

SPAD inversion of summer maize combined with multi-source remote sensing data

Fangjiang Pan^{1,2}, Wenhua Li^{1,2}, Yubin Lan^{1,2}, Xuguang Liu^{1,2}, Jianchi Miao^{1,2},
Xiao Xiao^{1,2}, Haiyu Xu^{1,2}, Liquan Lu^{2,3}, Jing Zhao^{1,2*}

(1. School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, China;

2. Shandong University of Technology Sub-center of National Center for International Collaboration Research on Precision Agricultural Aviation Pesticide Spraying Technology, Zibo 255000, China;

3. School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China)

Abstract: The chlorophyll content is an important indicator of corn growth and yield. In order to improve the prediction accuracy of chlorophyll content, this study combines ground hyperspectral characteristic parameters (original spectral characteristics, first-order differential, characteristic spectral position), vegetation index calculated by multispectral, and effective plant height (Canopy Height Model, CHM) of crops, etc. Through correlation analysis of sensitive characteristics of chlorophyll content, the study uses multiple linear regression (MLR), partial least squares regression (PLSR), classification and regression tree regression (CART), and random forest (RF) to construct a summer maize SPAD inversion model. Then, the accuracy of the model was evaluated through the root mean square error (RMSE) and coefficient of determination (R^2). The results show that the position of the red edge and the first-order differential values within the red edge D_r , CHM, SAVI, NDVI, RDVI, GNDVI, RVI, and DVI are significantly correlated with SPAD; the MLR model under a single data source is the best, the model's R^2 is 0.8281, RMSE is 2.136; the RF model under multi-source data is the best. The model's R^2 and RMSE are 0.9114 and 2.3955 respectively. The accuracy of the SPAD inversion model constructed based on multi-source data is better than that of a single data source. This study shows that the random forest model based on multi-source data can invert the SPAD of summer maize better. This method can provide theoretical support for summer maize growth monitoring and fine fertilization management.

Keywords: Multi-source remote sensing, summer corn, Unmanned Aerial Vehicle, chlorophyll, machine learning

DOI: 10.33440/j.ijpaa.20210402.174

Citation: Pan F J, Li W H, Lan Y B, Liu X G, Miao J C, Xiao X, Xu H Y, Lu L Q, Zhao J. SPAD inversion of summer maize combined with multi-source remote sensing data. *Int J Precis Agric Aviat*, 2021; 4(2): 45–52.

1 Introduction

As one of the main raw materials of national food and feed, the yield and quality of maize is very important to national food security. Chlorophyll, as the main pigment for photosynthesis, plays a core role in light absorption, energy conversion, and organic synthesis of crops, and is an important indicator for monitoring crop growth and yield^[1]. Studies had shown that crop spectral reflectance is closely related to the change of chlorophyll content^[2]. Therefore, it is one of the current research hotspots to explore the application of multi-source spectral remote sensing

technology to invert chlorophyll content and improve the inversion accuracy.

Remote sensing measurement of chlorophyll content is mainly divided into satellite remote sensing measurement and low-altitude remote sensing measurement. Although satellite remote sensing can obtain spectral information of crops in a wide range, it has problems such as low accuracy, poor timeliness and great difficulty in obtaining data, which make it impossible for satellite data to be applied to accurate field management^[3,4]. With its strong flexibility and good timeliness, UAV provides a new collection method for low-altitude remote sensing data acquisition^[5,6]. At present, many scholars have obtained crop canopy reflectance by spectral remote sensing and used the vegetation index method to invert chlorophyll content. Su et al.^[7] combined Sentinel-2 and MODIS data. They used the joint probability distribution method to retrieve LAI and chlorophyll content, and accurately obtained LAI and chlorophyll content of maize canopy through PROSAIL model and fusion Kalman filter. Su Wei et al.^[8] used an unmanned aerial vehicle to carry a multi-spectral camera to obtain the corn canopy image and built a chlorophyll inversion model based on vegetation index. The results showed that the optimal inversion resolution of the image was 0.1~0.3 m, and the inversion effect of the normalized red edge green index was the best. Hassanijalilian et al.^[9] extracted the dark green vegetation index from the collected soybean visible light images and found that the inversion accuracy of the support vector machine model was the highest when constructing the chlorophyll inversion model using

Received date: 2021-09-20 **Accepted date:** 2021-10-20

Biographies: **Fangjiang Pan**, Postgraduate student, research interests: remote sensing, mechanical design and computer programming. Email: 181831794@139.com; **Wenhua LI**, Postgraduate student, research interests: Unmanned aerial remote sensing, Email: 1446158794@qq.com; **Yubin Lan**, PhD, professor, Director, research interests: precision agricultural aviation application, Email: ylan@scau.edu.cn; **Xuguang Liu**, Postgraduate student research interests: Unmanned aerial remote sensing, Email: 1215151729@qq.com; **Jianchi Miao**, Postgraduate student, research interests: remote sensing, Email: 1393746086@qq.com; **Xiao Xiao**, Postgraduate student, research interests: remote sensing, Email: 1085751745@qq.com; **Haiyu Xu**, Postgraduate student research interests: Unmanned aerial remote sensing, Email: 903837514@qq.com; **Liquan Lu**, PhD, Associate Professor, research interest: agricultural aviation technology and fluid transmission technology research. Email: luliquanustb@163.com.

* **Corresponding author:** **Jing Zhao**, Associate professor, research interests: remote sensing, Mailing Address: Shandong University of Technology West Campus. Email: zbceozj@163.com

the machine learning method. Wen Yao et al.^[10] obtained multi-spectral canopy data of maize seedling stage by UAV and constructed the chlorophyll inversion model by vegetation index method combined with least squares-support vector machine model. The predicted accuracy of this model reached 0.56, which was a better-predicted SPAD change in the maize seedling stage. It is not difficult to see that UAV remote sensing is mainly based on visible light and multi-spectrum. Thus, more spectral information cannot be obtained, which makes it impossible to further improve the accuracy of SPAD inversion based on UAV remote sensing data.

