Remote sensing estimation of chlorophyll content in rape leaves in Weibei dryland region of China

Liheng Xia (18092729601@163.com)
Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Natural Resources
Panpan Zhang
Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Natural Resources

Article

Keywords: rape leaves, SPAD, genetic algorithm, BP, neural network

Posted Date: March 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2675708/v1

Additional Declarations: No competing interests reported.

Abstract

To explore the Hyperspectral Estimation Method for estimating the chlorophyll content of rape leaves, so as to provide a scientific basis for rapid and nondestructive monitoring of the chlorophyll content of rape crops in Northwest China.Taking the rapeseed crops in the northwest region as the research object, through the correlation analysis of the SPAD value and the spectral parameters of the rape leaves, the spectral parameters sensitive to SPAD were screened, and the single factor model,the partial least square regression model (PLSR) and BP neural network model optimized by genetic algorithm based on multiple linear stepwise regression based on the spectral parameters were constructed respectively and were compared.The results showed that: 1) The general trend of the spectral curve of rape leaves was the same, and the spectral reflectance decreased with the increase of chlorophyll content; 2) The correlation of seven spectral parameters involved in the modeling was above 0.770, all of which reached significant correlation at 0.01 level; 3) In each growth period, the BP neural network model optimized by genetic algorithm based on multiple linear stepwise regression is the optimal model. The modeling R^2 is above 0.77, and the maximum can reach 0.91. It is verified that R^2 is above 0.73, the maximum can reach 0.92, RMSE is between 1.32–3.22, RE is between 2.50% – 4.49%. BP neural network model optimized by genetic algorithm based on multiple linear stepwise regression is an inversion method which can estimate the SPAD value of rape leaves accurately and quickly.

Introduction

Chlorophyll, a pigment essential for photosynthesis in green plants, is closely related to the photosynthetic capacity of plants and their growth status and is an important indicator of plant nitrogen content [1–5]. Traditional methods for measuring chlorophyll content mainly include spectrophotometry and atomic absorption, which are time-consuming, labor-intensive, and cumbersome processes. The emerging hyper-spectral remote sensing has the advantage of real-time, rapid, and non-destructive monitoring of chlorophyll content, providing an effective means for quantitative diagnosis of crop chlorophyll content [6–9]. At present, many scientists at home and abroad have conducted research on the application of hyperspectral remote sensing technology for the determination of chlorophyll content. There has been more research on wheat and maize and less on oilseed rape in terms of crops. The prediction models have been extended from simple traditional linear models to complex machine learning algorithm models. Liu Lu et al searched for the optimal spectral index by constructing a traditional linear model of the spectral reflectance of summer maize and the SPAD value of the plant canopy[10–11]. Based on ground-based hyperspectral and real-world agronomic data, Sampson et al concluded that combining the continuous wavelet transform and partial least squares regression could achieve remote sensing estimation of chlorophyll in winter wheat[12–14].

In recent years, BP neural network algorithm models are emerging for numerous applications due to their ability to implement any complex non-linear mapping and their self-learning capability, generalization, and generalization[15], and their application in chlorophyll prediction is gradually increasing. Liu Wenya et al used a BP neural network to quantitatively simulate the chlorophyll content of Sargassum pine [16]. regression model, and a BP neural network model based on spectral reflectance and SPAD values of maize leaves, and concluded that the BP neural network had the better predictive ability for chlorophyll. Genetic algorithms are parallel and can make it easier for BP neural networks to converge to a global minimum, improving the inversion accuracy of BP neural network models[17–20].

Studies employing multiple spectral parameters and using a BP neural network model optimized by a genetic algorithm based on multiple linear stepwise regression to predict SPAD values are rare. In the present study, oilseed rape in northwest China a partial least squares regression model (PLSR) and a multiple linear stepwise regression genetic algorithm based optimized BP neural network model (MLSR-GA-BP) was constructed by analyzing SPAD values of oilseed rape leaves and spectral parameters of leaves. This observation would provide theoretical and technical support for remote sensing monitoring of oilseed rape growth in the northwest.

Materials And Methods

Overview of the study area and experimental design

The present experimental study area is in Xipo Village, Xinglin Town, Fufeng County, Baoji City, Shaanxi Province, China, at longitude 107°58'48" E and latitude 34°19'12" N (Fig. 1). The area is in the western part of the Guanzhong Plain and the central and western part of Shaanxi Province. The region has an average annual temperature of 12.4°C and average annual precipitation of 592 mm, and mainly grows wheat, corn, and oilseed rape.

This experiment followed four critical fertility stages of oilseed rape for sampling and observation: seedling stage (29 March 2019), bolting stage (16 April 2019), flowering stage (25 April 2019), and maturation stage (11 May 2019). Sampling was done in each plot with eight sampling points. The upper, middle, and lower two leaves of an oilseed rape plant were collected as the sample of this sampling point then immediately put into plastic bags, drain, air was put into a thermostat box with ice bags, and brought to the laboratory. The SPAD is the relative content index of plant chlorophyll, and the SPAD value is often used as an indicator of chlorophyll content in practical studies, so the SPAD value and spectral reflectance of oilseed rape leaves were determined for the present study.

Leaf SPAD measurement

SPAD values of oilseed rape leaves were determined using a SPAD-502 handheld chlorophyll meter, and the leaves were numbered. First, a total of six leaves from the upper, middle, and lower layers were selected for measurement, followed by ten measurements at different parts of each leaf and averaged, and finally, the SPAD values of the six leaves of each sample point were averaged as the final SPAD value of that sample point.

Leaf spectral reflectance determination

The spectral reflectance of oilseed rape leaves was measured using an American SVC HR-1024i spectrometer to study the spectral 400 ~ 1 000 nm band, as the green plant spectrum is mainly influenced by chlorophyll in the visible band. The spectral profile of oilseed rape leaves was resampled to 1 nm using the spectrometer's own software SVC HR-1024i PC. Each sample point was selected to measure a total of 6 leaves in the upper, middle and lower 3 layers (corresponding to the SPAD leaves under test). To make the measurement results more representative, 32 spectra were measured for each leaf, and an average of 12 spectra was taken as the final spectral reflectance of the sample points.

Spectral parameter selection

In the present study, eight feature parameters based on hyperspectral position, five area feature parameters composed of spectral curves, four new spectral feature parameters and 14 vegetation indices were selected (Table 1).

Table 1
Spectral parameters and their definition or calculation formula

Parameters	Spectral parameters	Definition or algorithm Definition or algorithm
Hyperspectral position characteristic parameters	D_b	Maximum value of first order differentiation within 490-530nm
	D_y	Maximum value of first order differentiation within 560-640nm
	D_r	Maximum value of first order differentiation within 680-760nm
	λ_b	D_b corresponding to the wavelength position
	λ_y	D_y corresponding to the wavelength position
	λ_r	D_r corresponding to the wavelength position
	R_g	Maximum spectral reflectance in the range 510-560nm
	R_r	Maximum spectral reflectance in the range 640-680nm
Characteristic parameters of hyperspectral area	S_{Db}	Integration of first-order derivative spectra from 490–530 nm
	S_{Dy}	Integration of first-order derivative spectra from 560-640nm
	S_{Dr}	Integration of first-order derivative spectra from 680-760nm
	S_{Rg}	Sum of first order differential band values within 510-560nm
	S_{Rr}	Sum of first order differential band values within 640-680nm
New spectral characteristic parameters	S_g	Skewness of spectral reflectance in the 510-560nm range
	K_g	Kurtosis of spectral reflectance in the 510-560nm range
	S_r	Skewness of spectral reflectance in the 640-680nm range
	K_r	Kurtosis of spectral reflectance in the 640-680nm range
Vegetation index	NDVI	$(R_{800}-R_{670})/(R_{800}+R_{670})$

