

Remote estimation of rapeseed phenotypic traits under different crop conditions based on unmanned aerial vehicle multispectral images

Bo Duan,^a Xiaolu Xiao,^a Xiongze Xie,^b Fangyuan Huang,^a Ximin Zhi,^a and Ni Ma^{a,*}

^aChinese Academy of Agricultural Sciences, Oil Crops Research Institute, Wuhan, China

^bXiangyang Academy of Agricultural Sciences, Xiangyang, China

ABSTRACT. Rapeseed is an essential oil crop and the third major source of edible oil in the world. Accurate estimation of rapeseed phenotypic traits at field scale is important for precision agriculture to improve agronomic management and ensure edible oil supply. Unmanned aerial vehicle (UAV) remote sensing technology has been applied to estimate crop phenotypic traits at field scale. Machine learning is one of the main methods to develop estimation models for phenotypic traits based on UAV data. However, the accuracy and adaptability of machine learning estimation models are constrained by the representativeness of the training data. Here, we explored the influence of growth stage and crop conditions on the estimation of rapeseed phenotypic traits by machine learning and provided an optimized strategy to construct training data for improving the estimation accuracy. Four machine learning methods were employed, including partial least squares regression, support vector regression (SVR), random forest (RF), and artificial neural network (ANN), with SVR showing the best performance in estimating rapeseed phenotypic traits. The models established for a certain cultivar, planting site, or planting density had low estimation accuracies for other cultivars, planting sites, and planting densities during the entire growth period. The results showed that cultivar and planting site had an unquantifiable influence on phenotypic traits. Integration of stratified sampling and developing estimation models for different growth stages respectively can improve the estimation accuracy for different cultivars and planting sites during the entire growth period. Planting density exhibited a quantifiable influence on phenotypic traits, and the construction of training data with samples of both low and high planting densities could improve the estimation accuracy for different planting densities. Overall, optimization of the training data by considering the influence of crop conditions on phenotypic traits can improve the estimation accuracy of rapeseed phenotypic traits based on machine learning.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.18.018503](https://doi.org/10.1117/1.JRS.18.018503)]

Keywords: rapeseed phenotyping; growth estimation; optical remote sensing; crop conditions; machine learning

Paper 230334G received Jul. 6, 2023; revised Dec. 24, 2023; accepted Feb. 19, 2024; published Mar. 9, 2024.

1 Introduction

Rapeseed (*Brassica napus L.*) is one of the most important oil crops, as well as the third major source of edible oil in the world.¹ It has great nutritional value and functional properties, and it plays important roles in both food and non-food applications,² such as animal feed, biofuel, and

*Address all correspondence to Ni Ma, mani@caas.cn

medicine. Timely and accurate monitoring of rapeseed growth is of great significance for improving field management, which can help improve rapeseed yield and ensure the oil supply.

Phenotypic traits, such as plant height, leaf number, leaf area, and above-ground biomass, are complicated traits with great temporal and spatial variations and can indicate the growth status of the crop.³ The dynamic changes in some phenotypic traits can indicate the relationship between plant growth and the surrounding environment.⁴ However, traditional measurement of phenotypic traits mostly depends on time-consuming and labor-intensive manual work, which cannot easily be implemented over large areas.⁵ Remote sensing (RS) is a non-destructive measurement technology, with the advantages of high accuracy and high throughput and therefore the potential to effectively estimate crop phenotypic traits.^{6,7} In recent years, the unmanned aerial vehicle (UAV) technique has achieved considerable advancement, providing a novel platform for RS and making it possible to collect data with unprecedented spatial, spectral, and temporal resolution.⁸ Owing to its high spatial-temporal resolution,⁹ UAV RS has been widely used to estimate phenotypic traits for crop growth monitoring, particularly at the regional scale. For example, Ren et al.¹⁰ employed normalized difference spectral indices obtained by UAV hyperspectral data to estimate the above-ground biomass of winter wheat from flowering to maturity and then retrieved the spatial information on the crop harvest index with normalized root mean square error (RMSE) below 15%. Qiao et al.¹¹ estimated leaf area index (LAI) of maize by combining morphological parameters (canopy height, canopy coverage, and canopy volume) and vegetation index (VI); as a result, the overall accuracy of RMSE in the seedling stage, jointing stage, tasseling stage, silking stage, blister stage, and milk stage was 0.26; and they further used the dynamic change of LAI to evaluate the maize growth status. Therefore, UAV RS has become an effective tool for retrieving data on phenotypic traits for crop growth monitoring.