Hyperspectral data can provide more detailed information and provide possibility for further improvement of accuracy of chlorophyll inversion. At present, many scholars have explored the use of hyperspectral data to invert chlorophyll content. An et al.^[11] carried out correlation analysis on hyperspectral data and used a gradient lifting tree algorithm to build an inversion model of rice chlorophyll content. Mao Zihui et al.^[12] pointed out that the vegetation index including the red-edge band and the green band was strongly correlated with chlorophyll content, and the red-edge band was more sensitive to SPAD changes. Liu Haojie et al.^[13] also showed that the two-exponential chlorophyll inversion model based on the red-edge band was the best. Sun Yangyang et al.^[14] conducted principal component analysis on maize hyperspectral data and found that the band such as red edge, near-infrared and green peak were more sensitive to SPAD, and the multivariate regression model constructed by this method obtained the best inversion accuracy.

At present, UAV multi-spectral remote sensing can quickly obtain field-scale information, but it lacks more spectral details. Ground hyperspectral can obtain more spectral details, but the scope of action is small. In addition, most studies are still dominated by a single data source, and there is a lack of SPAD inversion research under the condition of multi-source data. Therefore, this research is based on UAV multispectral data and ground hyperspectral data for SPAD inversion.

In this study, a variety of vegetation indices were calculated based on UAV multispectral data, and chlorophyll sensitive indices including red and green bands were selected as model parameters. First-order differentiation was performed on the hyperspectral data to screen and extract spectral sensitive features of original spectrum and the first derivative of the spectrum. The ground DSM model was built by using the UAV data, and canopy height model (CHM) was developed by subtracting the bare land values from the DSM data. This research used the above-mentioned spectral characteristics, effective plant height and other multi-source data as model parameters, using Python programming and Minitab to perform multiple linear regression (MLR), partial least squares regression (PLSR), classification regression tree (Classification and regression tree, CART) and random forest (Random forest, RF) regression to construct the model of summer maize SPAD inversion. By using the root mean square error and coefficient of determination to assess the accuracy of the model. It provides better theoretical support for summer maize growth monitoring and management that best inversion model of summer maize SPAD based on multi-source data.

2 Materials and methods

2.1 Experimental designs

The experimental area is located in the ecological unmanned farm of the Shandong University of Technology, Zhutai Town,

Zibo City, Shandong Province (36°57'15" N, 118°12'50" E), with an altitude of about 27 m. It has a temperate semi-humid continental monsoon climate, suitable for summer corn, Wheat growth. The total area of the design plot is 50 m×150 m. The 1 m×1 m sample plot was designed by using the five-point sampling method. Two representative plants were selected from each plot for the actual measurement of chlorophyll content. There 30 data samples were obtained of three keys growth stages. The samples are divided into the training set and testing set according to 2:1, and data is drawn by the random sampling method. Finally, there are 20 samples in the training set and 10 samples in the testing set.

2.2 Multi-spectral acquisitions of UAV

The multispectral data on July 20 (flaring stage), August 9 (tasselling stage) and August 21 (seed setting stage) were obtained by using a DJI M210 which equipped with Ysense MS600 multispectral camera. Data was obtained from 10:00 to noon of that day. The weather was fine. During aerial photography, the UAV flew at a height of 70 m, flight speed of 4 m/s, a heading overlap of 80%, and a side overlap of 70%. MS600 camera has a single-band channel and spectral resolution of 450 nm@35 nm, 555 nm@25 nm, 660 nm@22.5 nm, 720 nm@10 nm, 840 nm@30 nm and 940 nm@35 nm (Figure 1). The pixel resolution is 1280×960 and the storage format is TIFF. Two sets of standard whiteboard images were obtained before and after each data acquisition flight. Multi-spectral reflectivity correction was carried out for each period of data through Ysense-Ref, and orthophotos correction and image mosaicing were carried out for the collected images using Pix4D Mapper (Pix4D, Switzerland), a professional data processing software for UAV aerial photographs. Digital Surface Model (DSM) images and single-band orthophotos with a ground resolution of 4.45 cm per pixel were obtained^[15]. Finally, the ArcGIS calibration toolset is used to register images in different periods to ensure the consistency of image positions in different periods.



Figure 1 DJI M210 and MS 600 camera

2.3 Ground measurement data collection

PSR1100-f (Spectral Evolution, USA) is a non-imaging hyperspectral measuring instrument with a measurement range of 320~1100 nm and a sampling interval of 1.5 nm. Through its format conversion tool (SED to CSV Converter), the data can be resampled to 1nm resolution and stored directly as a .csv files. The measuring instrument was used for hyperspectral collection of corn at the sampling point, with a collection time of 11:00 to 14:00. Before collection, the standard whiteboard was used for correction, and the influence of the dark current was removed. During collection, the probe was 0.5 m away from the crop canopy and perpendicular to the ground, and each sample point was collected 5 times (Figure 2). The corn plants are relatively sparse in the trumpet period during the collection of hyperspectral data. The

healthy corn plants in the sampling points were selected for measurement three times, and the mean value was taken as the final spectral data of the sample points in that period. At the same time, the GPS location of the collection point, the total number of leaves and the total number of plants in the sample area were recorded.

During the grouting stage and the milk maturity stage, there are jagged fluctuations in the hyperspectral data. That means the hyperspectral data has noise interference, so hyperspectral filtering and smoothing are required. Local weighted regression (Lowess) can remove the noise points according to the fluctuation range of the data. By judging the distance between each data and the line, the true abnormal value is eliminated and a smooth curve conforming to fit the overall trend. After many tests, we found that smoothing the hyperspectral data within the range of 0.03 can better filter out the data noise points. Therefore, this experiment adopts the Lowess algorithm to flatten the hyperspectral data, with the smoothness is 0.03 and the number of iterations is 2. All the data are smoothed and filtered, and the obtained smoothed spectrum was used as the final spectrum data.

SPAD-502Plus was used for the actual measurement of chlorophyll content. During the measurement, two representative plants were selected from each sample area, which was divided into three layers according to 4:3:3 of the total number of leaves. Five points of a uniform leaf were selected from each layer for measurement. At the same time, the GPS position of the collection point was recorded, and the total number of leaves and plants in the sample area were recorded.