Parameters	Spectral parameters	Definition or algorithm Definition or algorithm
	GNDVI	$(R_{801}-R_{550})/(R_{801}+R_{550})$
	SR	R_{744}/R_{667}
	VOG_1	R_{740}/R_{720}
	VOG_2	$(R_{734}-R_{747})/(R_{715}+R_{726})$
	VOG_3	$(R_{734}-R_{747})/(R_{715}+R_{720})$
	CARI	$(R_{700}-R_{670})-0.2(R_{700}-R_{670})$
	TCARI	$3[(R_{700}-R_{670})-0.2(R_{700}-R_{550})(R_{700}/R_{670})]$
	MCARI	$[(R_{700}-R_{670})-0.2(R_{700}-R_{550})](R_{700}/R_{670})$
	OSAVI	$[(1+0.16)(R_{800}-R_{670})]/(R_{800}+R_{670}+0.16)$
	TCARI/OSAVI	TCARI/OSAVI
	MCARI/OSAVI	MCARI/OSAVI
	MTCI	$(R_{754}-R_{709})/(R_{709}-R_{681})$
	PRI	$(R_{570}-R_{531})/(R_{570}+R_{531})$

Model construction method

In this study, a multivariate linear stepwise analysis genetic algorithm-based BP neural network model (MLSR-GA-BP neural network) was used for optimization. The MLSR-GA-BP neural network model uses multiple linear stepwise regression to filter out variables with large influence on the dependent variable, and then uses genetic algorithms to optimize the BP neural network weights and thresholds to finally establish the optimal model. The specific genetic algorithm parameters and BP neural network parameters are shown in Table 2.

Table 2
Genetic algorithm parameters and BP neural network algorithm parameters

Parameters of genetic algorithm	Value	BP neural network parameters	Value
Population size	10	Training times	100
Maximum number of iterations	50	Network performance objectives	0.001
Crossover probability	0.8	Training function	trainlm
Mutation probability	0.1	Learning rate	0.1
Optimization scope	[-1,1]	Momentum coefficient	0.8
		Number of hidden layer nodes	3(Seedling stage is 2)
		Transfer function from input layer to output layer	Hyperbolic tangent S-shaped functions
		Transfer function from hidden layer to output layer	Log S functions

Model validation parameters

A total of three indicators, namely the coefficient of determination R^2, the root mean square error RMSE, and the relative error RE, were used to evaluate the fitting and forecasting ability of the model, where the larger the coefficient of determination R^2 (which cannot exceed 1) and the smaller the root mean square error RMSE and relative error RE, the better the fitting and forecasting ability of the model. In this experiment, the measured SPAD values were sorted by size and sampled by stratified sampling method according to 3:1 to construct the estimation model and validation model. Sixty samples of the estimation model and 20 samples of the validation model were taken for each fertility period[21–24]

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

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

$$RE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \quad (3)$$

Results And Discussion

Changes in SPAD and hyperspectral characteristics of oilseed rape leaves at different fertility stages

The mean SPAD of rapeseed leaves showed a decreasing trend, with a total decrease of 26.51% from seedling to maturation stage (Fig. 2). The results of the ANOVA showed that there were significant differences between seedling stage and bolting stage, flowering stage and maturity stage, respectively, and between maturity stage and bolting and flowering stage, respectively, but no significant differences between bolting and flowering stage.

The overall trend in spectral reflectance is consistent across all fertility stages and is in line with the pattern of hyperspectral reflectance of most green plants (Fig. 3). At 400–500 nm, a 1st absorption valley was formed due to the strong absorption of blue light by chlorophyll; At 600–700 nm, a 1st absorption valley is formed due to the strong absorption of red light by chlorophyll. The two absorption valleys form a strong reflectance peak around 550 nm, with a sharp increase in reflectance at 670 ~ 780 nm due to the strong absorption of red light by chlorophyll gradually changing to scattering of NIR light; At 780 to 1 000 nm, a persistently high reflectance plateau is formed due to the internal structure of the leaf and the concentration of nutrients within the leaf, among other things. From Fig. 3, it is apparent that the changes of reflectance in different fertility stages, the overall trend from seedling to maturity reflectance was on the rise, but in the vicinity of 400 nm-520 nm and 580 nm-690 nm, the flowering stage reflectance was higher greater than the maturity stage reflectance, the overall trend was the opposite of the average law of SPAD of rapeseed leaves in different fertility stages. The reason is that the higher the SPAD value of rape leaves, the higher the chlorophyll content and thus the stronger the light absorption, and the lower the reflectance.

Correlation analysis between SPAD values and raw spectral reflectance of oilseed rape leaves

The SPAD values of oilseed rape leaves correlated with the original spectral reflectance, showed that the SPAD values of oilseed rape leaves at all fertility stages were highly significantly correlated with the original spectra in the green-yellow and red light bands, with the greatest correlations around 550 nm and 710 nm, and both were significantly negatively correlated (Fig. 4). The correlation declined rapidly after 710 nm in all four fertility stages and then increased slowly, but only the maturity stage reached a significant positive correlation at the 0.01 level and not in the other three fertility stages.

Correlation analysis of SPAD values and spectral parameters in oilseed rape leaves

SPSS software was used to analyze the correlation between SPAD value and spectral parameters of rape leaves, and the results are shown in Table 3. Twenty-two spectral parameters were significantly correlated at the 0.01 level (specific spectral parameters were not identical), one spectral parameter was significantly correlated at the 0.05 level, and eight spectral parameters were not correlated (specific spectral parameters were not identical) at the seedling, bud and flowering stages, and 26 spectral parameters were significantly correlated at the 0.01 level, one spectral parameter was significantly correlated at the 0.05 level, and four spectral parameters were not correlated at the maturity stage. Therefore, the correlation between leaf SPAD values and spectral parameters was better at maturity compared to the other three fertility stages. At the seedling stage, the best correlated spectral parameter was MCARI, with a correlation of 0.808, of which S_{Rg}, CARI, and MCARI all correlated above 0.800; at the bud stage, the best correlated spectral parameters were VOG_2 and MTCI, with a correlation of 0.908, of which 14 spectral parameters correlated above 0.800; at the flowering stage, the best correlated The spectral parameter with the best correlation was CARI with a correlation of 0.839, of which $D_{b,}$ S_{Rg}, CARI, MCARI and MTCI correlated to 0.800 or more; at maturity, the spectral parameter with the best correlation was S_{Rg} with a correlation of 0.869, of which 13 spectral parameters correlated to 0.800 or more. From the perspective of spectral parameters, a total of seven spectral parameters, D_b, S_{Db}, S_{Rg}, CARI, MCARI, MCARI/OSAVI, and MTCI, all correlated above 0.770 in the four reproductive stages, with better correlations, and could better participate in model building.

