

Leaf Area Index Estimation Model for UAV Image Hyperspectral Data Based on Wavelength Variable Selection and Machine Learning Methods

Juanjuan Zhang

Henan Agricultural University

Tao Cheng (✉ 15038156398@163.com)

Henan Agricultural University <https://orcid.org/0000-0001-6250-2581>

Wei Guo

Henan Agricultural University

Xin Xu

Henan Agricultural University

Xinming Ma

Henan Agricultural University

Yimin Xie

Henan Agricultural University

Hongbo Qiao

Henan Agricultural University

Research

Keywords: Winter wheat, Leaf area index, Unmanned aerial vehicle, Hyperspectral imaging data, Characteristic bands, Machine learning, Model.

Posted Date: December 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-131883/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background In order to accurately estimate leaf area index (LAI) of winter wheat by using unmanned aerial vehicle (UAV) hyperspectral imagery.

Methods The UAV hyperspectral imaging data, alternating slice-wise diagonalization (ASD) spectral data, and LAI were simultaneously obtained at main growth stages (jointing stage, booting stage, and filling stage) of various winter wheat varieties under various nitrogen fertilizer treatments. The characteristic bands related to LAI were extracted from UAV hyperspectral data with different algorithms including first derivative (FD), successive projections algorithm (SPA), competitive adaptive reweighted sampling (CARS), and competitive adaptive reweighted sampling combined with successive projections algorithm (CARS_SPA). Furthermore, three modeling machine learning methods including partial least squares regression (PLSR), support vector machine regression (SVR), and extreme gradient boosting (Xgboost) were used to build LAI estimation models.

Results Our results show that the correlation coefficient between UAV and ASD hyperspectral data is greater than 0.99, indicating the UAV data can be used for estimation of wheat growth information. The LAI bands selected by using different algorithms were slightly different among the 15 models built in this study. The Xgboost model using nine consecutive characteristic bands selected by CARS_SPA algorithm as input was proved to have the best performance. This model yielded identical results of coefficient of determination (0.89) for both calibration set and validation set, indicating a high accuracy of this model.

Conclusions The Xgboost modeling method in combine with CARS_SPA algorithm can reduce input variables and improve the efficiency of model operation. Our results provide reference and technical support for nondestructive and rapid estimation of winter wheat LAI by using UAV.

Background

Leaf area index (LAI), which is defined as half of the all-sided leaf area per unit ground area [1], is a key biophysical parameter to determine the photosynthesis, respiration, and transpiration of vegetation canopy [2,3]. Winter wheat (*Triticum aestivum* L.) is a main food crop in China. It is of great importance to obtain the winter wheat LAI rapidly and effectively to monitor wheat growth, manage water and fertilizer application, and predict the yield. Destructive methods for measuring winter wheat LAI normally provide more precise results, but the assessment is time-consuming, labor-intensive, and expensive. The basic requirement of modern agriculture is to conduct real-time, fast, and accurate measurement of winter wheat LAI in the field, yet it is difficult to actualize. Remote sensing is a reliable, fast, and non-destructive way to monitor growth parameters of crops. Specifically, the newly emerged low-altitude remote sensing detection technology based on UAV exhibits characteristics of high spatial resolution, strong timeliness, low cost, and low flight altitude that does not require a flight permit. This technology can fill the gap between ground-based monitoring and satellite remote sensing of dynamically monitoring crop growth at multiple scales, therefore, has been widely used in precision agriculture [4–9].

UAV remote sensing platform can carry digital cameras featured with simultaneous interpreting and multispectral or hyperspectral sensors. The accuracy of crop phenotypic information by remote sensing are varied due to inconsistent performance among devices [10]. Modern hyperspectral sensors that continuously cover all spectral regions can successfully reflect the characteristics of crops. UAV has been used in earlier studies to analyze the characteristic bands related to LAI. The estimation models of wheat LAI built based on vegetation indices (VIs) are used to extract wheat growth information from low-altitude UAV hyperspectral data. For example, by combining UAV-based hyperspectral data, Fu et al. [11] used RESAVI to effectively invert winter wheat LAI. Xie et al. [12] estimated the wheat LAI using six spectral indices of UAV hyperspectral data. However, spectral indices show different degrees of saturation [13,14], and their universality and accuracy are easily disturbed by external factors [15]. Therefore, to improve the modeling accuracy, some studies combined UAV hyperspectral data with machine learning (ML) methods, such as support vector machine regression (SVR), partial least squares regression (PLSR), neural network, and random forest (RF), for modeling. For example, by using UAV hyperspectral data, Gao et al. [16] extracted spectral features based on UAV hyperspectral data, and constructed the winter wheat LAI estimation model using PLSR. Yue et al. [17] constructed estimation model of wheat LAI by using VIs, RF and PLSR. In aforementioned studies, the extracted spectral parameters and ML algorithms are combined to construct inversion models of physiological parameters. In recent years, algorithms for spectral features extraction such as principal component analysis (PCA), variable projection importance (VIP), genetic algorithm (GA), and continuous projection algorithm (SPA) have been widely used in studies of ground monitoring or satellite remote sensing [18,19]. These algorithms can effectively remove the redundancy in hyperspectral data, thus reduce the risk of overfitting, and finally obtained a model of robust and high prediction accuracy [20–22]. Due to the inconsistency in performance of sensors, spectral data obtained are varied. However, at present, limited research has been conducted on dimensionality reduction and characteristic band extraction based on UAV hyperspectral imaging data and ML methods for winter wheat LAI estimation.

In this study, the UAV hyperspectral data, ASD non-imaging hyperspectral data, and LAI measurements of winter wheat at various key growth stages were obtained. Various variables extraction algorithms were used to extract characteristic bands related to LAI. Next, three ML methods (PLSR, SVR, and Xgboost) were employed to construct estimation models of winter wheat LAI based on selected spectral variables. Last, these models were comprehensively compared and a most suitable model for estimating winter wheat LAI was determined. This study provides the methodology and technical support for UAV remote sensing on winter wheat LAI estimation.

1 Materials And Methods

1.1 Study area and experimental design

Experiments were conducted from 2017 to 2018 at Xindian Regional Test Station, Yancheng District, Luohe City, Henan Province, China (113°53'1"E, 33°41'60"N) (Fig. 1). The area has a warm wet monsoon climate, with precipitation mostly occurring in summer and autumn, an average annual temperature of

about 14.6°C. The contents of soil organic matter, total nitrogen, alkali hydrolyzed nitrogen, available phosphorus, and available potassium are 13.33 g kg⁻¹, 1.02 g kg⁻¹, 92.33 g kg⁻¹, 54.02 mg kg⁻¹, and 299 mg kg⁻¹, respectively. The winter wheat varieties used in this study include Zhoumai 27 (ZM27), Yumai 49–198 (YM49-198), Xinong 509 (XN509), and Aikang 58 (AK58). Winter wheats were planted in 44 plots (each of size 16 m × 9 m) including three repeats (Fig. 2). Different nitrogen was used in each plot. A set of four nitrogen treatments were used: 0 (N0), 120 kg.hm⁻² (N8), 225 kg.hm⁻² (N15), and 330 kg.hm⁻² (N22). The ratio of base fertilizer topdressing is 6:4. The base fertilizer was applied before sowing and the topdressing was applied at the jointing stage. Winter wheats were machine-sown on October 23, 2017, at a sowing rate of 180 kg hm⁻². Other cultivation and management measures were generally the same as those applied in high yield wheat fields. The specific design is shown in Fig. 2.

1.2 Data Collection and Index Determination

All experimental data including UAV-based hyperspectral imaging data, field hyperspectral reflectance, and wheat LAI, were collected at three key growth stages of winter wheat in 2018. The specific dates for data collection were March 11 (jointing stage), April 8 (booting stage), and May 12 (filling stage).

1.2.1 UAV-Based Imaging Hyperspectral Data

The UAV remote sensing platform is composed of an AZUP-T8 eight-propeller UAV (XIRO company, China), and a UHD185 high-resolution imaging spectrometer (Cubert, Germany). The UHD185 acquires wavelengths from the visible to the near-infrared (450–950 nm) from 125 spectral bands. The sampling interval of the hyperspectral data is 4 nm, and the spectral resolution is 8 nm. Hyperspectral data acquisition was carried out on cloudless and windless days. The reference plate was used to calibrate UHD185 before data collection. The height, moving speed, course overlap, and side overlap were set as 50 m, 6 m s⁻¹, 80%, and 60%, respectively. The data processing flow is shown in Fig. 3.