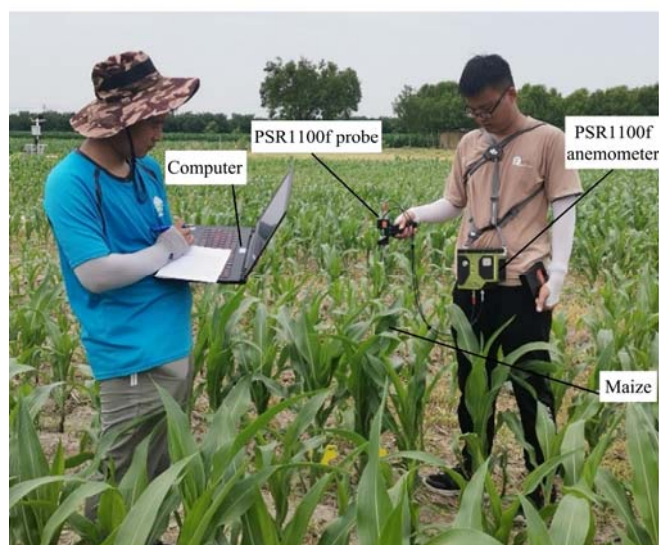


Figure 2 The ground measurement data collection of maize

2.4 Selection of effective plant height and spectral features

The maize height was obtained by using the method of Niu Qinglin et al.^[16]. In ArcGIS, 20 bare field data of DSM images in the trumpet stage were firstly obtained, and the mean value was taken as the true value of a bare field. Then, the bare field value was subtracted from the plant DSM value to obtain the final plant effective height (CHM). Gao Mingyang et al.^[17] extracted the plant height of wheat through this method, which proved the reliability of this method.

Previous studies on the relationship between hyperspectral and maize SPAD indicated that the hyperspectral red edge position λ_r , the highest point of near-infrared reflectance position λ_{nir} , the yellow edge position λ_y , the first-order differential value Dy within the yellow edge, the first-order differential value Db within the

blue edge, and the first-order differential value Dr within the red edge were all highly correlated with SPAD^[18]. Studies have shown that crops show obvious green peak and red valley features in the hyperspectral spectrum^[19]. Therefore, in this study, green peak reflectance Rg, red valley reflectance Rr and the above spectral position features are taken as spectral characteristic parameters to screen SPAD sensitive features and build inversion models.

Many kinds of research have shown that the chlorophyll in a red band (R), green band (G) and near-infrared band (NIR) has a good response^[19]. Therefore, this study chooses six good vegetation indexes between chlorophyll and multispectral vegetation index (Table 1), which contain R, G and NIR. This study used ArcGIS map algebra tool to calculate each index, and combined the measured points in GPS location information. The vegetation index of each sampling point was obtained for chlorophyll correlation analysis and inversion model construction.

2.5 SPAD inversion model construction and evaluation method

Multivariate data often have collinearity and correlation, so many studies mostly use the multiple linear regression model and partial least squares regression model to describe the relationship between each variable and the target variable. A classification regression tree is a decision tree algorithm optimized based on ID3 and C4.5. In CART regression modelling, its nodes are divided into binary trees based on the principle of minimum variance of samples, and the best division points and optimal output values are searched through recursion continuously, and the pruning method is adopted to reduce model overfitting^[20,21]. The random forest algorithm is based on the decision tree. Through the various decision to vote or unrelated weak joint, the results of a decision tree were obtained as a model of the final result^[22,23]. This method is often used for data classification or regression, which has good adaptability, fast training speed and data is not easy to have the advantages of the fitting. Through Minitab and Python programming, based on multiple linear regression, partial least square regression, classified regression tree and random forest method, the 30 samples obtained were divided into the training set and testing set according to 2:1, and the SPAD inversion model of summer corn was constructed based on that. The accuracy of the model was evaluated by the root mean square error (RMSE) and the coefficient of determination (R^2).

Table 1 Vegetation index selection and calculation formulas

Vegetation index	Name	Equation
NDVI	Normalized difference vegetation index	$(R_{nir} - R_{red}) / (R_{nir} + R_{red})$
GNDVI	Green normalized difference vegetation index	$(R_{nir} - R_{green}) / (R_{nir} + R_{green})$
DVI	Difference vegetation index	$R_{nir} - R_{red}$
RVI	Ratio vegetation index	R_{nir} / R_{red}
SAVI	Soil-adjusted vegetation index	$\frac{(R_{nir} - R_{red})(1 + L)}{(R_{nir} + R_{red} + L)}$, $L=0.5$
RDVI	Renormalized difference vegetation index	$\sqrt{NDVI \cdot DVI}$

3 Results and analysis

3.1 Hyperspectral feature selection

By analyzing the spectral mean values of each period, it can be seen that the hyperspectral characteristics of maize are typical of green plants. There are obvious green peaks (540-560 nm),

red valleys (670-685 nm) and near-infrared (740-900 nm) high reflectance platforms. The chlorophyll content increases with the growth period. And the spectral reflectance also shows an increasing trend (Figure 3a). According to the analysis of the first derivative of the original spectral mean (Figure 3b), the spectral derivative presents a peak value at the first derivative (515-530 nm) within the blue edge and the first derivative (720-750 nm) within the red edge, and the derivative curve presents a valley value at the first derivative within the yellow edge. According to the correlation curve between the original spectral mean value and SPAD (Figure 3c), the yellow edge position (570-580 nm), the red edge position (680-700 nm) and the highest near-infrared reflectance position (870-900 nm) of the curve, show a high correlation with SPAD. In Figure 3d, the first derivative of spectral mean and SPAD also show a high correlation at the blue edge, yellow edge, red edge and near-infrared position. Therefore, in this study, the hyperspectral red edge position λ_r , the highest point of near-infrared reflectance λ_{nir} , the yellow edge position λ_y , the first-order differential value Dy in the yellow edge, the first-order differential value Db in the

blue edge, the first-order differential value Dr in the red edge, the green peak reflectance R_g and the red valley reflectance R_r were taken as SPAD sensitive parameters for correlation analysis and model construction.