Table 3
Correlation between SPAD value and spectral parameters of rape leaves in different growth stages

Spectral parameters	Seedling stage	Bolting stage	Flowering stage	Maturation stage
D_b	-0.783**	-0.887**	-0.817**	-0.853**
D_y	0.440**	0.721**	0.579**	0.690**
D_r	-0.247*	-0.479**	-0.529**	-0.507**
λ_b	0.295**	0.620**	0.425**	0.385**
λ_y	0.311**	0.286*	0.07	0.11
λ_r	0.721**	0.849**	0.779**	0.771**
R_g	-0.387**	-0.636**	-0.533**	-0.670**
R_r	0.087	0.105	0.077	-0.018
S_{Db}	-0.773**	-0.870**	-0.791**	-0.850**
S_{Dy}	0.710**	0.853**	0.734**	0.800**
S_{Dr}	-0.046	0.138	-0.078	0.289**
S_{Rg}	-0.800**	-0.882**	-0.813**	-0.869**
S_{Rr}	0.403**	0.618**	0.555**	0.777**
S_g	0.203	0.521**	0.342**	0.535**
K_g	-0.306**	-0.585**	-0.408**	-0.632**
S_r	-0.310**	-0.156	-0.252*	0.429**
K_r	-0.173	-0.087	-0.095	0.455**
NDVI	-0.103	-0.061	-0.1	0.129
GNDVI	0.466**	0.776**	0.599**	0.767**
SR	-0.186	-0.19	-0.176	0.089
VOG_1	0.692**	0.894**	0.754**	0.835**
VOG_2	-0.737**	-0.908**	-0.779**	-0.835**

Note: ** Significantly correlated at 0.01 level (both sides), * Significantly correlated at 0.05 level (both sides).

Spectral parameters	Seedling stage	Bolting stage	Flowering stage	Maturation stage
VOG_3	-0.733**	-0.907**	-0.778**	-0.833**
CARI	-0.805**	-0.877**	-0.839**	-0.865**
TCARI	-0.748**	-0.842**	-0.755**	-0.823**
MCARI	-0.808**	-0.855**	-0.802**	-0.855**
OSAVI	-0.100	0.014	-0.094	0.221*
TCARI/OSAVI	-0.699**	-0.811**	-0.717**	-0.801**
MCARI/OSAVI	-0.795**	-0.841**	-0.779**	-0.845**
MTCI	0.776**	0.908**	0.801**	0.855**
PRI	-0.087	-0.163	-0.21	-0.514**
Note: ** Significantly correlated at 0.01 level (both sides), * Significantly correlated at 0.05 level (both sides).				

Comparison of single-factor model construction and accuracy of SPAD values in oilseed rape leaves

The results of model modeling are shown in Table 4, which shows that at the seedling stage, the MCARI power function model has the highest coefficient of determination R^2 and its validation coefficient R^2 is larger, while RMSE and RE are smaller. Similarly, the MCARI/OSAVI quadratic polynomial model had the best prediction ability at the bud stage; at the flowering stage, the exponential model developed by CARI had the best prediction ability; at the maturity stage, the MCARI power function model had the best prediction ability compared with the models developed by other parameters during the same period. Comparing the optimal models for the four fertility stages, it can be seen that the modeling coefficient of determination and the validation coefficient of determination were the highest for the bud shoots stage, and the RMSE and RE were both lower, so from the perspective of univariate modeling it was suitable to estimate the SPAD values of rape leaves at the bud shoots stage using the quadratic polynomial model developed by MCARI/OSAVI.

Table 4
Single factor optimal estimation model of spectral parameters and SPAD values in different growth periods

Growth period	Spectral parameters	Model	Modeling accuracy	Verification accuracy		
			R^2	R^2	RMSE	RE/%
Seedling stage	D_b	$y = -25.65\ln(x) - 94.702$	0.61	0.81	3.61	4.78
	S_{Db}	$y = -21.58\ln(x) - 4.9315$	0.59	0.76	3.62	4.81
	S_{Rg}	$y = -22.6\ln(x) - 9.737$	0.64	0.75	3.45	4.65
	CARI	$y = 69.858e^{-8.73x}$	0.64	0.72	3.25	4.63
	MCARI	$y = 29.96x^{-0.318}$	0.65	0.74	3.33	4.57
	MCARI/OSAVI	$y = 33.808x^{-0.315}$	0.63	0.76	3.35	4.50
	MTCI	$y = 0.8265x^2 + 7.9612x + 38.246$	0.57	0.62	3.40	4.33
Bolting stage	D_b	$y = 2E+06x^2 - 20966x + 100.39$	0.85	0.61	3.86	6.88
	S_{Db}	$y = 3118.5x^2 - 816.18x + 93.073$	0.84	0.80	3.04	4.73
	S_{Rg}	$y = 3847.9x^2 - 909.94x + 93.263$	0.84	0.74	2.84	4.65
	CARI	$y = 3856.3x^2 - 695.32x + 71.093$	0.83	0.80	3.59	5.58
	MCARI	$y = 319.86x^2 - 239.31x + 84.629$	0.84	0.80	3.40	5.38
	MCARI/OSAVI	$y = 157.92x^2 - 169.68x + 84.942$	0.85	0.84	3.35	5.08
	MTCI	$y = -1.208x^2 + 19.044x + 23.883$	0.83	0.65	3.52	5.63
Flowering stage	D_b	$y = -6231.5x + 72.445$	0.68	0.35	3.41	5.25
	S_{Db}	$y = 1521.8x^2 - 484.35x + 77.58$	0.66	0.32	3.28	4.95

Note: x represents the spectral parameter, y represents the SPAD value of rape leaves, the same below.

	S_{Rg}	$y = 2157.5x^2 - 592.36x + 79.345$	0.69	0.34	3.13	4.69
	CARI	$y = 63.157e^{-6.189x}$	0.71	0.42	2.85	4.67
	MCARI	$y = 29.603x^{-0.32}$	0.68	0.35	2.98	4.63
	MCARI/OSAVI	$y = 33.688x^{-0.312}$	0.66	0.35	3.06	4.69
	MTCI	$y = 41.659x^{0.4104}$	0.61	0.41	2.99	4.68
Maturation stage	D_b	$y = 77.732e^{-130.4x}$	0.74	0.56	2.07	3.64
	S_{Db}	$y = 68.228e^{-4.623x}$	0.74	0.56	2.04	3.49
	S_{Rg}	$y = 75.008e^{-6.24x}$	0.76	0.58	1.83	3.16
	CARI	$y = 61.592e^{-5.848x}$	0.78	0.66	2.14	3.64
	MCARI	$y = 27.512x^{-0.363}$	0.77	0.66	2.02	3.38
	MCARI/OSAVI	$y = 31.352x^{-0.35}$	0.76	0.65	2.07	3.50
	MTCI	$y = 38.741x^{0.4455}$	0.75	0.61	2.13	3.73

Note: x represents the spectral parameter, y represents the SPAD value of rape leaves, the same below.

Comparison of partial least squares regression model construction and accuracy

The Bolting stage modeling coefficient of determination R^2 and the validation coefficient of determination R^2 were the highest, the RMSE was below 3 and the RE was below 4.5%, both of which were small (Table 5), so the Bolting stage model had the highest accuracy and the best fitting predictive ability among the four fertility stages.

Table 5
Partial least squares regression model for different growth periods

Growth period	Model	Modeling accuracy	Verification accuracy		
		R^2	R^2	RMSE	RE/%
Seedling stage	y = 75.7-13446.5*Db + 2564.1*SDb-1702.5*SRg-52.7*CARI + 29*MCARI-227.7*MCARI/OSAVI + 1.5*MTCI	0.73	0.75	2.97	4.01
Bolting stage	y = 45.9-15160.8*Db + 356.2*SDb + 281.8*SRg + 86.8*CARI-143*MCARI + 66.3*MCARI/OSAVI + 11.3*MTCI	0.87	0.86	3.00	4.48
Flowering stage	y = 53.5-11223.7*Db + 738.8*SDb-297.2*SRg-737.2*CARI + 211.13*MCARI-74*MCARI/OSAVI + 4.8*MTCI	0.83	0.52	3.58	5.61
Maturation stage	y = 62.91 + 3726.06*Db + 98.11*SDb-555.77*SRg-219.93*CARI-57.46*MCARI + 71.36*MCARI/OSAVI + 3.19*MTCI	0.77	0.80	2.13	4.01

Construction and accuracy comparison of BP neural network model (MLSR-GA-BP neural network) optimized by genetic algorithm based on multivariate linear stepwise analysis

The model modeling results are shown in Table 6. from Table 6, it can be seen that the highest modeling coefficient of determination R^2 and the best modeling results were obtained for the budburst stage, followed by the flowering stage, maturity stage, and seedling stage. However, the validation coefficient of determination R^2 was the highest for the maturity stage, and the RMSE and RE were the lowest, so the validation accuracy was the highest for the maturity stage, followed by the bud stage, seedling stage, and flowering stage. Combining the validation accuracy and prediction accuracy, both the bud shoots and maturity stage fit predictions were better.