1.2.2 Measurement of Field Hyperspectral Reflectance

Canopy reflectance measurement of each plot was collected with an ASD FieldSpec4 Portable high-resolution spectrometer (ASD Inc., USA) with the spectral range of 350–2500 nm. ASD measurement was carried out before UAV data acquisition. The measurements were taken 1 m above the canopy with a 25° field of view optic. Three spectral measurements were acquired per plot (10 spectra as a sampling interval) and averaged to a single measurement for further analysis. The sampling interval of the hyperspectral data is 1.4 nm at 350–1000 nm, and 2 nm at 1000–2500 nm, and the spectral resolutions are 3 and 2 nm, respectively. Standard whiteboard correction was performed before and after the measurement. The exact position of each plot was located by a GPS device.

1.2.3 Measurement of LAI

Winter wheat LAIs were measured at fixed sample points of each plot. Ten winter wheat plants were randomly sampled from each plot, immediately sealed into a paper bag, and brought back to the laboratory for separation of stems and leaves. LAIs were measured according to the method proposed by Liu et al. [23].

1.3 Data Analysis

A total of 132 winter wheat LAIs were collected in this study. Eight were removed due to test errors, and the resulting 124 LAIs were used for the following data analysis. To evaluate the models' robustness of temporal variation, the LAIs of area 1 and 3 (80 samples) were used as the calibration set, and the LAIs of area 2 (44 samples) were used as the standalone validated set. The LAI data are summarized in Table 1.

Table 1
Summary of leaf area index (LAI) of winter wheat

Sample type	Sample Number	Maximum Value	Minimum Value	Mean Value	Standard Deviation	Coefficient Of Variation
Total Sample	124	8.38	1.78	5.43	1.70	0.32
Calibration set	80	8.38	1.78	5.33	1.74	0.33
Validation set	44	8.24	3.10	5.62	1.62	0.30

The characteristic bands of LAI in UAV hyperspectral data were extracted with various algorithms including the first derivative (FD), successive projections algorithm (SPA), competitive adaptive reweighted sampling (CARS) and competitive adaptive reweighted sampling combined with successive projections algorithm (CARS_SPA). In combine with the full spectrum information, three ML methods including partial least squares regression (PLSR), support vector machine regression (SVR), and extreme gradient boosting (Xgboost) were used to construct LAI estimation models. The flowchart of UAV hyperspectral imaging data processing and data analysis for constructing winter wheat LAI estimation model is presented in Fig. 3.

1.3.1 Algorithms for Spectral Variables Extraction

Four algorithms for spectral variables extraction were used in the present study. (1) FD can effectively remove the interference of linear and near-linear background noise in the raw spectrum, enhancing spectral characteristic differences. (2) SPA selects the combination of variables with minimum redundancy from the spectral matrix to minimize the collinearity among variables, resulting to a great reduction in the number of variables used as input for modeling. Therefore, the complexity of the model is reduced whereas the stability and accuracy of the model are improved [24]. (3) CARS is a new variable selection algorithm, which follows the principle of "survival of the fittest". The exponential decreasing function (EDF) and adaptive reweighting sampling (ARS) are used to remove the wavelength variables with low absolute value of regression coefficients in PLSR model. After that, root mean square error of cross-validation (RMSECV) calculation are used to select important variables with good stability in several iterations, and a final subset of variables is obtained. The subset with the lowest RMSEV is considered as the optimal variable subset [25]. (4) CARS_SPA. By combining the advantages of the two algorithms (SPA and CARS) [26], the CARS_SPA algorithm not only minimize the spectral variable redundancy, but also reduces the interference of invalid variables during the SPA calculation process.

1.3.2 Modeling Methods

Three ML methods were used for data mining and matrix recognition. (1) PLSR is a regression modeling method to predict a set of dependent variables from a set of independent variables. It combines features of principal component analysis, canonical correlation analysis, and multiple regression analysis. PLSR can solve the multi-collinearity among independent variables and is suitable for small sample datasets. (2) SVR. The use of Lagrangian multipliers for data regression analysis is an extension of support vector machine classification to solve regression problems. It has a strong learning ability from small datasets and solves high-dimensional and nonlinear problems by transforming them into linear problems through nonlinear transformation. In this study, the Gaussian kernel function is used as the kernel function, and the GridSearchCV function is used to discover the optimal parameters (penalty coefficient cost and gamma). (3) Xgboost is a new efficient ensemble learning algorithm proposed by Chen [27]. It is an improved algorithm of gradient boosting, and uses the Taylor expansion to obtain the second derivative as independent variable. By separating the selection of loss function from the optimization of model algorithm and parameter selection, the applicability of Xgboost is increased, which makes it select loss function on demand. A strategy similar to RF is adopted to support data sampling, and make full use of the advantages of multi-core CPU parallel computing, which greatly improves the operation speed and prediction accuracy of the model [28]. The main kernel parameters are as follows: n_estimators, the maximum depth of the tree (max_depth), regularization parameters (min_child_weight), gamma, sampling method (subsample and colsample_bytree), and learning_rate.

1.4 Accuracy Evaluation

The accuracy of the LAI estimation models were evaluated using the coefficient of determination (R^2), root mean square error (RMSE) and relative percent deviation (RPD). The formulas are as follows:

$$R^2 = \frac{\sum_{i=1}^n (xi - \bar{x})^2 \times (yi - \bar{y})^2}{\sum_{i=1}^n (xi - \bar{x})^2 \times \sum_{i=1}^n (yi - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (yi - xi)^2}{n}} \quad (2)$$

$$RPD = \sqrt{\frac{\sum_{i=1}^n (xi - \bar{x})^2}{n-1}} / RMSE \quad (3)$$

Where, xi , \bar{x} , yi , and \bar{y} are the measured LAI, the mean measured LAI, the LAI predicted by the model, and the mean LAI predicted by the model, respectively; n is the number of data points. RPD can reflect the predictive power of the model, generally speaking, when $1.0 < RPD < 1.4$, the prediction ability of the model is poor; When $1.4 < RPD < 1.8$, the model can be used for correlation assessment; When $1.8 < RPD < 2.0$, the model can be used for quantitative prediction; When $2.0 < RPD < 2.5$, a better quantitative prediction can be made; When $RPD > 2.5$, the result shows that the model prediction is better. A larger R^2 and RPD indicates a better model fit, while a smaller RMSE indicate a higher model accuracy.

2 Results And Analysis

2.1 Reliability Verification of UAV Hyperspectral Imaging Data

In order to verify the reliability of UAV hyperspectral imaging data, the winter wheat hyperspectral data collected by ASD were resampled into UHD185 bands to calculate the average reflectance for each plot. Correlation analysis showed that the spectral reflectance of UHD185 and ASD are in high consistence at spectral bands between 458 and 830 nm, overlapping in both green peak position and red edge region (Fig. 4a). At spectral bands between 830 and 950 nm, the spectral reflectance of UHD185 gradually decreased, while that of ASD deviated from UHD185 and exhibited a flat curve (Fig. 4a). This may be due to relatively high noise between 830 and 950 nm since this range is close to the edge of UHD185 sensor spectrum. We compared spectral reflectance of UHD185 and ASD at 458–830 nm and found a high correlation ($R^2 > 0.99$) between them (Fig. 4b). These results indicate that the spectral reflectance of UHD185 is reliable between 458 and 830 nm (3–96 wavebands) and can be used to estimate winter wheat LAI.

2.2 Correlation between Winter Wheat LAI and UAV Hyperspectral Imaging Data

Comparisons were made between the winter wheat LAI and raw UHD185 spectral reflectance (458–830 nm), and between LAI and FD-transformed UHD185 spectral reflectance (458–830 nm). Between LAI and UHD185 spectral reflectance, the maximum negative and positive correlation coefficients were at 654 nm ($r = -0.80$) and 802 nm ($r = 0.49$), respectively (Fig. 5). Between LAI and FD-transformed spectral reflectance, the maximum negative and positive correlation coefficient were at 546 nm ($r = -0.74$) and 774 nm ($r = 0.83$), respectively. Spectral bands with absolute correlation greater than 0.6 are 498–506 nm, 542 nm, 546 nm, 738–786 nm, and 830 nm (Fig. 5).