Further correlation analysis was conducted between the acquired hyperspectral features and the measured SPAD (Table 2). It was found that, among the eight kinds of hyperspectral features, only the position of the red edge λ_r and the first-order differential value Dr in the red edge, were significantly correlated with summer corn SPAD, with correlations of 0.424 and 0.653, respectively. Other spectral features had poor correlations. The analysis found that with the change of growth period, the maize growth healthily. The red band reflectivity and the red edge first-order differential value of the leaves both increase and the “red edge” redshift phenomenon appears. That means the maize chlorophyll content and spectral red edge parameters exist in significant correlation. The results of this study are relatively similar to Li yuanyuan et al.^[18] Therefore, the position of the red edge and the first-order differential value Dr in the red edge were finally selected as the sensitive spectral features to build the model.

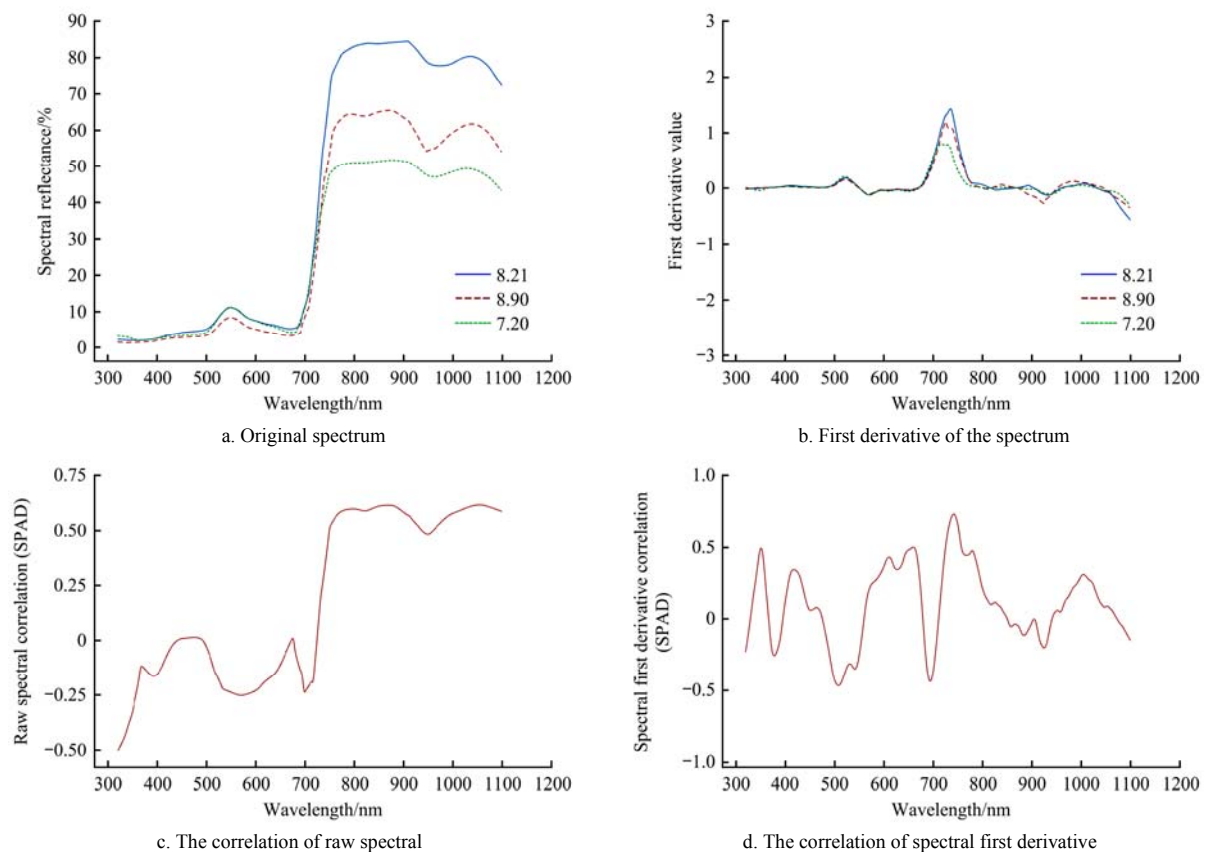


Figure 3 The correlation of chlorophyll between spectral first derivative and raw spectral

Table 2 Hyperspectral feature correlation analysis

Hyperspectral features	Correlation (SPAD)
Red edge position λ_r	0.424*
The highest point of near-infrared reflectance λ_{nir}	0.008
Yellow edge position λ_y	-0.139
The first-order differential value Dy in the yellow edge	0.203
The first-order differential value Db in the blue edge	-0.214
The first-order differential value Dr in the red edge	0.653**
Green peak reflectance R_g	-0.22
Red valley reflectance R_r	-0.016

Note: ** Significant correlation at the level of 0.01 (double-tailed).

* Significant correlation at the level of 0.05 (double-tailed). The same as below.

3.2 Correlation analysis of vegetation index and CHM

The UAV multi-spectral vegetation index and effective plant height of the sampling points were obtained from the ground GPS information as the feature data. Correlation analysis of the selected feature data and summer maize SPAD was conducted (Table 3). It was shown that all parameters were highly correlated with SPAD, and the correlation of CHM, SAVI, NDVI, RDVI and GNDVI was higher than 0.7. The highest was 0.786 for CHM. The correlation between RVI, DVI and summer maize SPAD was greater than 0.65, and all parameters were significantly correlated with summer maize SPAD. Tiahs^[24] research showed that there is a good correlation between chlorophyll and vegetation indices such as DNVI. While Zhang Xuezhi et al.^[25] also pointed out that

NDVI and RVI had a good correlation with maize leaf pigment content. Therefore, all the six vegetation indices and CHM were taken as parameters to construct the inversion model.