Table 6
MLSR-GA-BP neural network model in different growth stages

Growth period	Modeling accuracy			Verification accuracy		
	R^2	RMSE	RE/%	R^2	RMSE	RE/%
Seedling stage	0.77	2.83	3.78	0.78	2.94	3.95
Bolting stage	0.91	2.63	3.61	0.87	2.96	4.47
Flowering stage	0.83	2.22	3.54	0.73	3.22	3.22
Maturation stage	0.82	1.94	3.62	0.92	1.32	2.50

Comparison of the effect of each model and model accuracy

To select the optimal model for estimating SPAD values of oilseed rape leaves, the above models were compared and shown in Table 7. The modeling coefficient of determination R^2 of MLSR-GA-BP was the largest at all four fertility stages and had the best modeling effect, PLSR was the second and the single-factor model was the worst. In terms of model accuracy, the validation coefficient of determination R^2 of MLSR-GA-BP was the largest in all four fertility stages, and the root mean square error RMSE of validation was the lowest in seedling, budburst and maturity stages, slightly higher in flowering stage than the one-factor model, and the relative error RE of validation was the smallest in all. The MLSR-GA-BP neural network model had the highest prediction accuracy when the three indices were combined. Therefore, the spatial distribution of the measured and predicted values of the validation samples of the MLSR-GA-BP neural network model was plotted as a scatter plot (Fig. 5), in which the solid line is the regression equation of the measured values and the dashed line is the 1:1 line, and when the regression equation is closer to the 1:1 line, it indicates that the model works better. From Fig. 5, it can be seen that the maturity model estimates the best results with a slope of the curve of 0.97, which is closest to 1.

Table 7
Comparison of models in different growth periods

Growth period	Model	Modeling accuracy	Verification accuracy		
		R^2	R^2	RMSE	RE/%
Seedling stage	Single Factor	0.65	0.74	3.33	4.57
	PLSR	0.73	0.75	2.97	4.01
	MLSR-GA-BP neural network	0.77	0.78	2.94	3.95
Bolting stage	Single Factor	0.85	0.84	3.35	5.08
	PLSR	0.87	0.86	3.00	4.48
	MLSR-GA-BP neural network	0.91	0.87	2.96	4.47
Flowering stage	Single Factor	0.71	0.42	2.85	4.67
	PLSR	0.83	0.52	3.58	5.61
	MLSR-GA-BP neural network	0.83	0.73	3.22	3.22
Maturation stage	Single Factor	0.77	0.66	2.02	3.38
	PLSR	0.77	0.80	2.13	4.01
	MLSR-GA-BP neural network	0.82	0.92	1.32	2.50

Discussion

In the visible band, SPAD values of oilseed rape leaves showed a negative correlation with leaf spectral reflectance, indicating that the higher the chlorophyll content, the lower the spectral reflectance and the stronger the absorption of light. This is because in the visible band, the reflectance spectrum of green plants is mainly influenced by phytochromes, especially chlorophyll, in which the absorption of light is strong and the reflectance is low[25–28].

The choice of model is crucial when estimating leaf SPAD values [29–31]. Many researcher have applied various models to the estimation of SPAD values of various crops, and most of them believe that multi-factor models are better than single-factor models, machine learning algorithm models are better than traditional linear models, and BP neural network algorithm models are better than partial least squares algorithm models in machine learning algorithms. The possible reasons are as follows: firstly, the SPAD value reflects the chlorophyll content, and the difference of chlorophyll content is reflected in the reflectance spectrum from visible to near-infrared multiple spectral bands, and the spectral parameters are the combination of different spectral bands, and the single-factor modeling only considers a single spectral parameter, while the multi-factor modeling involves multiple spectral parameters in modeling, i.e., multiple spectral bands are involved in modeling, so the multi-factor model is better than the single-factor model. Secondly, the partial least squares algorithm model concentrates the features of principal component analysis, typical correlation analysis and linear regression analysis [32–36]. The BP neural network algorithm model has the function of realizing any complex nonlinear mapping, which makes it especially suitable for solving problems with complex internal mechanisms, and it has self-learning ability, generalization and generalization ability, and the parallelism of genetic algorithm can make it easier to converge to the global minimum. These advantages make the machine learning algorithm have advantages that the traditional linear model does not have. So the machine learning algorithm model is better than the traditional linear model, and the BP neural network algorithm model is better than the partial least squares algorithm model. When studying BP neural networks, there is no fixed method for calculating the number of nodes in its hidden layer, and several methods have been used by previous authors[37–38], but after training, none of them is suitable for this model, so the number of nodes in the hidden layer of this model is selected after several attempts to find the optimal number of nodes.

At present, there is no unified definitive model for estimating SPAD values at home and abroad, and the selection of the optimal model varies[39–42]. This is because chlorophyll content varies from crop to crop, and chlorophyll content of the same crop can be affected by varietal, planting location, water and fertilizer status, and instrumental differences, which have led to the absence of a uniform model standard for estimating SPAD values to date. The SPAD values of oilseed rape leaves measured in this paper are also limited to respond to the situation in the northwest arid region, and the model results affected by this cannot be applied to the estimation of SPAD values of oilseed rape leaves in other regions yet, and need to be further studied before being extended[42–45].

Conclusions

In this paper, the SPAD estimation models based on partial least squares regression model (PLSR) and BP neural network based on multiple linear stepwise regression with genetic algorithm optimization (GA) were constructed and compared to the following conclusions by analyzing the correlation between SPAD values and spectral parameters of oilseed rape leaves in the northwest region.

1)The mean SPAD values of oilseed rape showed a gradual decrease at the seedling, bolting, flowering and maturation stages. The overall trend of spectral reflectance of oilseed rape leaves showed a gradual increase, but near 400 nm-520 nm and 580–690 nm, the reflectance at flowering stage was greater than that at maturation, and the overall trend was due to the higher SPAD value of oilseed rape leaves, the higher chlorophyll content and thus the stronger absorption of light and lower reflectance, resulting in the opposite trend of the two. The mean values of SPAD in oilseed rape were not significantly different at the bud and flowering stages, and showed significant differences in all other reproductive stages.

2)From the perspective of spectral parameters, a total of seven spectral parameters, D_b, SD_b, SRg, CARI, MCARI, MCARI/OSAVI, and MTCI, were correlated above 0.770 at all four fertility stages, which were well correlated and could better participate in model building.

3)From the optimal model, the MLSR-GA-BP neural network model was the optimal model, and the modeling R2 was above 0.77 and the maximum could reach 0.91, and the validation R2 was above 0.72 and the maximum could reach 0.92, which could better predict the SPAD value of rape leaves.

Declarations

Acknowledgements

The authors are grateful to the authorities of driving forces (DJNY2022-36) for financial support from Analysis of the evolution of spatial and temporal patterns of arable land in the northern Weibei dry plateau area. for this study.