2.3 Selection of Characteristic Bands related to Winter Wheat LAI from UAV Hyperspectral Reflectance Data

Four methods including FD, SPA, CARS, and CARS_SPA were used to select characteristic bands related to wheat LAI. The FD correlation analysis demonstrated that spectral bands of 498–506 nm, 542 nm, 546 nm, 738–786 nm, and 830 nm were in high correlation with winter wheat LAI (correlation coefficients >0.6). We selected bands corresponding to the maximum value at the inflection point, and discarded multiple highly correlated bands that are close to the inflection point. In this way, spectral bands at 506, 546, 774, and 830 nm were selected by FD, which account for 4.25% of the total variables.

During the process of SPA algorithm, the minimal and maximal numbers of characteristic bands extracted were set as 5 and 94, respectively. At the minimum RMSE of 1.049, a total of 28 optimal characteristic bands (458, 466, 474, 482, 498, 502, 506, 510, 518, 526, 530, 534, 542, 546, 558, 566, 570, 574, 610, 626, 650, 658, 686, 698, 710, 762, 814, and 830 nm), accounting for 29.8% of the total variables, were selected (Fig. 6).

The number of characteristic bands gradually decreased with the increase in CARS iteration (Fig. 7a). With the increase in sampling time, the 10-fold RMSE cross validation first slightly decreased, then increased significantly at the CARS iteration of 47 (Fig. 7b). This result indicates that certain key information was lost after performing 47 iterations of CARS, resulting to a poor performance of the model. At the iteration of 24, RMSE reached its minimum of 0.9674. Thirteen variables (566, 586, 602, 610, 634, 682, 698, 710, 730, 734, 790, 802, and 814 nm) accounting for 13.8% of the total variables were therefore selected (Fig. 7c).

The 28 characteristic bands selected using SPA algorithm may contain noise due to the complexity in SPA calculation process CARS algorithm can effectively remove the variables with small weight, and effectively select variables closely related to LAI. Therefore, the 28 characteristic bands selected by SPA algorithm were filtered by CARS algorithm to obtain the optimal combination of characteristic bands. The result showed the minimum RMSE (0.9718) appeared at the CARS iterations of 15 (Fig. 8). Therefore, nine variables (466, 474, 518, 526, 610, 658, 710, 814, and 830 nm) accounting for 9.57% of the total variables were selected.

We compared the LAI characteristic bands selected by the four algorithms, the distributions of characteristic bands selected by different algorithms were in consistent to a large extent, yet differences were also observed (Fig. 9).

2.4 Construction of Winter Wheat LAI Estimation Model by Different Modeling Methods

Based on characteristic bands of wheat LAI selected by using different variable selection algorithms and full spectrum information, three modeling methods including PLSR, SVR, and Xgboost were employed to construct LAI estimation models. Independent samples including calibration and validation sets were used to test these models. Table 2 summarizes the results for the models obtained by different modeling methods.

Table 2
Regression analysis of characteristic bands and winter wheat LAI

Modeling method	Variable extraction	Wavelengths numbers	Calibration			Validation		
			R ²	RMSE	RPD	R ²	RMSE	RPD
PLSR	Full_spectrum	94	0.79	0.81	2.15	0.78	0.79	2.06
	FD	4	0.81	0.77	2.26	0.81	0.75	2.15
	SPA	28	0.79	0.81	2.14	0.78	0.81	2.01
	CARS	13	0.80	0.79	2.22	0.81	0.78	2.08
	CARS_SPA	9	0.83	0.73	2.39	0.83	0.74	2.19
SVR	Full_spectrum	94	0.80	0.79	2.20	0.77	0.79	2.06
	FD	4	0.80	0.78	2.22	0.80	0.72	2.24
	SPA	28	0.81	0.75	2.31	0.79	0.76	2.14
	CARS	13	0.79	0.79	2.20	0.77	0.79	2.04
	CARS_SPA	9	0.82	0.75	2.31	0.84	0.65	2.75
Xgboost	Full_spectrum	94	0.93	0.50	3.48	0.80	0.79	2.04
	FD	4	0.88	0.65	2.69	0.81	0.75	2.17
	SPA	28	0.84	0.73	2.37	0.76	0.82	1.96
	CARS	13	0.82	0.82	2.12	0.82	0.84	1.93
	CARS_SPA	9	0.89	0.63	2.51	0.89	0.55	2.92

A reliable PLSR model was constructed using 9 characteristic bands selected by CARS_SPA algorithm as input. In this model, similar results of model evaluation indices (R^2 , RMSE, and RPD) were obtained for the calibration and validation sets. The R^2 , RMSE, and RPD of the calibration set were 0.83, 0.73, and 2.39, respectively, and those of the validation set were 0.83, 0.74, and 2.19, respectively. The PLSR model constructed based on 28 characteristic bands selected by SPA algorithm showed a poor performance.

The R^2 , RMSE, and RPD of the calibration set were 0.79, 0.81, and 2.14, respectively, and those of the validation set were 0.78, 0.81, and 12.01, respectively.

The SVR method based on different combinations of characteristic bands yield similar results of evaluation indices. Among these SVR models, the one using 9 characteristic bands selected by CARS_SPA algorithm as input showed the best performance. The R^2 , RMSE, and RPD of the calibration set were 0.82, 0.75, and 2.31, and those of the validation set were 0.84, 0.65, and 2.75.

Further analysis was performed on the model constructed by using Xgboost method. The Xgboost model based on the 28 characteristic bands selected by SPA algorithm showed a poor performance. The R^2 , RMSE, and RPD of the calibration set were 0.84, 0.73, and 2.37, respectively, and those of the validation set were 0.76, 0.82, and 1.96, respectively. The Xgboost model showed the best performance when using 9 characteristic bands selected by CARS_SPA algorithm. The R^2 , RMSE, and RPD of the calibration set were 0.89, 0.63, and 2.51, respectively, and those of the validation set were 0.89, 0.55, and 2.92, respectively.

In summary, among all characteristic bands combinations, the one containing 9 characteristic bands selected by CARS_SPA algorithm outperformed with either of the ML modeling methods, followed by models constructed using the four characteristic bands selected by FD algorithm. This may be due to the fact that these 9 characteristic bands selected by CARS_SPA algorithm are uniformly distributed within the spectral range of 458–830 nm, which thus well maintain the spectral information of the reflectance corresponding to LAI inversion. The three LAI estimation models constructed based on characteristic bands selected by CARS_SPA were superior to LAI models constructed based on the full spectrum. However, the accuracies of LAI models constructed based on selected characteristic band were different from those of the models constructed based on full spectrum information. These results demonstrated that characteristic bands extraction could greatly reduce the number of variables used for modeling, thus reduces the modeling complexity which improves the modeling efficiency while ensures its accuracy. Comparing the three ML modeling methods, the Xgboost models performed the best, followed by PLSR and SVR. The calibration and validation results of the best-performed model (LAI_CARS_SPA_Xgboost) are shown in Fig. 10.

3 Discussion

In combine with VI or other modeling methods, a large number of wheat LAI estimation models have been built based on selected characteristic bands related to wheat LAI. However, the characteristic bands used in different studies were quite varied. For example, Cheng et al. [29] quantitatively analyzed UAV hyperspectral data, and found that 516, 636, 702, 760, and 867 nm were the most sensitive bands to LAI changes. Gao et al. [30] constructed a ratio VI model for estimating wheat LAI using UAV hyperspectral data of 494 and 610 nm, resulting to a modeling R^2 of 0.74 and a prediction R^2 of 0.78. Xie et al. [31] used airborne hyperspectral data to study the normalized VI and found that models constructed at 660 and 785 nm show better estimation winter wheat LAI. By using the hyperspectral data of UAV, Tao et al.