Table 3 The correlation of SPAD between vegetation index and effective plant height

Characteristic Parameters	Correlation (SAPD)
NDVI	0.772**
DVI	0.691**
RVI	0.66**
GNDVI	0.742**
SAVI	0.772**
RDVI	0.752**
CHM	0.786**

3.3 Results of the summer maize SPAD inversion model

In this study, multiple linear regression, partial least square regression, classified regression tree and random forest were selected to construct a summer maize SPAD inversion model, which was based on hyperspectral sensitive features, vegetation index, effective plant height and all bands. The R^2 and RMSE were used to evaluate the accuracy of the model. By analyzing the results of each model (Table 4), it can be seen that the inversion model based only on hyperspectral sensitive features have the lowest accuracy, among which the linear model has the lowest accuracy. The R^2 of the training set model is 0.4233, the R^2 of the testing set model is 0.5613, the RMSE is 3.9422, and the highest RMSE is 4.3919 for the CART model. Comparing the RMSE of each model with the model's R^2 , it was found that the errors of each model were large. Therefore, the inversion model based on hyperspectral sensitive features could not perform a good inversion of summer maize's SPAD.

When the vegetation index and effective plant height were used as the input parameters, the accuracy of the random forest on the training set was good, but there was a big difference between the model's R^2 on the training set and testing set, so the model's stability was poor and it could not be used as the final inversion model. The R^2 of partial least squares, CART and multiple linear regression models on the training set and the testing set all exceed 0.7, and the difference between them is small. The R^2 of the multiple linear regression model is 0.8281, which is higher than the CART model and slightly lower than the PLSR model. The RMSE is 2.136, which is lower than the CART and PLSR model. Therefore, the multiple linear regression model is the best inversion

model under this parameter. When all the parameters were taken as the model input, all the models showed higher model's R^2 and lower RMSE. Among them, the R^2 of the random forest model in training set and testing set were better, which were 0.9114 and 0.8104 respectively, and the RMSE of the model was 2.3955. By comparing all the models, it can be seen that the random forest model is the best inversion model under this parameter.

Analyzing the models respectively, we found that the model based only on hyperspectral features in the same model has the lowest inversion accuracy. Compared with the former, the inversion accuracy of the SPAD model has been improved, which is based on vegetation index and plant height. Among them, the multilinear model and the partial least square model are the most obvious. On the training set, the model's accuracy has increased about 40%, and the accuracy of the testing set has improved by about 30%. The RMSE both decreased by more than 1.7. On the training set, the accuracy of the classification regression tree and random forest model increased by about 10%, and the accuracy of the testing set increased by 20%-40%. That means the vegetation index and effective plant height are at a certain level. To a certain extent, it can improve the accuracy of the model.

Analyzing the SPAD inversion model based on multi-source features, we found that except the random forest and classification regression tree, the inversion accuracy of the other two models both increased by 2% to 4% on the training set, and the RMSE decreased by about 0.2. On the training set, the inversion accuracy of the random forest model decreased slightly, but the accuracy of the test set increased by more than 10%, and the RMSE decreased by about 1.2. There may be two reasons. The one is that the training set has fewer data. The other is that there is a certain overfitting phenomenon when only modelling based on the vegetation index. That makes the training set model more accurate, and the testing set model's accuracy is lower. However, multi-source data increases the sensitivity of model recognition, which further improves the accuracy of the model. The overall correlation between hyperspectral features and SPAD is lower than the vegetation index. When the classification regression tree is initially classified, the model uses vegetation index as the major classification feature. So the model inversion accuracy under multi-source data has not been further improved. The overall analysis can be considered that multi-source features are beneficial to the further improvement of model accuracy.

Table 4 Accuracy of each SPAD inversion model under different parameters

Model	Input value	Training set R^2	Testing set R^2	RMSE
Multiple linear regression	$\lambda r, Dr$	0.4233	0.5613	3.9422
	NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM	0.8281	0.8764	2.136
	$\lambda r, Dr, NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM$	0.8667	0.8606	1.9917
Partial least squares regression	$\lambda r, Dr$	0.4291	0.5085	3.9936
	NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM	0.8434	0.7824	2.2779
	$\lambda r, Dr, NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM$	0.8666	0.8605	1.9917
Classification and regression tree	$\lambda r, Dr$	0.6104	0.345	4.3919
	NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM	0.7316	0.708	2.9327
	$\lambda r, Dr, NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM$	0.7316	0.708	2.9327
Random forest	$\lambda r, Dr$	0.8295	0.454	3.8122
	NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM	0.9292	0.6837	3.6006
	$\lambda r, Dr, NDVI, GNDVI, DVI, RVI, SAVI, RDVI, CHM$	0.9114	0.8104	2.3955

3.4 Analysis of the results of each inversion model

The overall analysis of the model found that the model built only based on hyperspectral features had a lower R^2 and a higher RMSE, so this model built will not be analyzed separately. Comparing Figure 4 and Figure 5, it can be found that the MLR model fitting effect is better than that of the PLSR model which under the same parameters. Comparison of the same model under different parameters that the multi-source data of the model fitting effect are good was made. The model predicted more evenly distributed on either side of the 1:1 line, and a single data model fitting effect is relatively close, model prediction is uniformly distributed in the straight line on both sides. That means the

hyperspectral features increases the spectral detail information and the sensitive features of SPAD, which is conducive to the further improvement of model accuracy. By comprehensive analysis of the R^2 of the two models in the training set and testing set (Table 4), the PLSR model has a large R^2 difference in the two data sets under a single data source, while the MLR model has the same R^2 . The RMSE of the MLR model is 0.14 smaller than that of the PLSR model, indicating that the model is relatively stable. However, the inversion effect of the two models under multi-source data is similar. Therefore, the MLR model with multi-source data and single data has the best inversion effect.