Author Contributions

The manuscript was reviewed and approved for publication by all authors. L.X. conceived and designed the experiments. L.X. performed the experiments, analyzed the data, drew the figures and wrote the paper. P.P.Z. and P.P.Z. revised the paper.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

StatementThe use of plants in the present study complies with international, national and/or institutional guidelines.

Data availability The datasets generated and analysed during the current study are not publicly available due this experiment was a collaborative effort, the trial data does not belong to me alone but are available from the corresponding author on reasonable request.

References

1. Delegido J, Alonso L, González G, Moreno J. Estimating chlorophyll content of crops from hyperspectral data using a normalized area over reflectance curve (NAOC). Int. J. Appl. Earth Observ. Geoinforma. 12, 165–174 (2020).
2. Francisco Flores-de-Santiago, John M, Kovacs Francisco, Flores-Verdugo. The influence of seasonality in estimating mangrove leaf chlorophyll-a content from hyperspectral data. Wetlands Ecol. Manage. 21, 1232–1245 (2013).
3. Gopal KNSR, Prafull S, Himesh P, Vaishangi B, Bappa D, Sudhir K, Raju D, Chandrapal V and Madan PVC. Application of thermal imaging and hyperspectral remote sensing for crop water deficit stress monitoring. Geocarto Int. 36, 223–234 (2021).
4. Liu L, Peng Z.G, Zhang B.Z, et al. Hyperspectral-based monitoring of SPAD values in summer maize canopies. J*ournal of Soil and Water Conservation*,2019,33, 353–360 (2013).
5. Masood S, Zhao XQ and Shen RF. Bacillus pumilus increases boron uptake and inhibits rapeseed growth under boron supply irrespective of phosphorus fertilization. AOB Plants, 11, z36 (2019).
6. Ustin SL, Gitelson AA, Jacquemoud S, Schaepman M, Asner GP, Gamon JA and Zarco-Tejada P. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sens. Environ. 113, S67-S77 (2009).
7. Delegido J, Fernandez G, Gandia S, Moreno J. Retrieval of chlorophyll content and LAI of crops using hyperspectral techniques: application to PROBA/CHRIS data. Int. J. Remote Sens.29, 7107–7127 (2008).
8. Gu Y, Liu T, Gao G, Ren G, Ma Y, Chanussot J and Jia X. Multimodal hyperspectral remote sensing: an overview and perspective. Science China Inform. Sci. 64, 1–24 (2021).
9. Kittipornkul P, Treesubsuntorn C and Thiravetyan P. Effect of exogenous catechin and salicylic acid on rice productivity under ozone stress: the role of chlorophyll contents, lipid peroxidation, and antioxidant enzymes. Environ. Sci. Pollut. Res. 27, (2020).
10. Banerjee S, Shanmugam P.. Atmospheric correction of airborne hyperspectral remote sensing data for inland water applications. Hyperspectral Imaging and Sounding of the Environ. 55, 125–134 (2018).
11. Dash J, Curran P J. Evaluation of the MERIS terrestrial chlorophyll index. Adv. Space Res. 39, 221–243 (2004).
12. Dong Z, Yang W, Zhang M, Zhu H, Wang C. Estimation models of maize leaf SPAD value based on hyperspectral remote sensing. Crops 27, 126–131 (2019).
13. Kawamura, K.; Watanabe, N.; Sakanoue, S.; Inoue, Y. Estimating forage biomass and quality in a mixed sown pasture based on partial least squares regression with waveband selection. Jpn. Soc. Grassl. Sci. 54, 131–145 (2008).
14. Nakaji, T.; Oguma, H.; Nakamura, M.; Kachina, P.; Asanok, L.; Marod, D.; Aiba, M.; Kurokawa, H.; Kosugi, Y.;Kassim, A.R.; et al. Estimation of six leaf traits of East Asian forest tree species by leaf

spectroscopy and partial least square regression. Remote Sens. Environ. 233, 111381 (2019).

15. Serbin, S.P.; Singh, A.; Desai, A.R.; Dubois, S.G.; Jablonski, A.D.; Kingdon, C.C.; Kruger, E.L.; Townsend, P.A. Remotely estimating photosynthetic capacity, and its response to temperature, in vegetation canopies using imaging spectroscopy. Remote Sens. Environ.167, 78–87 (2015).
16. Weiss, M.; Jacob, F.; Duveiller, G. Remote Sensing of agricultural applications: A meta-review. Remote Sens. Environ. 236, 111402 (2020).
17. Schellberg, J.; Hill, M.J.; Gerhards, R.; Rothmund, M.; Braun, M. Precision agriculture on grassland: Applications, perspectives and constraints. Eur. J. Agron. 29, 59–71 (2008).
18. Obermeier, W.A.; Lehnert, L.W.; Pohl, M.J.; Gianonni, S.M.; Silva, B.; Seibert, R.; Laser, H.; Moser, G.; Miller, C.; Luterbacher, J.; et al. Grassland ecosystem services in a changing environment: The potential of hyperspectral monitoring. Remote Sens. Environ. 232, 111273 (2019).
19. Cho, M.A.; Skidmore, A.; Corsi, F.; Van Wieren, S.E.; Sobhan, I. Estimation of green grass/herb biomass for airborne hyperspectral imagery using spectral indices and partial least squares regression. Int. J. Appl. Earth Obs. 9, 414–424 (2007).
20. Dechant, B.; Ryu, Y.; Kang, M. Making full use of hyperspectral data for gross primary productivity estimation with multivariate regression: Mechanistic insights from observations and process-based simulations. Remote Sens. Environ. 234, 111435 (2019).
21. Fu, Y.; Yang, G.; Wang, J.; Song, X.; Feng, H. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. Comput. Electron. Agric. 100, 51–59 (2014).
22. Li, F.; Mistele, B.; Hu, Y.; Chen, X.; Schmidhalter, U. Reflectance estimation of canopy nitrogen content in winter wheat using optimized hyperspectral spectral indices and partial least squares regression. Eur. J. Agron. 52, 198–209 (2014).
23. Meach-Hensold, K.; Montes, C.M.; Wu, J.; Guan, K.; Fu, P.; Ainsworth, E.A.; Pederson, T.; Moore, C.E.; Brown, K.L.; Raines, C.; et al. High-throughput field phenotyping using hyperspectral reflectance and partial, least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. Remote Sens. Environ. 231, 111176 (2019).
24. De Jong, S. SIMPLS: An alternative approach to partial least squares regression. Chemom. Intell. Lab. 18, 251–263 (1993).
25. Geladi, P.; Kowalski, B. Partial least squares regression: A tutorial. Anal. Chim. Acta 1986, 185, 1–17.
26. Caporaso, N.; Whitworth, M.B.; Fisk, I.D. Protein content prediction in single wheat kernels using hyperspectral imaging. Food Chem. 240, 32–42 (2018).
27. Flynn, K.C. Site suitability analysis for tef (Eragrostis tef) within the contiguous United States. Comput. Electron. Agric. 159, 119–128 (2019).
28. Boe, A.; Somerfieldt, J.; Wynia, R.; Thiex, N. A preliminary evaluation of the forage potential of teff. *Proc. S. Dak. Acad. Sci*. 65, 75–82 (1986).