[32] confirmed that the linear combination index (LCI) constructed at 670, 710, and 850 nm and the plant biochemical index (PBI) constructed at 560 and 810 nm were coupled with red edge parameters respectively, which can be used to accurately estimate winter wheat LAI. Based on the UAV hyperspectral data and optimal index algorithm, Chen et al. [33] determined 454, 754, and 834 nm were the optimal bands for winter wheat LAI estimation model during the flowering period. The red and near-infrared bands used in the above studies can better reflect the dynamic changes of LAI [34]. Bands of 518, 610, 658, 710, 810, and 830 nm selected in this study are similar to those in the aforementioned studies. In addition, the 9 characteristic bands selected by CARS_SPA algorithm in this study also include 466, 474, and 526 nm, which are the convoluted absorption bands of chlorophyll and carotenoids [35]. These bands are in the blue-green light spectral range, which would perform better in LAI prediction if combined with red and near-infrared bands [36,37]. The three models built based on the 9 characteristic bands selected by CARS_SPA algorithm all showed outstanding prediction results, indicating that the nine bands may contain effective information related to winter wheat LAI. These results may provide reference for other related studies.

The ML methodologies can be used to effectively analyze and utilize information-rich datasets as well as high-dimension observation data [38]. It has been used in the analysis and modeling of remote sensing data, but the accuracies of different ML methods are varied. Based on UAV hyperspectral data, Tao et al. [32] combined the VI and red edge parameters, and constructed an estimation model of winter wheat LAI using PLSR method. This model yielded a modeling R^2 of 0.80, and a prediction R^2 of 0.75. Based on UAV hyperspectral data, Tao et al. [39] built a winter wheat LAI inversion model in combination with multiple linear regression of random forest in the flowering period, which performed the best, resulting to a modeling R^2 of 0.68, and a prediction R^2 of 0.85. In this study, three ML methods were employed to construct winter wheat LAI estimation model. In general, the Xgboost models perform the best. This may be due to the fact that Xgboost is fast, highly effective, suitable for large-scale data processing, and has custom loss function [40], Xgboost can tackle majority of the flaws that emerged from the existing modeling methods [41]. In this study, the Xgboost model built based on the 9 characteristic bands selected by the CARS_SPA algorithm has been proved to be the best model for estimating winter wheat LAI, the accuracy of which (calibration $R^2 = 0.89$; validation $R^2 = 0.89$) is higher than those constructed by Tao et al. [33, 39]. Therefore, Xgboost may be a reliable modeling method, which can be used for UAV remote sensing in combine with wheat LAI modeling and prediction. The findings in this study provide technical support for rapid and nondestructive estimation of winter wheat LAI.

4 Conclusion

Based on different nitrogen treatments and field experiments, hyperspectral images, and LAI at key growth stages of winter wheat were obtained by the UAV platform equipped with hyperspectral imaging sensor. Four algorithms were used to extract characteristic bands related to LAI, and three ML methods were used to construct models for estimating winter wheat LAI. We compared the reliability and accuracy of these models, and found out that the Xgboost model constructed based on 9 characteristic bands

selected by CARS_SPA algorithm exhibited the highest accuracy. In this model, the R^2 , RMSE, and RPD of the calibration set were 0.89, 0.63, and 2.51, respectively, and those of the validation set were 0.89, 0.55, and 2.92, respectively. The combination of CARS_SPA algorithm and Xgboost modeling method can reduce input variables, improve the operation efficiency of the model, and ensure a higher accuracy. Our results provide a reference for the nondestructive and rapid acquisition of winter wheat LAI by UAV remote sensing.

Methods:

Aim: In order to accurately estimate leaf area index (LAI) of winter wheat by using unmanned aerial vehicle (UAV) hyperspectral imagery. **Processes:** In this study, the UAV hyperspectral data, ASD non-imaging hyperspectral data, and LAI measurements of winter wheat at various key growth stages were obtained. Various variables extraction algorithms were used to extract characteristic bands related to LAI. Next, three ML methods (PLSR, SVR, and Xgboost) were employed to construct estimation models of winter wheat LAI based on selected spectral variables. Last, these models were comprehensively compared and a most suitable model for estimating winter wheat LAI was determined. This study provides the methodology and technical support for UAV remote sensing on winter wheat LAI estimation.

Abbreviations

UAV: Unmanned aerial vehicle

LAI: Leaf Area Index

ASD: ASD FieldSpec4 Portable high-resolution spectrometer

FD: First derivative

SPA: Successive projections algorithm

CARS: Competitive adaptive reweighed sampling

CARS_SPA: Competitive adaptive reweighed sampling combined with successive projections algorithm

PLSR: Partial least squares regression

SVR: Support vector machine regression

Xgboost: Extreme gradient boosting

Vis: Vegetation indices

EDF: Exponential decreasing function

R^2 : Coefficient of determination

RMSE: Root mean square error

RPD: Relative percent deviation

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The data sets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: This work was supported by grants from the National Key Research and Development Program of China (2016YFD0300609), the Key Scientific and Technological Projects of Henan Province (192102110012) and Henan Modern Agriculture (Wheat) Research System (S2010-01-G04).

Authors' contributions: Juanjuan Zhang and Tao Cheng conceived and designed the research; Juanjuan Zhang, Tao Cheng, Wei Guo, Xin Xu and Yimin Xie performed the experiments; Juanjuan Zhang and Tao Cheng wrote the manuscript; Juanjuan Zhang, Tao Cheng and Hongbo Qiao analyzed the test results and revised the manuscript; and Xinming Ma, Juanjuan Zhang and Hongbo Qiao reviewed and edited the manuscript.

Acknowledgements: The authors of this research would like to thank the graduate students at the College of Information and Management Science and Ministry of Agriculture in Henan Agricultural University for their continued support of our research.

Authors' information: Affiliations

Henan Agricultural University, Zhengzhou, 450000, China

Juanjuan Zhang, Tao Cheng, Wei Guo, Xin Xu, Xinming Ma, Yimin Xie, Hongbo Qiao

Henan Grain Crops Collaborative Innovation Center, Zhengzhou, 450000, China

Juanjuan Zhang, Tao Cheng, Wei Guo, Xin Xu, Xinming Ma, Yimin Xie, Hongbo Qiao

References

1. Jin X, Yang G, Xu X, et al. Combined multi-temporal optical and radar parameters for estimating LAI and biomass in winter wheat using HJ and RADARSAR-2 data[J]. Remote Sensing, 2015;7(10): 13251-13272. doi: [10.3390/rs71013251](https://doi.org/10.3390/rs71013251).

