMSE measures prediction error magnitude, with smaller values indicating better predictive capability:

$RMSE = \sqrt{(1/n \Sigma(\hat{y}_i - y_i)^2)}$

where $\hat{y}_i$ is the predicted value, y_i is the measured value, and n is the sample number.

2.1 Spectral Feature Variable Selection Results and Analysis

In this study, five variable screening methods were applied to spectral data after fractional-order differentiation. The original spectral reflectance was processed at 0.2 intervals from 0 to 2. As shown in [Figure 3: see original paper], after fractional-order differentiation transformation, 17 bands passed significance tests at the 0.01 level across all orders, with the maximum absolute correlation coefficient reaching 0.65. Pearson correlation analysis was conducted between original spectra and fractional-order derivative spectra (using 1.5 order as an example) and LWC, revealing that no bands in the original spectrum passed the 0.01 significance test. Therefore, fractional-order differentiation was employed to improve correlation levels. This preprocessing avoided information omission, reduced noise effects, and enhanced spectral expression capability.

CARS method: As iterations progressed, the number of characteristic wavelengths decreased at a decreasing rate, indicating that the algorithm first removed bands weakly correlated with LWC, then eliminated strongly correlated bands. After CARS operation, many characteristic wavelength variables remained, with potential multicollinearity. Therefore, CARS was coupled with SPA for further screening.

SPA method: Through forward variable selection, SPA extracted 8 characteristic wavelengths, accounting for 2.2% of original bands, effectively reducing spectral information redundancy.

GA method: Compared with SPA, GA further reduced computational load. As the number of selected variables increased, RMSE rose rapidly. When the variable number was 38, RMSE stabilized, indicating this as the optimal subset. GA extracted 38 characteristic wavelengths, accounting for 10.5% of original bands.

MC-UVE method: This method calculated stability values for all bands through Monte Carlo sampling, ultimately selecting 8 wavelengths as the characteristic band subset.

CARS-SPA method: The coupling method selected the fewest variables. After CARS operation, SPA screening extracted 8 characteristic wavelengths, accounting for 2.2% of original bands. The selected bands were concentrated in the ultraviolet and near-infrared regions, with the near-infrared region being the sensitive area for leaf water content.

The distribution of bands obtained by the five variable screening methods is shown in [Figure 10: see original paper]. Analysis revealed that SPA, GA, and CARS-SPA methods showed relatively consistent band positions, primarily distributed in the near-infrared region.

2.2 Model Establishment and Analysis

To investigate the impact of different variable screening methods on model accuracy, RFR and WOA-RFR models were constructed using full bands and the five screening methods. Correlation analysis after fractional-order differentiation showed that the 1.5-order differential model achieved $R^2 \geq 0.881$ and $RMSE \leq 0.042$, indicating good preprocessing effects. Therefore, 1.5-order differentiation was selected for subsequent analysis.

As shown in , full-band models showed poor prediction performance. In contrast, models based on characteristic bands selected by all five methods demonstrated significantly improved accuracy. The CARS-SPA-WOA-RFR model achieved the best performance with $R^2 = 0.93$ and $RMSE = 0.032$. The validation set R^2 and RMSE were similar to the calibration set, indicating good model stability.

To verify model stability, Dataset II was used as an independent validation dataset for the CARS-SPA-WOA-RFR model. As shown in [Figure 11: see original paper] and [Figure 12: see original paper], the model's fitted values were evenly distributed on both sides of the 1:1 line, indicating good fitting performance. However, other models showed lower R^2 values and larger errors between measured and predicted values. The CARS-SPA-WOA-RFR model achieved the highest R^2, enabling accurate estimation of cotton leaf water content.

3 Discussion

Spectral preprocessing is essential for improving data quality and modeling accuracy. This study used fractional-order differentiation to process cotton leaf hyperspectral data, significantly enhancing the correlation between spectral data and leaf water content. This finding aligns with previous research by Yu Lei et al. and Hasan Umut et al., who reported improved correlation levels after processing soil and wheat leaf spectra.