29. Bultosa, G.; Taylor, J.N.R. Encyclopedia of Grain Science; Wringley, C., Corke, H., Walker, C., Eds.; Academic Press: Oxford, UK.pp. 281–289 (2004).
30. Dekking, L.S.; Winkelaar, Y.K.; Koning, F. The Ethiopian cereal tef in celiac disease. N. Engl. J. Med. 353, 1748–1749 (2005).
31. Gerbremariam, M.M.; Zarnkow, M.; Becker, T. Teff (Eragrostis tef) as a raw material for malting, brewing and manufacturing of gluten-free foods and beverages: A review. J. Food Sci. Tech. 51, 2881–2895 (2014).
32. Hopman, G.D.; Dekking, E.H.A.; Blokland, M.L.J.; Wuisman, M.C.; Zuijderduin, W.M.K.F.; Schweizer, J.J. Tef in the diet of celiac patients in the Netherlands. Scand. J. Gastroenterol. 43, 277–282 (2008).
33. Twidwell, E.K.; Boe, A.; Casper, D.P. Tef: A New Annual Forage Grass for South Dakota? South Dakota State University Extra Extension: Brookings, SD, USA.p. 8071 (2002).
34. Flynn, K.C.; Frazier, A.E.; Admas, S. Performance of chlorophyll prediction indices for Eragrostis tef at Sentinel-2 MSI and Landsat-8 OLI spectral resolutions. Precis. Agric. 1–15 (2020).
35. National Forage Testing Association. Forage Analyses Procedures; South Dakota State University Extra Extension: Brookings, SD, USA, pp. 79–81 (1993).
36. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639 (2021).
37. Dawson, C.T.; Curran, P.J. A new technique for interpolating the reflectance red edger position. Int. J. Remote Sens. 19, 2133–2139 (1998).
38. Frazier, A.E.; Wang, L.; Chen, J. Two new hyperspectral indices for comparing vegetation chlorophyll content. Geo Spat. Inf. Sci. 17, 17–25 (2014).
39. Kokaly, R.F.; Asner, G.P.; Ollinger, S.V.; Martin, M.E.; Wessman, C.A. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sens. Environ.,113, 578–591 (2009).
40. Khan Z, Zhang K, Khan MN, Fahad S and Hu L. Coupling of biochar with nitrogen supplements improve soil fertility, nitrogen utilization efficiency and rapeseed growth. Agron. 10, 1334–1355 (2020).
41. Liu W.Y., Pan J. A neural network-based model for hyperspectral estimation of chlorophyll content in Pinus equisetum[J]. Journal of Applied Ecology.28, 1128–1136 (2017).
42. Lockwood RB, Chrisp MP, Parameswaran L, Thome KJ and Babu SR. Remote sensing using VNIR/SWIR dispersive imaging spectrometers: historical development, current state-of-the-art, and future trends. IEEE Int. Geosci. *Remote Sens. Symp*. 67, 34–44 (2018).
43. Nian, Wang, Biyun, Chen, Kun, Xu, Guizhen, Gao and Feng, Li. Association mapping of flowering time QTLs and insight into their contributions to rapeseed growth habits. Front. Plant Sci. 7, 755–768 (2016.).
44. Sampson, PH, Mohammed, GH, Zarco-Tejada, PJ, Miller, JR and Noland TL. The bioindicators of Forest Condition Project: A physiological, remote sensing approach. Forest. Chronicle 76, 941–952

(2000).

45. Wesley J, Moses Anatoly A, Gitelson, Richard L and Perk Daniela. Estimation of chlorophyll-a concentration in turbid productive waters using airborne hyperspectral data. Water Res. 46, 588–602 (2012).