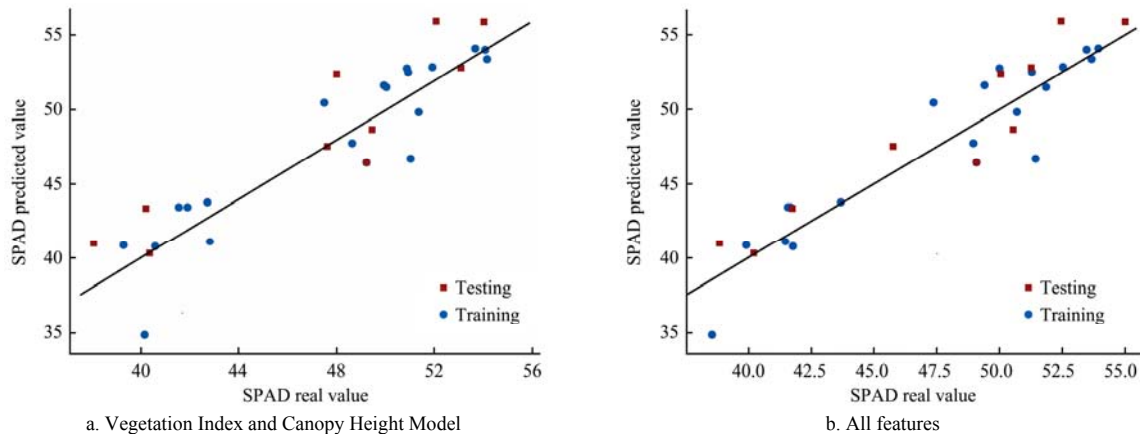


Figure 4 Partial least squares regression model under each parameter

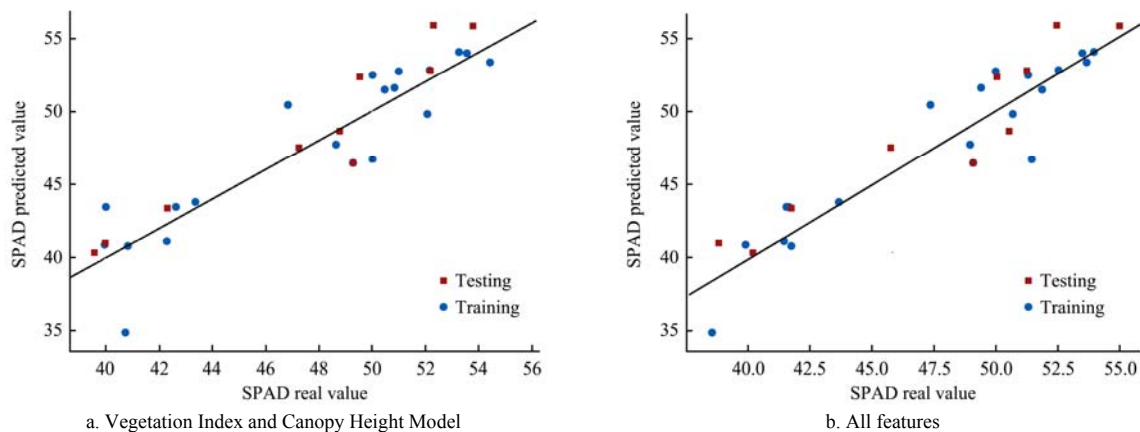


Figure 5 Multiple linear regression model under each parameter

According to the CART model results (Table 4), the model based on single parameters and multi-source parameters has the same prediction accuracy and error. By comparing the intermediate results, it is found that the CART model only keeps two terminal nodes through pruning operation, and the CART model under this node can explain most of the relationship between the predicted values and the true values. By analyzing the relative importance of the parameters to the model, we found (Figure 6) that in the case of two parameters, NDVI, RVI and SAVI had the highest importance, while the importance of effective plant height was up to 89%, and the importance of hyperspectral features was only 48.6%. Finally, only the NDVI terminal node is retained, which makes the model have the same accuracy and error under the two data parameters.

The random forest model has good data adaptability. Therefore, it's widely used in the regression and classification of data. This study used the random forest model to perform SPAD inversion modelling on summer maize under the two parameters

(Figure 7). It was found that the model under two kinds of parameters was stably after basic training 100 times, but when modelling only based on vegetation index and effective plant height, the model's R^2 of the training set is more than 0.9, while the R^2 of the testing set is less than 0.75. There is a big difference between the two sets, and the model has an overfitting phenomenon. When the model is built based on multi-source data, the R^2 of the model on the training set reaches about 0.9, and the R^2 on the testing set exceeds 0.8, indicating that the model has good generalization ability. This phenomenon indicates that multi-source data are helpful to increase the correlation degree between different features and targets which can improve the recognition accuracy of models. The over-fitting of the model is that when learning only based on vegetation index and effective plant height, the model identifies both vegetation index and effective plant height as SPAD sensitive features. Some irrelevant noises (such as some abnormal data and collinearity data) are also identified as SPAD sensitive features. Another one is that the

training set data is not large enough. The model cannot further determine the correlation between this information and SPAD,

which make the model unable to correctly identify the target features on the new data set, and finally leads to the over-fitting of the model.

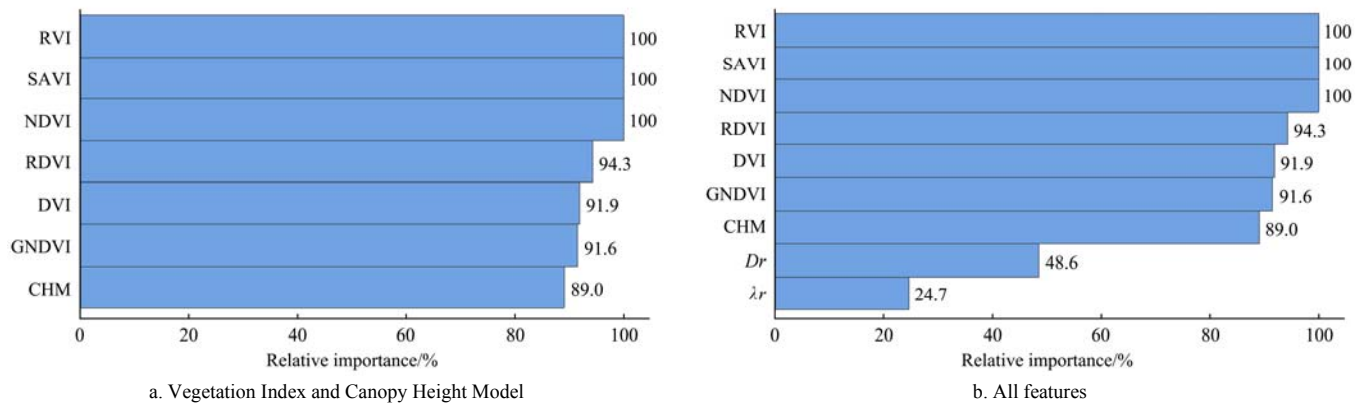


Figure 6 The importance of each parameter of the classification regression tree model