Due to the large number of bands, hyperspectral data contains numerous bands unrelated to water content with low contribution levels, resulting in poor full-band modeling performance. This study effectively screened characteristic wavelengths through five variable selection methods, reducing data redundancy and improving modeling effectiveness. This conclusion is consistent with Zhang et al.'s findings on characteristic wavelength selection. Particularly, the coupling of CARS and SPA in this study more effectively screened characteristic wavelengths, yielding the optimal final model, which aligns with conclusions from Yu Lei et al.

Different machine learning models produce varying prediction results for the same data. The WOA-RFR model demonstrated superior accuracy for cotton leaf water content inversion compared with the standard RFR model. After validation with independent sample sets, the model remained stable and achieved good inversion performance. This result is consistent with Mohammadi et al.'s finding that WOA-improved models can effectively enhance prediction accuracy.

This study used fractional-order differentiation to preprocess original spectral data, effectively reducing environmental impacts on spectral measurements. However, field data collection remains influenced by soil, atmosphere, and surrounding canopy effects, with regional differences causing slight variations in cotton leaf hyperspectral characteristics across different areas. Consequently, the selected characteristic bands differ somewhat from water-sensitive bands identified in previous studies, showing some offset and concentrating more in the ultraviolet and visible regions. Additionally, this study used datasets from two different growth stages for repeated model validation, partially overcoming the limitation of single-dataset studies in previous research. However, the estimation results still exhibited some "low-value overestimation and high-value underestimation" phenomena. Future research should employ more optimized machine learning algorithms for model validation and correction to further improve model stability and applicability.

4 Conclusions

1) Fractional-order differentiation spectral preprocessing can improve correlation levels, with 1.5-order processing showing particularly noticeable effects.

2) Different feature band selection methods yield varying numbers and positions of bands. MC-UVE extracted the most bands (38), while SPA, GA, and CARS-SPA showed relatively consistent band positions. CARS-SPA extracted the fewest variables (8), with significant differences among methods.

3) The WOA-RFR model achieved good inversion performance. The CARS-SPA-WOA-RFR model demonstrated high inversion accuracy, with model prediction $R^2 = 0.93$ and RMSE = 0.032, indicating that this model can achieve good precision for predicting cotton leaf water content across different growth stages and varieties.

References

[1] Liu Wenjing, Fan Yongsheng, Dong Yan Qi, et al. Analysis and suggestions on the current situation of cotton production in China[J]. China Seed Industry, 2022(1): 21-25.

[2] Ma Chunyue, Sawut Mamat, Ablet Ershat, et al. Characteristics and influencing factors of geographical agglomeration of cotton plantation in Xinjiang[J]. Acta Agronomica Sinica, 2019, 45(12): 1859-1867.

[3] Sun Jun, Cong Sunli, Mao Hanping, et al. CARS model for predicting leaf moisture of leaf used lettuce based on hyperspectral[J]. Transactions of the

Chinese Society of Agricultural Engineering, 2017, 33(5): 178-184.

[4] Junttila S, Hölttä T, Saarinen N, et al. Close range hyperspectral spectroscopy reveals leaf water content dynamics[J]. Remote Sensing of Environment, 2022, 277: 113071.

[5] Li L, Ustin S L, Riano D. Retrieval of fresh leaf fuel moisture content using genetic algorithm partial least squares (GA-PLS) modeling[J]. IEEE Geoscience and Remote Sensing Letters, 2007, 4(2): 261-265.

[6] Sun J, Zhou X, Hu Y G, et al. Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging[J]. Computers and Electronics in Agriculture, 2019, 160: 153-159.

[7] Li X L, Wei Z X, Peng F F, et al. Estimating the distribution of chlorophyll content in CYVCV infected lemon leaf using hyperspectral imaging[J]. Computers and Electronics in Agriculture, 2022, 198: 107036.

[8] Su Yi, Wang Keru, Li Shaokun, et al. Monitoring models of the plant water content based on cotton canopy hyperspectral reflectance[J]. Cotton Science, 2010, 22(6): 554-560.

[9] Yang Baohua, Chen Jianlin, Chen Linhai, et al. Estimation model of wheat canopy nitrogen content based on sensitive bands[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(22): 176-182.