Figures

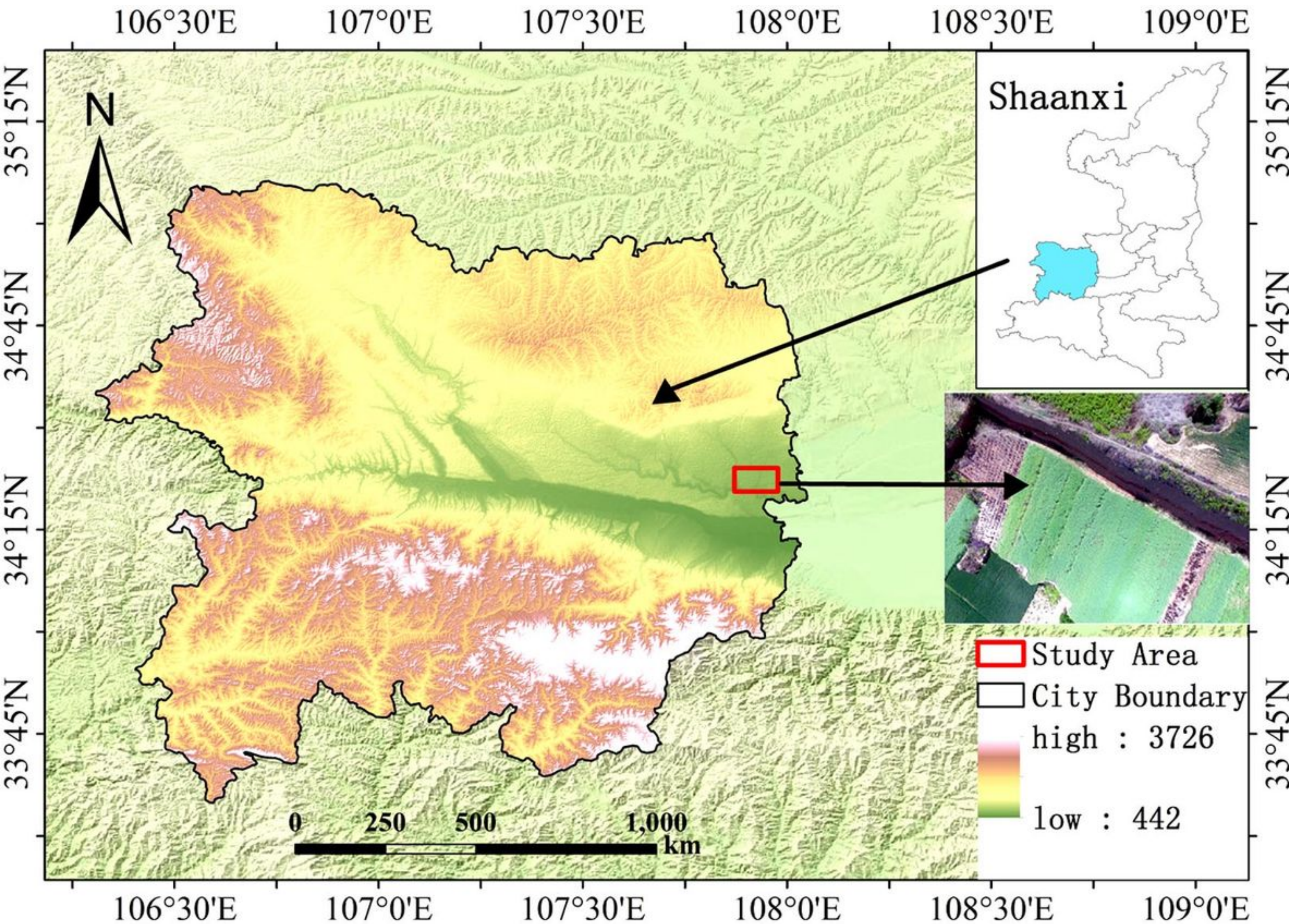

Figure 1

Sketch map of rape experimental field location

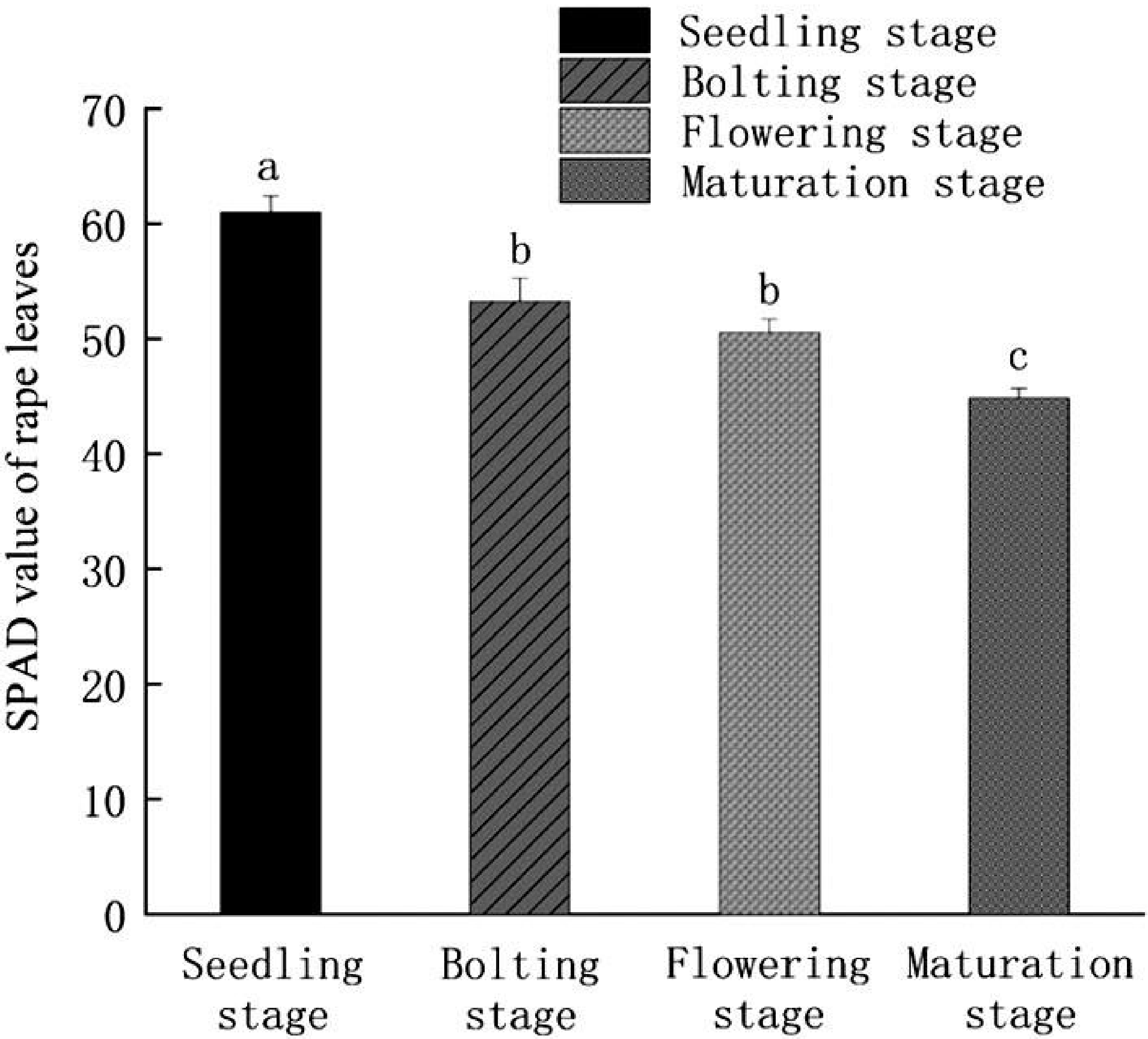

Figure 2

Changes of SPAD average value of rape leaves in different growth stages.

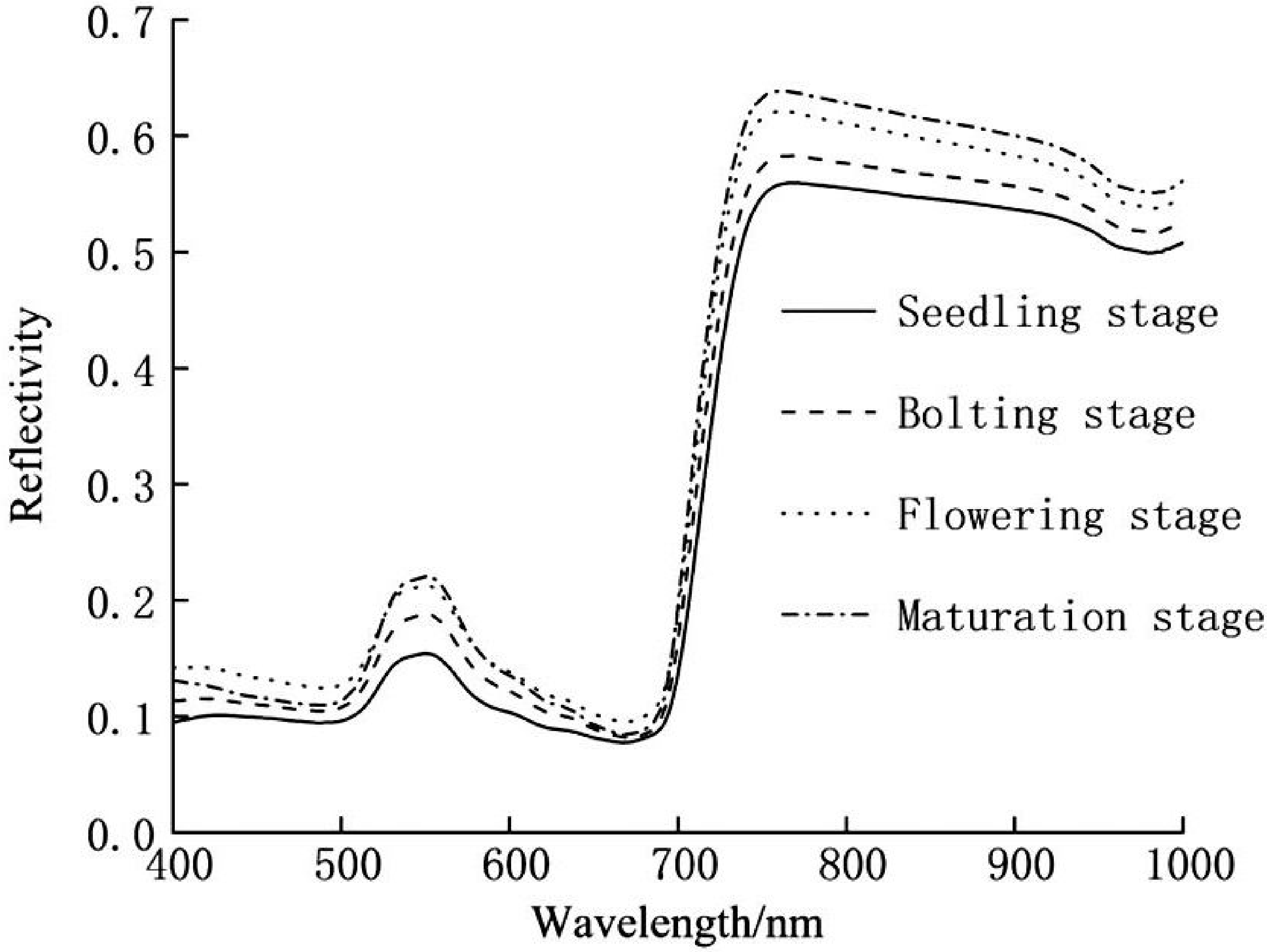

Figure 3

Hyperspectral reflectance curve of rape leaves at different growth stages.