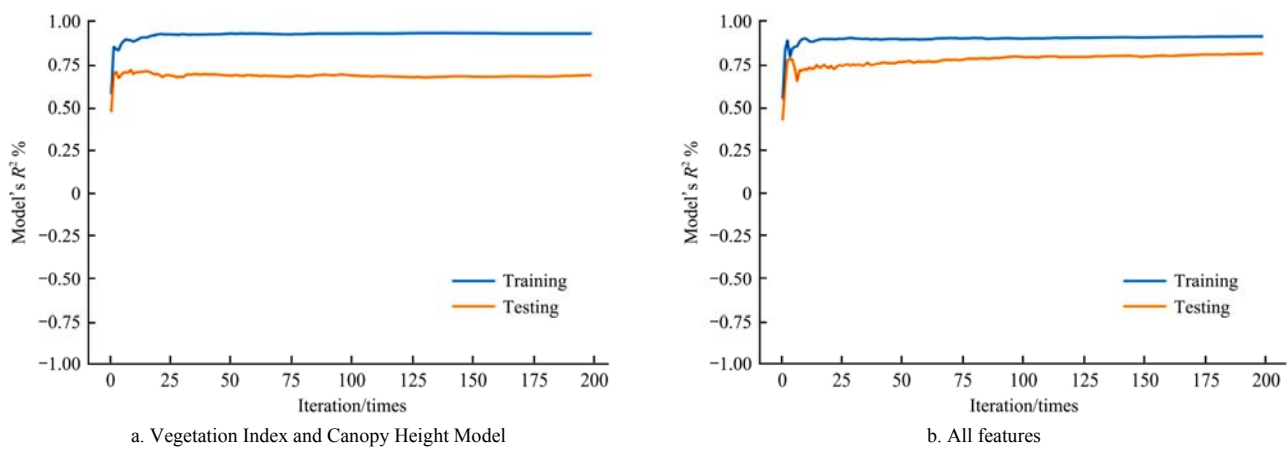


Figure 7 Accuracy rate of random forest model under each parameter

In conclusion, under a single data source, both the MLR model and the RF model have good accuracy, but the accuracy of the RF model on the testing set is greatly different from that of the training set, indicating the instability of the model, while the accuracy of the MLR model on the two data sets is relatively similar, so the best inversion model under a single data source is the MLR model. The MLR model, PLSR model and RF model all have good accuracy under multi-source data. The accuracy of the random forest model on the training set is 0.05 higher than that of the PLSR and MLR model, and the difference between the random forest model and MLR and PLSR model on the test set is only about 0.05. Moreover, the error of the random forest model is only about 0.4 of that of the MLR model and PLSR model. Therefore, the optimal inversion model under multi-source data is the random forest model.

Compared with previous SPAD inversion studies on a single data source^[26,27], this study combined hyperspectral data with multi-spectral data and adopted traditional linear regression and machine learning methods to compare and analyze the inversion accuracy of the SPAD model under single data source and multi-source data. The results show that the accuracy of the inversion model with multi-source data is better than that with a single data source, and the optimal inversion model of summer maize SPAD with multi-source data is the random forest model. Although the inversion accuracy of the model in this study is relatively high, the test results are only based on the one-year data of one maize variety in the test area, and the model training data are relatively small, so the generalization ability of the model needs to

be improved. Therefore, the data of multiple regions, long time series and multiple varieties data are still needed to improve the generalization ability and stability of the model.

4 Conclusions

According to the correlation between multiple parameters and summer maize SPAD, it can be seen that the position of red edge λ_r and the first-order differential value Dr , the UAV multi-spectral vegetation index and plant effective height in hyperspectral characteristics are all sensitive features of summer maize SPAD inversion.

For the SPAD inversion model with a single data source, the multiple linear regression model is the best. The model's R^2 is 0.8281 and RMSE is 2.136. The optimal SPAD inversion model under multi-source data is the random forest model, with a model's R^2 of 0.9114, RMSE is 2.3955. Moreover, the model based on multi-source data is superior to a single data source.

By comparing various models, it can be seen that the random forest model based on multi-source data has the best inversion effect. This study provides a SPAD estimation method for summer maize based on multi-source data, which can provide theoretical support for summer maize growth monitoring and fine management.

Acknowledgements

This study was supported by Top Talents Program for One Case One Discussion of Shandong Province and the Academy of Ecological Unmanned Farm (Grant No. 2019ZBXC200).

[References]