2. Gower, S.T.,Kucharik, C.J.,Norman, J.M. Direct and indirect estimation of leaf area index, f APAR, and net primary production of terrestrial ecosystems[J]. Remote Sens. Environ,1999; **70**, 29– 51. doi: [10.1016/S0034-4257\(99\)00056-5](https://doi.org/10.1016/S0034-4257(99)00056-5).
3. Laliberte A S, Goforth M A, Steele C M, et al. Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments[J]. Remote Sensing, 2011; 3(11): 2529-2551. doi: [10.3390/rs3112529](https://doi.org/10.3390/rs3112529).
4. Yao X, Wang N, Liu Y, et al. Estimation of wheat LAI at middle to high levels using unmanned aerial vehicle narrowband multispectral imagery[J]. Remote Sensing, 2017; 9(12): 1304. doi: [10.3390/rs9121304](https://doi.org/10.3390/rs9121304).
5. Zhou X, Zheng H B, Xu X Q, et al. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017; 130: 246-255. doi:[10.1016/j.isprsjprs.2017.05.003](https://doi.org/10.1016/j.isprsjprs.2017.05.003).
6. Bendig J, Yu K, Aasen H, et al. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley[J]. International Journal of Applied Earth Observation and Geoinformation, 2015; 39: 79-87. doi: [10.1016/j.jag.2015.02.012](https://doi.org/10.1016/j.jag.2015.02.012).
7. Li Z, Chen Z, Wang L, et al. Area extraction of maize lodging based on remote sensing by small unmanned aerial vehicle[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014; 30(19): 207-213. doi: [3969/j.issn.1002-6819.2014.19.025](https://doi.org/10.1016/j.issn.1002-6819.2014.19.025).
8. Han L, Yang G, Dai H, et al. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data[J]. Plant Methods, 2019; 15(1): 10. doi: [10.3390/rs11111261](https://doi.org/10.3390/rs11111261).
9. Zheng H, Ma J, Zhou M, et al. Enhancing the nitrogen signals of rice canopies across critical growth stages through the integration of textural and spectral information from unmanned aerial vehicle (UAV) multispectral imagery[J]. Remote Sensing, 2020; 12(6): 957. [org/10.3390/rs12060957](https://doi.org/10.3390/rs12060957).
10. Alheit K V, Busemeyer L, Liu W, et al. Multiple-line cross QTL mapping for biomass yield and plant height in triticale (\times Triticosecale Wittmack)[J]. Theoretical and applied genetics, 2014; 127(1): 251-260. doi: [1007/s00122-013-2214-6](https://doi.org/10.1007/s00122-013-2214-6).
11. Fu Z, Jiang J, Gao Y, et al. Wheat growth monitoring and yield estimation based on multi-rotor unmanned aerial vehicle[J]. Remote Sensing, 2020; 12(3): 508. doi: [10.3390/rs12030508](https://doi.org/10.3390/rs12030508).
12. Xie Q,Huang W,Liang D,et al. Leaf area index estimation using vegetation indices derived from airborne hyperspectral images in winter wheat[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014; 7(8): 3586-3594. doi: [1109/JSTARS.2014.2342291](https://doi.org/10.1109/JSTARS.2014.2342291).
13. Liu J, Pattey E, Jégo G. Assessment of vegetation indices for regional crop green LAI estimation from Landsat images over multiple growing seasons[J]. Remote Sensing of Environment, 2012; 123: 347-358. doi: [10.1016/j.rse.2012.04.002](https://doi.org/10.1016/j.rse.2012.04.002).
14. Mutanga O, Skidmore A K. Narrow band vegetation indices overcome the saturation problem in biomass estimation[J]. International journal of remote sensing, 2004; 25(19): 3999-4014. doi: [10.1080/01431160310001654923](https://doi.org/10.1080/01431160310001654923).

15. Liang L, Di L, Zhang L, et al. Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method[J]. *Remote Sensing of Environment*, 2015; 165: 123-134. doi: [10.1016/j.rse.2015.04.032](https://doi.org/10.1016/j.rse.2015.04.032).
16. Gao L, Yang G J, Li C C, et al. Application of an improved method in retrieving leaf area index combined spectral index with PLSR in hyperspectral data generated by unmanned aerial vehicle snapshot camera[J]. *Acta Agronomica Sinica*, 2017; 43(4): 549-557. doi: [10.3724/SP.J.1006.2017.00549](https://doi.org/10.3724/SP.J.1006.2017.00549).
17. Yue J, Feng H, Jin X, et al. A comparison of crop parameters estimation using images from UAV-mounted snapshot hyperspectral sensor and high-definition digital camera[J]. *Remote Sensing*, 2018; 10(7): 1138. doi: [10.3390/rs10071138](https://doi.org/10.3390/rs10071138).
18. Chen Z, Jia K, Xiao C, et al. Leaf Area Index Estimation Algorithm for GF-5 Hyperspectral Data Based on Different Feature Selection and Machine Learning Methods[J]. *Remote Sensing*, 2020; 12(13): 2110. doi: [10.3390/rs12132110](https://doi.org/10.3390/rs12132110).
19. Jia M, Li W, Wang 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. doi: [10.1016/j.compag.2019.104942](https://doi.org/10.1016/j.compag.2019.104942).
20. Khaled A Y, Abd Aziz S, Bejo S K, et al. A comparative study on dimensionality reduction of dielectric spectral data for the classification of basal stem rot (BSR) disease in oil palm[J]. *Computers and Electronics in Agriculture*, 2020; 170: 105288. doi:[10.1016/j.compag.2020.105288](https://doi.org/10.1016/j.compag.2020.105288).
21. Samsudin S H, Shafri H Z M, Hamedianfar A, et al. Spectral feature selection and classification of roofing materials using field spectroscopy data[J]. *Journal of Applied Remote Sensing*, 2015; 9(1): 095079. doi: [10.1117/1.JRS.9.095079](https://doi.org/10.1117/1.JRS.9.095079).
22. Kumar A, Patidar V, Khazanchi D, et al. Optimizing feature selection using particle swarm optimization and utilizing ventral sides of leaves for plant leaf classification[J]. *Procedia Computer Science*, 2016; 89: 324-332. doi: [10.1016/j.procs.2016.06.079](https://doi.org/10.1016/j.procs.2016.06.079).
23. Liu R, Wang J, Yang G, et al. Comparison of ground-based LAI measuring methods on winter wheat[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011; 27(3): 220-224. doi: [3969/j.issn.1002-6819.2011.03.042](https://doi.org/10.3969/j.issn.1002-6819.2011.03.042).
24. Zhang J, 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. doi: [10.3390/s8021321](https://doi.org/10.3390/s8021321).
25. Jiang H, Zhang H, Chen Q, et al. Identification of solid state fermentation degree with FT-NIR spectroscopy: comparison of wavelength variable selection methods of CARS and SCARS[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2015; 149: 1-7. doi: [10.1016/j.saa.2015.04.024](https://doi.org/10.1016/j.saa.2015.04.024).
26. Fan S, Huang W, Guo Z, et al. Prediction of soluble solids content and firmness of pears using hyperspectral reflectance imaging[J]. *Food analytical methods*, 2015; 8(8): 1936-1946. doi: [10.1007/s12161-014-0079-1](https://doi.org/10.1007/s12161-014-0079-1).

27. Mousa S R, Ishak S. An extreme gradient boosting algorithm for freeway short-term travel time prediction using basic safety messages of connected vehicles[R]. 2017.
28. Can Y. Road network extraction from remote sensing images based on XGBoost[J]. *Microcomputer & Its Applications*, 2017; 2017(24): 9. doi:19358/j.issn.1674-7720.2017.24.009.
29. Cheng Xue, He Bingyan, Huang Yaohuan, et al. Estimation of Corn Leaf Area Index based on UAV Hyperspectral Image. *Remote Sensing Technology and Application*, 2019; 34(4): 775-784. (in Chinese) doi: 10.11873/j.issn.1004-0323.2019.4.0775.
30. Gao L, Yang G, Yu H, et al. Retrieving winter wheat leaf area index based on unmanned aerial vehicle hyperspectral remote sensing[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016; 32(22): 113-120. doi: **11975/j.issn.1002-6819.2016.22.016**.
31. Xie Q, Huang W, Liang D, et al. Leaf area index estimation using vegetation indices derived from airborne hyperspectral images in winter wheat[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014; 7(8): 3586-3594. doi: [10.1109/jstars.2014.2342291](https://doi.org/10.1109/jstars.2014.2342291) .
32. Tao H, Feng H, Xu L, et al. Estimation of Crop Growth Parameters Using UAV-Based Hyperspectral Remote Sensing Data[J]. *Sensors*, 2020; 20(5): 1296. doi: [10.3390/s20051296](https://doi.org/10.3390/s20051296).
33. Chen R, Feng H, Yang F, et al. Estimation of Leaf Area Index of Winter Wheat Based on Hyperspectral Data of Unmanned Aerial Vehicles[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019; 1785-1788. doi: [10.1109/igarss.2019.8897844](https://doi.org/10.1109/igarss.2019.8897844).
34. Darvishzadeh R, Skidmore A, Atzberger C, et al. Estimation of vegetation LAI from hyperspectral reflectance data: Effects of soil type and plant architecture[J]. *International journal of applied Earth observation and geoinformation*, 2008; 10(3): 358-373. doi: [10.1016/j.jag.2008.02.005](https://doi.org/10.1016/j.jag.2008.02.005).
35. Sims D A, Gamon J A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages[J]. *Remote sensing of environment*, 2002; 81(2-3): 337-354. doi: [10.1016/S0034-4257\(02\)00010-X](https://doi.org/10.1016/S0034-4257(02)00010-X).
36. Liang D, Guan Q, Huang W, et al. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013; 29(7): 117-123. doi: **1002-6819 (2013) -07-0117-07**.
37. Wang F M, Huang J F, Tang Y L, et al. New vegetation index and its application in estimating leaf area index of rice[J]. *Rice Science*, 2007; 14(3): 195-203. doi: [10.1016/S1672-6308\(07\)60027-4](https://doi.org/10.1016/S1672-6308(07)60027-4).
38. Houborg R, McCabe M F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018; 135: 173-188. doi: [10.1016/j.isprsjprs.2017.10.004](https://doi.org/10.1016/j.isprsjprs.2017.10.004).
39. TAO Huilin, FENG Haikuan, YANG Guijun, YANG Xiaodong, LIU Mingxing, LIU Shuaibing. Leaf Area Index Estimation of Winter Wheat Based on UAV Imaging Hyperspectral Imagery [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020; 51(01): 176-187. (in Chinese) doi: [6041/j.issn.1000-1298.2020.01.019](https://doi.org/10.1000-1298.2020.01.019).
40. Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.