[10] Wang Qiang, Yi Qiuxiang, Bao Anming, et al. Discussion on hyperspectral index for the estimation of cotton canopy water content[J]. Spectroscopy and Spectral Analysis, 2013, 33(2): 507-512.

[11] Zhang Wenxu, Tong Xuanmeng, Zhou Tianhang, et al. Remote sensing estimation of cotton leaf nitrogen content based on hyperspectral imaging[J]. Journal of Shenyang Agricultural University, 2021, 52(5): 586-596.

[12] Yi Xiang, Zhang Lifu, Lü Xin, et al. Estimation of cotton above-ground biomass based on unmanned aerial vehicle hyperspectral and successive projections algorithm[J]. Cotton Science, 2021, 33(3): 224-234.

[13] Mamat Yusup, Hasan Tursun, Gulam Magpirat. Remote sensing of cotton plantation areas monitoring in delta oasis of Ugan-Kucha River, Xinjiang[J]. Research of Agricultural Modernization, 2014, 35(2): 240-243.

[14] Li L L, Sun J, Tseng M L, et al. Extreme learning machine optimized by whale optimization algorithm using insulated gate bipolar transistor module aging degree evaluation[J]. Expert Systems with Applications, 2019, 127: 58-67.

[15] Zhou J, Zhu S Li, Qiu Y G, et al. Predicting tunnel squeezing using support vector machine optimized by whale optimization algorithm[J]. Acta Geotechnica, 2022, 17: 1343-1366.

[16] Zhao F, Li W D. A combined model based on feature selection and WOA for PM2.5 concentration forecasting[J]. Atmosphere, 2019, 10(4): 223.

[17] Li Qiyong, Zhang Weibin, Zhao Xinzhe, et al. Global maximum power point tracking for PV array based on support vector regression optimized by improved whale algorithm[J]. Transactions of China Electrotechnical Society, 2021, 36(9): 1771-1781.

[18] Hasan Umut, Sawut Mamat, Ma Chunyue. Hyperspectral estimation of wheat leaf water content using fractional differentials and successive projection algorithm-back propagation neural network[J]. Laser & Optoelectronics Progress, 2019, 56(15): 251-259.

[19] Chen Peng. Application of fractional differential algorithm in medical ultrasonic elastic image denoising[D]. Kunming: Kunming University of Science and Technology, 2018.

[20] Li Changchun, Shi Jinjin, Ma Chunyan, et al. Estimation of chlorophyll content in winter wheat based on wavelet transform and fractional differential[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(8): 172-182.

[21] Wang Jiawen, Peng Jie, Ji Wenjun, et al. Soil pH inversion based on electromagnetic induction data in cotton field of southern Xinjiang[J]. Arid Zone Research, 2022, 39(4): 1293-1302.

[22] Li H D, Liang Y Z, Xu Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648(1): 77-84.

[23] Zhang J K, Rivard B, Rogge D M. The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data[J]. Sensors, 2008, 8(2): 1321-1342.

[24] Cheng Jiehong, Chen Zhengguang, Yi Shujuan. Wavelength selection algorithm based on minimum correlation coefficient for multivariate calibration[J]. Spectroscopy and Spectral Analysis, 2022, 42(3): 719-725.

[25] Song Xiangzhong. Research of three new wavelength selection methods in near infrared spectroscopy quantitative analysis area[D]. Beijing: China Agricultural University, 2017.

[26] Mirjalili S, Lewis A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51-67.

[27] Mohammadi B, Mehdizadeh S. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm[J]. Agricultural Water Management, 2020, 237: 106145.

[28] Jia M, Li W, Wang K K, et al. A newly developed method to extract the optimal hyperspectral feature for monitoring leaf biomass in wheat[J]. Computers and Electronics in Agriculture, 2019, 165: 104942.

[29] Han Z Z, Deng L M. Application driven key wavelengths mining method for aflatoxin detection using hyperspectral data[J]. Computers and Electronics

in Agriculture, 2018, 153: 248-255.

[30] Zhang M J, Zhang S Z, Iqbal J. Key wavelengths selection from near infrared spectra using Monte Carlo sampling recursive partial least squares[J]. Chemometrics and Intelligent Laboratory Systems, 2013, 128: 17-24.

Note: Figure translations are in progress. See original paper for figures.