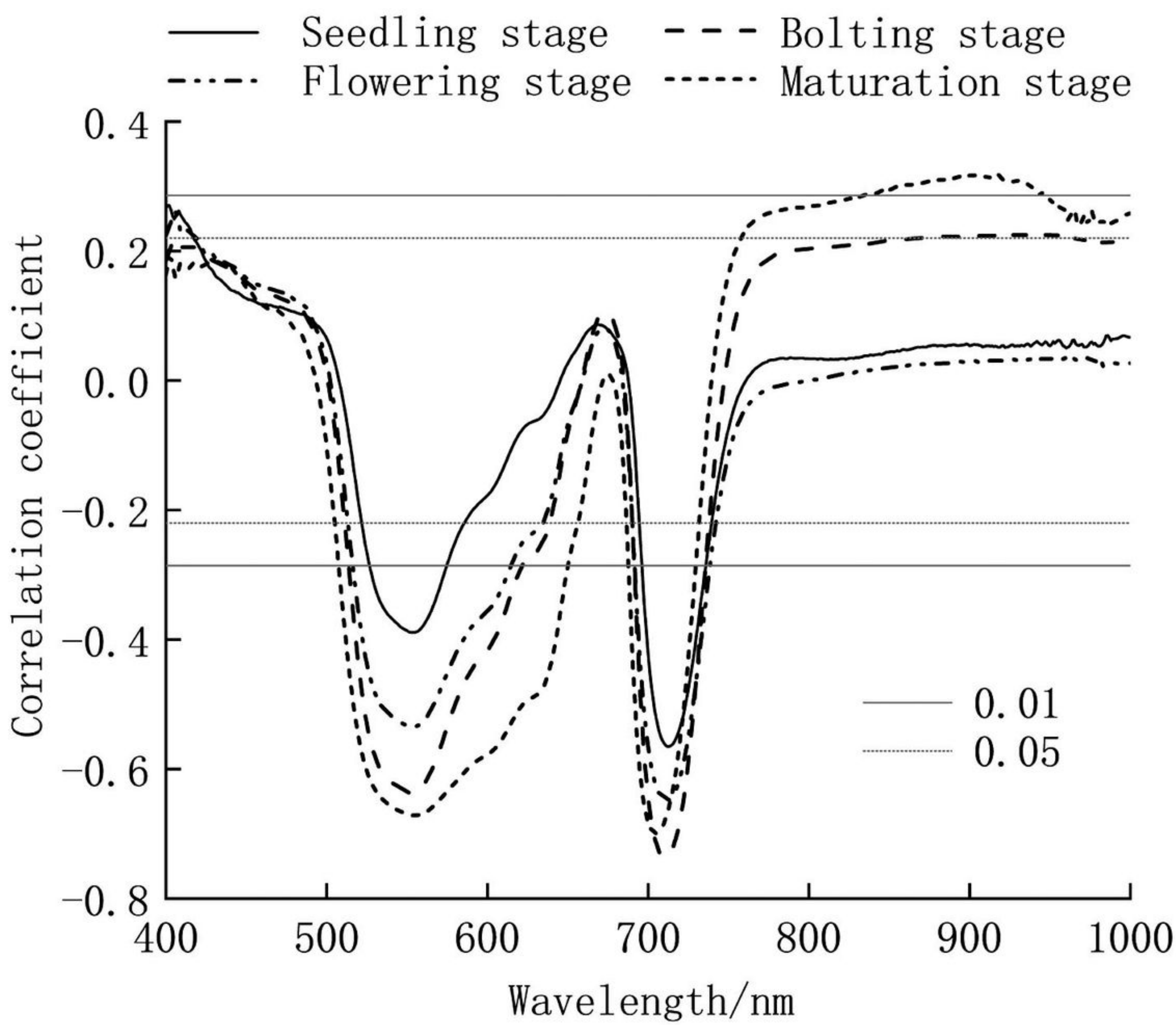

Figure 4

Correlation coefficient between SPAD value and original spectrum of rape leaves in different growth stages.

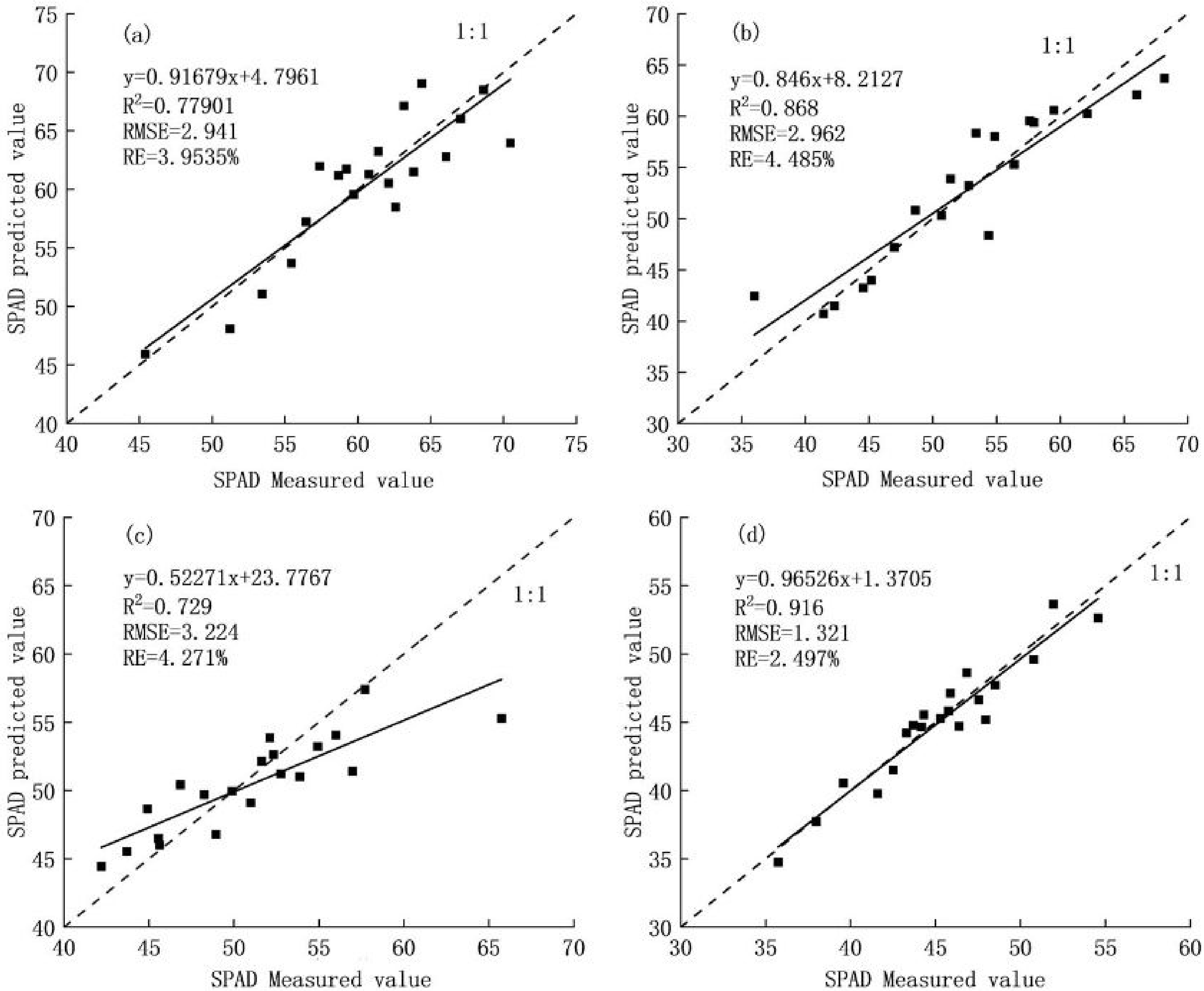

Figure 5

Distribution of estimated value and measured value of SPAD in validation sample rape leaves.

(a) stands for the seedling stage; (b) stands for the bud stage; (c) stands for the flowering stage; (d) stands for the mature stage.