41. Dhaliwal S S, Nahid A A, Abbas R. Effective intrusion detection system using XGBoost[J]. Information, 2018; 9(7): 149. doi: 10.3390/info9070149.

Figures

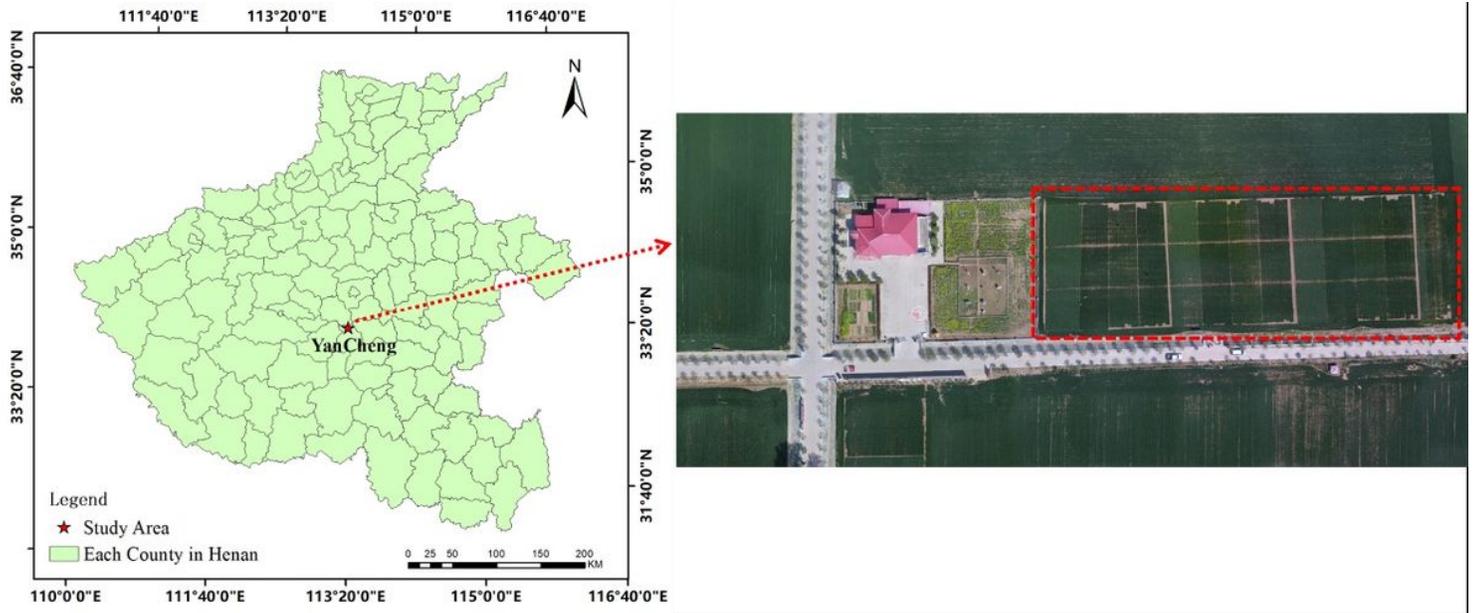


Figure 1

Study area location. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

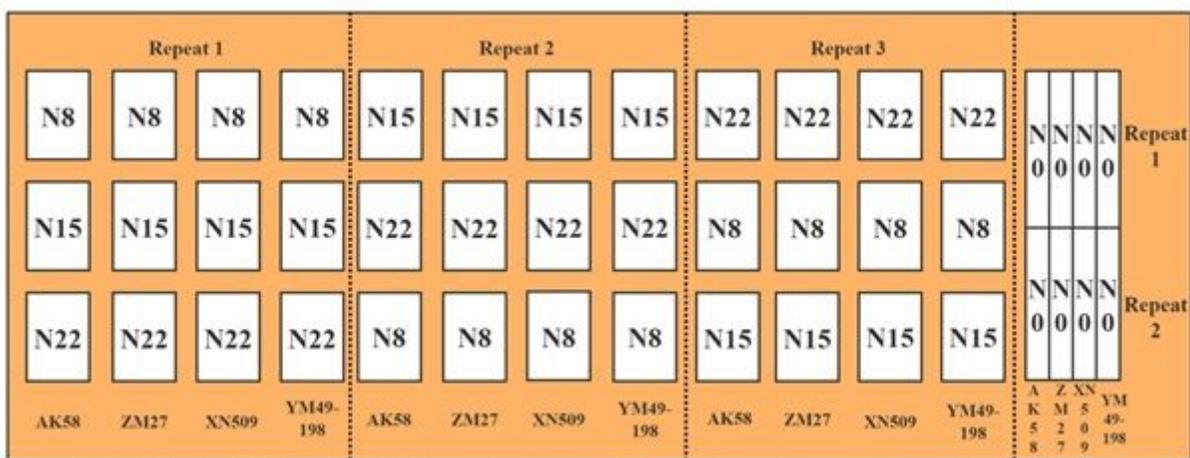


Figure 2

Experimental design. Note: AK58, ZM27, XN509, YM49-198 represented wheat cultivars, AK58 was Aikang 58, ZM27 was Zhoumai 27, XN509 was Xinong509, YM49-198 was Yumai49-198. N0, N8, N15, N22

represented nitrogen treatments of 0, 120, 225, and 330 kg.hm⁻², respectively.