- [1] Zhang F, Zhang Y K, Mao P J, et al. Status and development of measuring method in plant Chlorophyll content, *Journal of Agricultural Mechanization Research*, 2014(4): 238–241. doi: 10.13427/j.cnki.njyi.2014.04.057. (in Chinese)
- [2] Yuan X K, Zhou G S, Wang Q L, et al. Hyperspectral characteristics of chlorophyll content in summer maize under different water irrigation conditions and its inversion. *Acta Ecologica Sinica*, 2021, 41(2): 543–552. doi: 10.5846/stxb201901110095. (in Chinese)
- [3] Su W, Zhao X F, Sun Z P, et al. Estimating the corn canopy chlorophyll content using the Sentinel-2A image. *Spectroscopy and Spectral Analysis*, 2019, 39(5): 1535–1542. doi: 10.3964/j.issn.1000-0593(2019)05-1535-08. (in Chinese)
- [4] Xi X, Zhao G X. Chlorophyll content in winter wheat: inversion and monitoring based on UAV multi-spectral remote sensing. *Chinese Agricultural Science Bulletin*, 2020, 36(20): 119–126. (in Chinese)
- [5] Du X Y, Wan L, Cen H Y, et al. Multi-temporal monitoring of leaf area index of rice under different nitrogen treatments using UAV images. *International Journal of Precision Agricultural Aviation*, 2020, 3(1): 7–12. doi: 10.33440/j.ijpaa.20200301.57.
- [6] Cai N, Zhou X G, Yang Y B, et al. Use of UAV images to assess narrow brown leaf spot severity in rice. *International Journal of Precision Agricultural Aviation*, 2019, 2(2): 38–42. doi: 10.33440/j.ijpaa.20190202.47.
- [7] Su W, Sun Z P, Chen W H, et al. Joint retrieval of growing season corn canopy LAI and leaf chlorophyll content by fusing sentinel-2 and MODIS images. *Remote Sensing*, 2019(11): 2409. doi: 10.3390/rs11202409.
- [8] Su W, Wang W, Liu Z, et al. Determining the retrieving parameters of corn canopy LAI and chlorophyll content computed using UAV image. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2020, 36(19): 58–65. doi: 10.11975/j.issn.1002-6819.2020.19.007. (in Chinese)
- [9] Hassanijalilian O, Igathinathane C, Doetkott C, et al. Chlorophyll estimation in soybean leaves infield with smartphone digital imaging and machine learning. *Computer and Electronics in Agriculture*, 2020(174): 105433. doi: 10.1016/j.compag.2020.105433.
- [10] Wen Y, Li Z M, Zhao Y, et al. Multispectral reflectance inversion and chlorophyll content diagnosis of maize at seeding stage. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2015, 31(Supp.2): 193–199. doi: 10.11975/j.issn.1002-6819.2015.z2.030. (in Chinese)
- [11] An G Q, Xing M F, He B B, et al. Using machine learning for estimating rice chlorophyll content from in situ hyperspectral data. *Remote Sensing*, 2020(12): 3104. doi: 10.3390/rs12183104.
- [12] Mao Z H, Deng L, Sun J, et al. Research on the application of UAV multispectral remote sensing in the maize chlorophyll prediction. *Spectroscopy and Spectral Analysis*, 2018, 38(9): 2923–2931. doi: 10.3964/j.issn.1000-0593(2018)09-2923-09. (in Chinese)
- [13] Liu H J, Zhao Y, Wen Y, et al. Diagnosis of chlorophyll content in corn canopy leaves based on multispectral detector. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(Supp.1): 228–233. doi: 10.6041/j.issn.1000-1298.2015.S0.037. (in Chinese)
- [14] Sun Y Y, Wang G P, Yang K M, et al. Research on linear model of high spectral retrieval of chlorophyll content in corn. *Shandong Agricultural Sciences*, 2015, 47(7): 117–121. doi: 10.14083/j.issn.1001-4942.2015.07.027. (in Chinese)
- [15] Raj R., Walker J P., Pingale R, et al. Leaf area index estimation using top-of-canopy airborne RGB image. *International Journal of Applied Earth Observations and Geoinformation*, 2021(96), 102282. doi: rg/10.1016/j.jag.2020.102.
- [16] Niu Q L, Feng H K, Yang G J, et al. Monitoring plant height and leaf area index of maize breeding material based on UAV digital images. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018, 34(5): 73–82. doi: 10.11975/j.issn.1002-6819.2018.05.010. (in Chinese)
- [17] Gao M Y, Zhang J S, Pan Y Z, et al. Retrieval of winter wheat leaf area index based on vegetation and crop height. *Chinese Journal of Agricultural Resources and Regional Planning*, 2020, 41(8): 49–57. doi: 10.7621/cjarrp.1005-9121.20200806. (in Chinese)
- [18] Li Y Y, Chang Q R, Liu X Y, et al. Estimation of maize leaf SPAD value based on hyperspectrum and BP neural network. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2016, 32(16): 135–142. doi: 10.11975/j.issn.1002-6819.2016.16.019. (in Chinese)
- [19] Yuan X K, Zhou G S, Wang Q L, et al. Hyperspectral characteristics of chlorophyll content in summer maize under different water irrigation conditions and its inversion. *Acta Ecologica Sinica*, 2021, 41(2): 543–552. doi: 10.5846/stxb201901110095. (in Chinese)
- [20] Wang H, Zhang Q, Wang Y, et al. Improved CART decision tree regression algorithm based on ELM. *Computer Systems & Applications*, 2021, 30(2): 201–206. doi: 10.15888/j.cnki.csa.007789. (in Chinese)
- [21] Zhang Y K, Wang J L, Su H, et al. Study on landform classification of Cui-Scale watershed units based on CART: A case study of the tropic of cancer (Yunnan section). *Geography and Geo-information Science*, 2021, 37(2): 84–92. doi:10.3969/j.issn.1672-0504.2021.02.012. (in Chinese)
- [22] Li C C, Niu Q L, Yang G J, et al. Estimation of leaf area index of soybean breeding materials based on UAV digital images. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(8): 147–158. doi: 10.6041/j.issn.1000-1298.2017.08.016. (in Chinese)
- [23] Shao G M, Wang Y J, Han W T. Estimation method of leaf area index for summer maize using UAV-Based multispectral remote sensing. *Smart Agriculture*, 2020, 2(3): 118–128. doi: 10.12133/j.smartag.2020.2.3.202006-SA001. (in Chinese)
- [24] Tian J C, Yang Z F, Feng K P, et al. Prediction of tomato canopy SPAD based on UAV multispectral image. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(8): 178–188. doi: 10.6041/j.issn.1000-1298.2020.08.020. (in Chinese)
- [25] Zhang X Z, Zheng G Q, Dai Y B, et al. Estimation models of summer maize leaf pigment content based on canopy reflectance spectra. *Journal of Maize Sciences*, 2010, 18(6): 55–60. doi: 10.13597/j.cnki.maize.science.2010.06.019. (in Chinese)
- [26] Tahir M N, Naqvi S Z A, Lan Y B, Zhang Y L, Wang Y K, Afzal M, et al. Real time monitoring chlorophyll content based on vegetation indices derived from multispectral UAVs in the kinnow orchard. *Int J Precis Agric Aviat*, 2018; 1(1): 24–31. doi: 10.33440/j.ijpaa.20180101.0001.
- [27] Yu F H, Du W, Guo Z H, Zhou C X, Wang D K, Xu T Y. UAV hyperspectral inversion modeling of rice nitrogen content based on WOA-ELM. *Int J Precis Agric Aviat*, 2019; 2(2): 43–48. doi: 10.33440/j.ijpaa.20190202.39.