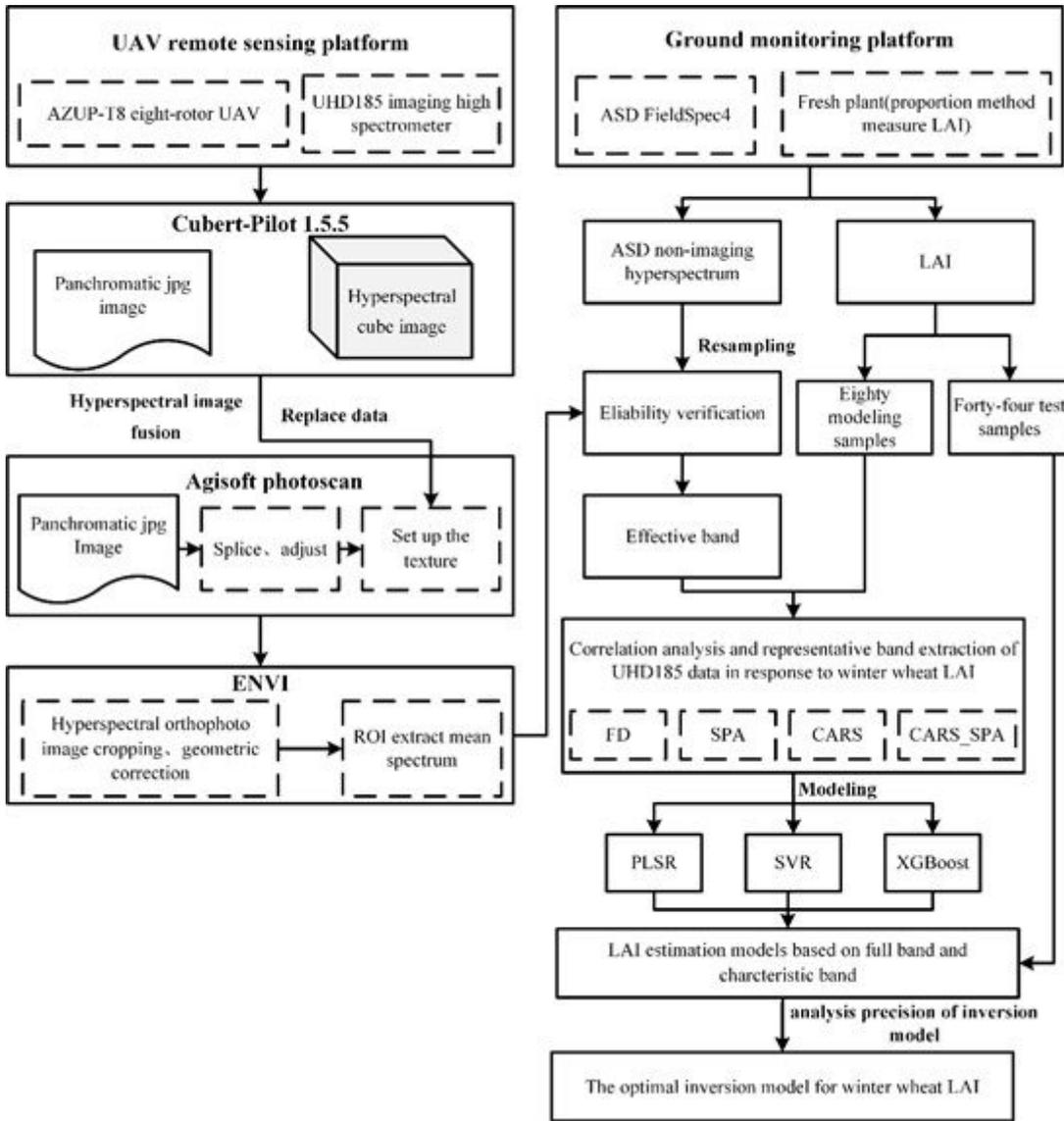


Figure 3

Flow chart of winter wheat LAI estimation modelling method

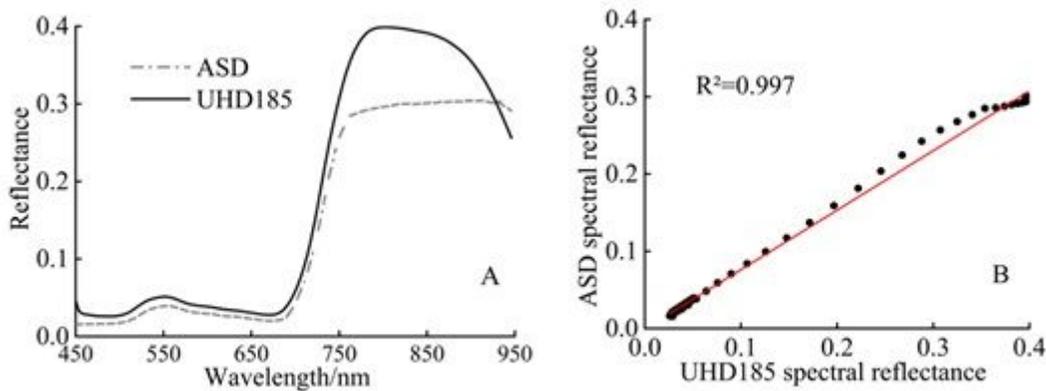


Figure 4

Reliability verification of UHD185 hyperspectral data. Note: Comparison of spectral reflectance curves of UHD185 and resampled ASD(A), Correlation between UHD185 and resampled ASD spectral reflectance(B).

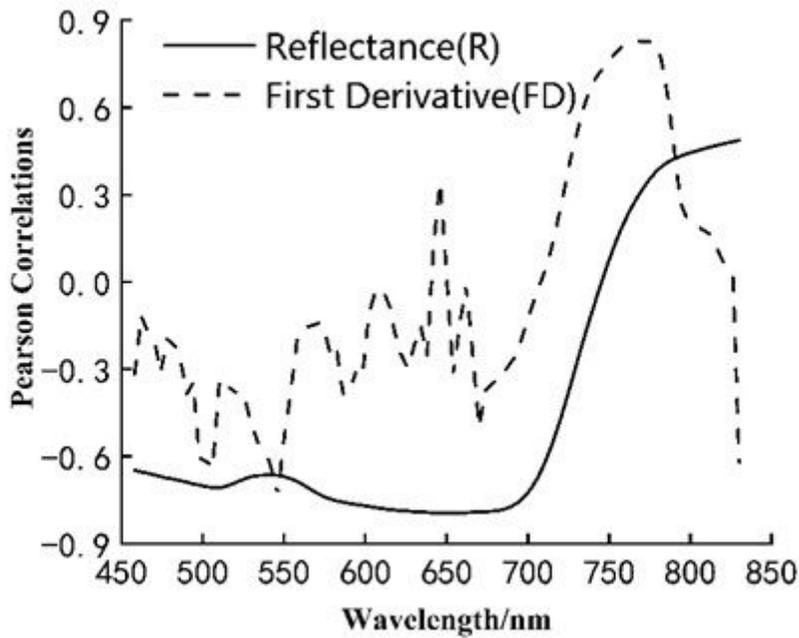


Figure 5

Correlation between UHD185 hyperspectral data and wheat LAI

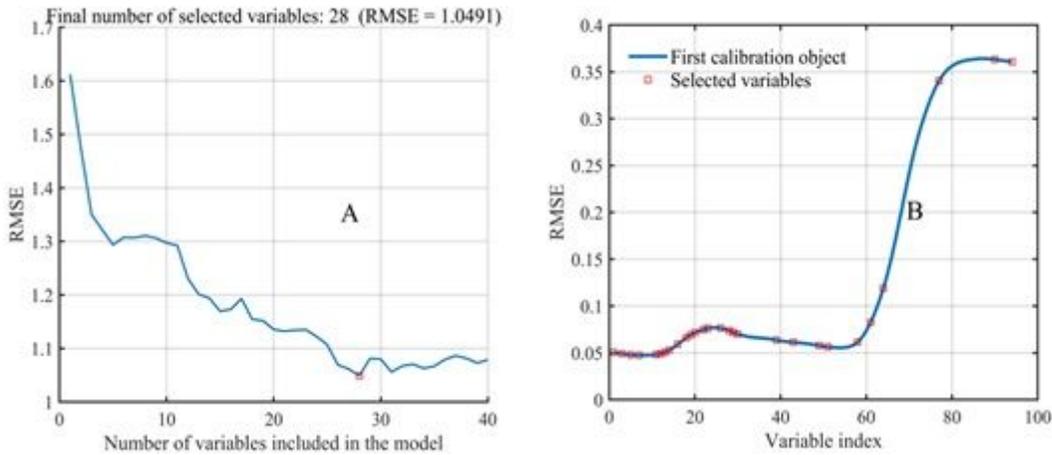


Figure 6

SPA Variable filtering process. Note: Number of variables(A), variable indices(B).

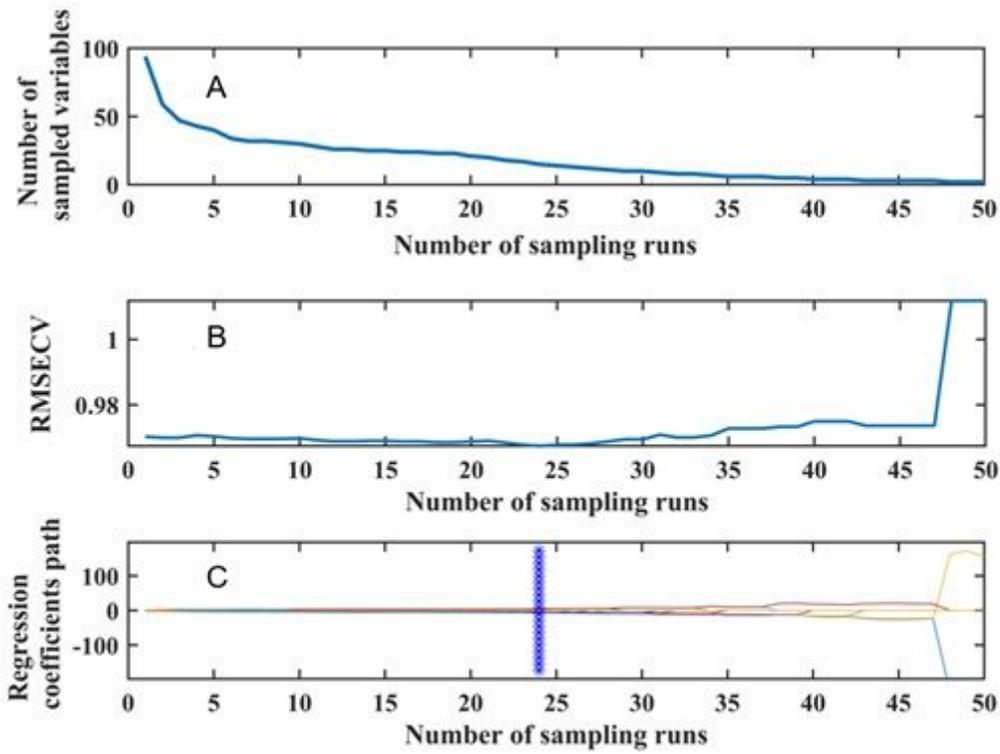


Figure 7

Selection of variables by CARS. Note: Variation of variables(A), 10-fold RMSE values(B), regression coefficient of variables(C). The number of variables corresponding to the minimum RMSE was marked with an asterisk.

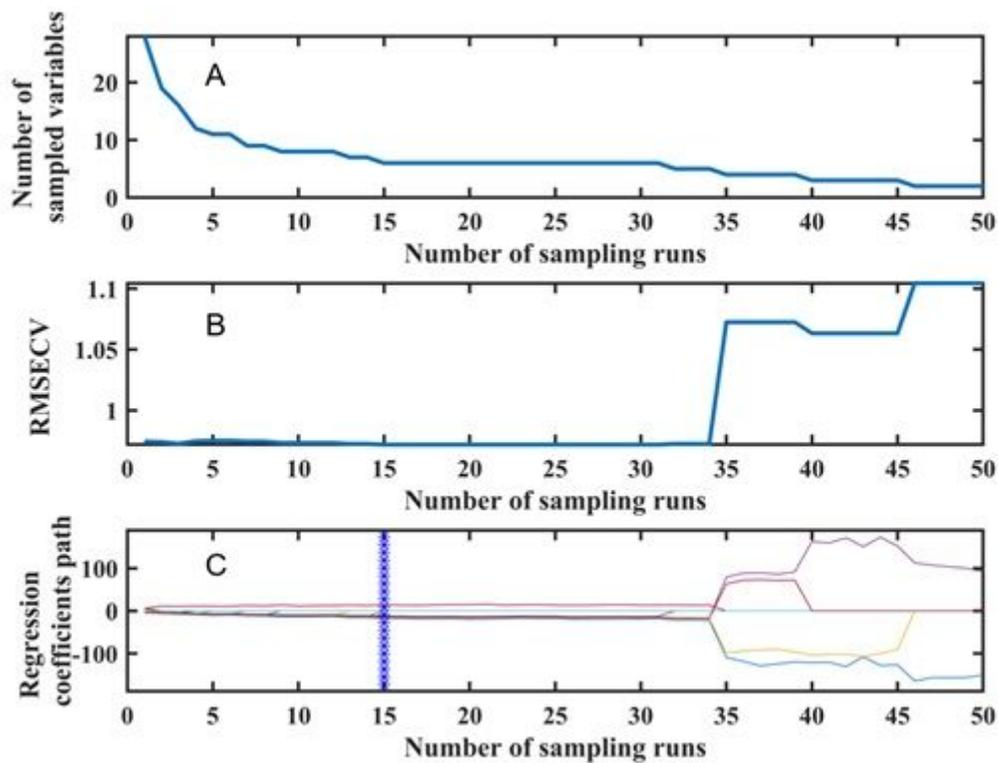


Figure 8

Variable selection process by CARS_SPA. Note: Variation of variables(A), 10-fold RMSE values(B), regression coefficient of variables(C). The number of variables corresponding to the minimum RMSE was marked with an asterisk.

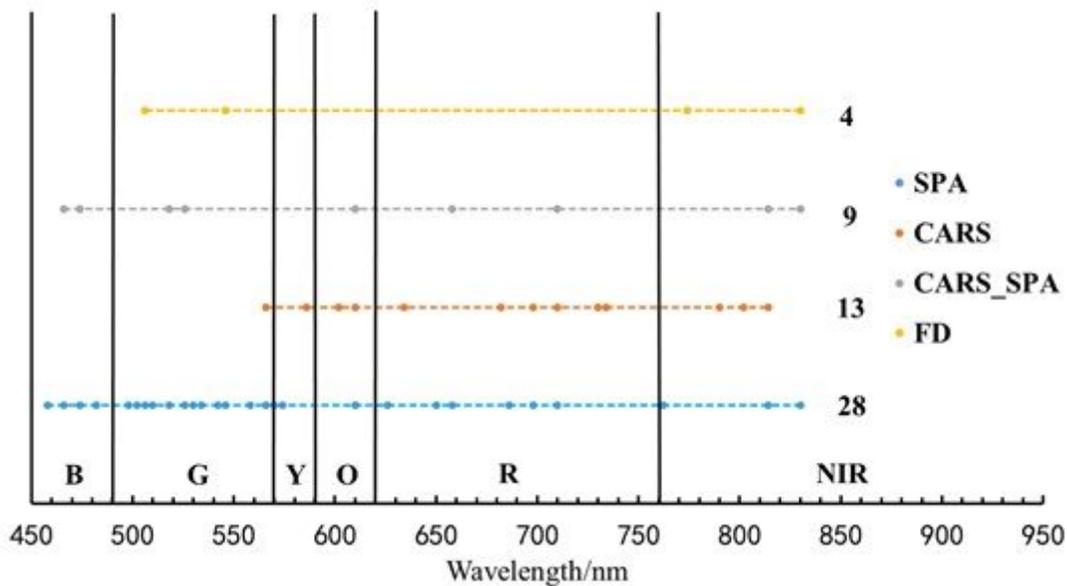


Figure 9

Distribution of characteristic bands selected by using different variable extraction algorithms. Note: SPA: Successive projections algorithm, CARS:Competitive adaptive reweighed sampling, CARS_SPA: competitive adaptive reweighed sampling combined with successive projections algorithm, FD: First derivative, Reflectance: The original reflectivity curve. B: Blue, G: Green, Y: Yellow, O: Orange, R: Red, NIR: Near-infrared.

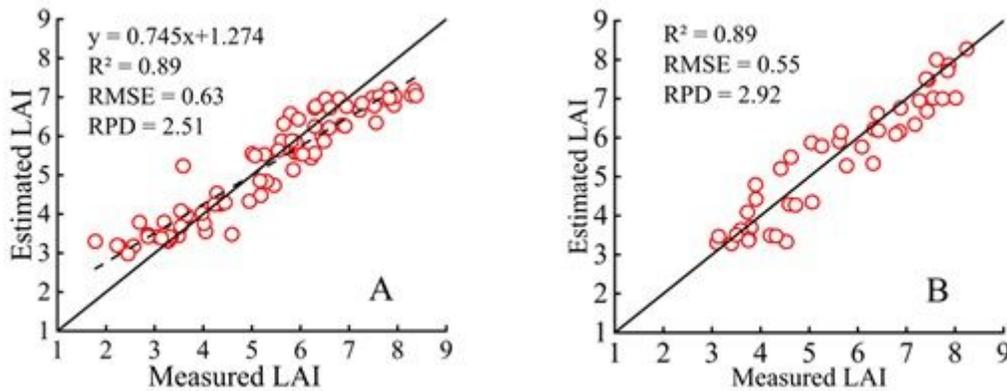


Figure 10

Calibration and validation results of best-performed wheat LAI estimation model (LAI_CARS_SPA_Xgboost). Note: Scatter diagram of calibration results(A), scatter diagram of Validation results(B).