Two methods are generally employed to build the relationship between crop phenotypic traits and UAV images, namely the statistical method and the radiative transfer model. The radiative transfer model was developed on the basis of the interaction between solar radiation and plant tissues, thus possessing more flexibility as it involves physical mechanisms.^{12,13} However, the structure and input parameters of the radiative transfer model are rather complicated,¹⁴ making it difficult to realize and limiting its further application. In contrast, the statistical method (also called regression) consists of calibration of a numerical relationship between one or several ground-measured phenotypic traits and the features of UAV images,¹⁵ and is characterized by an easy process of development and operation. Linear and non-linear regression with a simple regression process are the most commonly used statistical methods, but they suffer from weak robustness and inferior estimation accuracy.^{16,17} Machine learning has been developing rapidly, providing a more advanced statistical method to establish the relationship between phenotypic traits and UAV image features. It functions as a “black box” with limited process-based interpretation¹⁸ and can characterize complex relationships between variables without an explicit equation. Compared with simple regression, machine learning is more robust and adaptive and can better utilize the vegetation information in UAV images.¹⁹ Machine learning has exhibited powerful performance in estimating phenotypic traits. Teodoro et al.²⁰ used four machine learning methods, including deep learning (DL), random forest (RF), support vector regression (SVR), and linear regression, to estimate the plant height and grain yield of soybean based on UAV multispectral data. They found that DL, RF, and SVR performed better than linear regression at early growth stages with the highest r value for plant height (0.77) and grain yield (0.44). Teshome et al.²¹ estimated sweet corn biomass at the entire growth stage by using RF, SVR, and k-nearest neighbor with UAV multispectral data and found that SVR outperformed other algorithms with an R^2 value of 0.77. As a more sophisticated statistical technique, machine learning promotes the application of statistical methods in the estimation of crop phenotypic traits based on UAV images. However, the estimation model developed by machine learning is typically trained by ground-measured phenotypic traits,²² and its performance is largely determined by the representativeness of the training data. Generally, a higher estimation accuracy requires the construction of training data and test data under the same or similar crop conditions at the same phenological stage. In this case, phenological stages and crop conditions, such as crop type, cultivar, planting density, and fertilizer level, have important impacts on the structure of the training and test data datasets, and play an important role in the development of estimation

models.^{23,24} Therefore, it is necessary to analyze the influence of phenological stage and crop conditions on the estimation of phenotypic traits by using machine learning approach.

Some studies have compared the estimation results of phenotypic traits under different crop conditions based on UAV images, such as rapeseed LAI at different nitrogen (N) levels,²⁵ rice N nutrition at different N levels and plant densities,²⁶ and the above-ground biomass of winter wheat at different N levels.²⁷ Obviously, most studies have been focused on the uncertainties of fertilizer treatment and involved few other crop conditions. However, except for fertilizer treatment, other cultivation practices (such as planting density), growth environment, and cultivar also have complex influence on phenotypic traits and deserve more attention.²⁸ In addition, crop growth is affected by multiple factors including the crop itself and growth environment.²⁹ Hence, the difference and interaction of different crop conditions on the estimation of phenotypic traits should be determined.

In this study, multispectral images of rapeseed under different crop conditions (planting site, planting density, and cultivar) were collected by UAV during the entire growth period, and the estimation models of rapeseed phenotypic traits were developed by four machine learning algorithms, including partial least squares regression (PLSR), SVR, RF, and artificial neural network (ANN), respectively. Then, the optimal algorithm was used to develop estimation models with different training data and test data to analyze the influence of growth stage and crop conditions on the estimation accuracy. Specifically, this study aims to (1) explore the influence of growth stage and crop conditions on the estimation of phenotypic traits by using machine learning approaches; (2) explain the mechanisms for the influence of growth stage and crop conditions; (3) provide an optimized strategy to improve the estimation accuracy of phenotypic traits.

2 Materials and Methods

2.1 Study Area and Experimental Design

The study involved two study areas, namely the rapeseed experiment base of oil crops research institute, Chinese Academy of Agricultural Sciences in Jingzhou City (30°14'39"N, 112°21'14" E) and Xiangyang City (31°53'32"N, 112°1'35"E), Hubei province, China (Fig. 1). Two field-plot experiments involving different rapeseed cultivars and planting densities were respectively conducted in these two areas in 2021. The experiment at Jingzhou involved two cultivars (Zhongyouza 19 and Dadi 199) and three planting densities (22.5, 45.0, and 67.5 plants/m²) and therefore six treatments altogether. Each treatment was conducted in a separate field plot, and the distance between two adjacent plots was 40 cm. Each treatment was repeated three times, and there were 18 plots in this experiment. The experiment was conducted twice at Jingzhou to induce plant growth difference with a sowing date of September 27 2021 and October 12 2021, respectively. As for Xiangyang, the experiment involved other two cultivars (Zhongyouza 19 and Flower type (FT) cultivar) and three planting densities (15.0, 45.0, and 75.0 plants/m²).

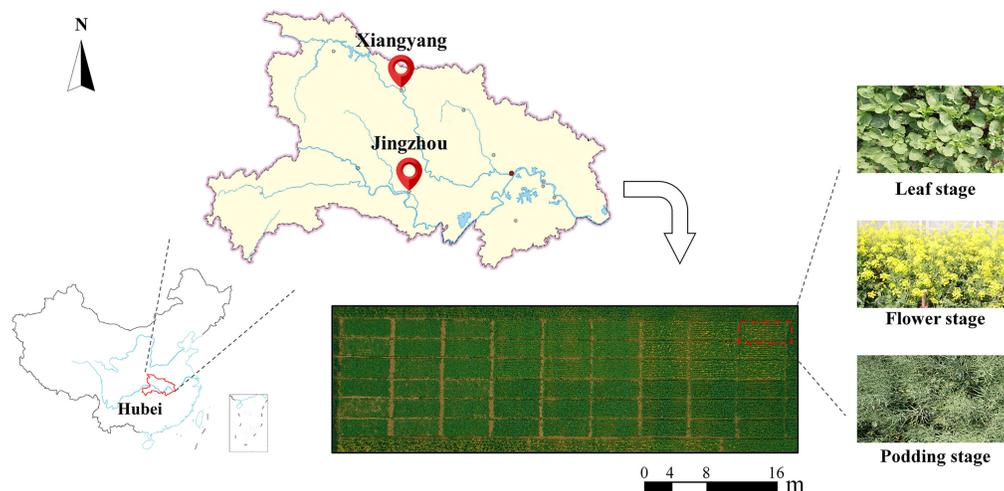


Fig. 1 Location of the study area.

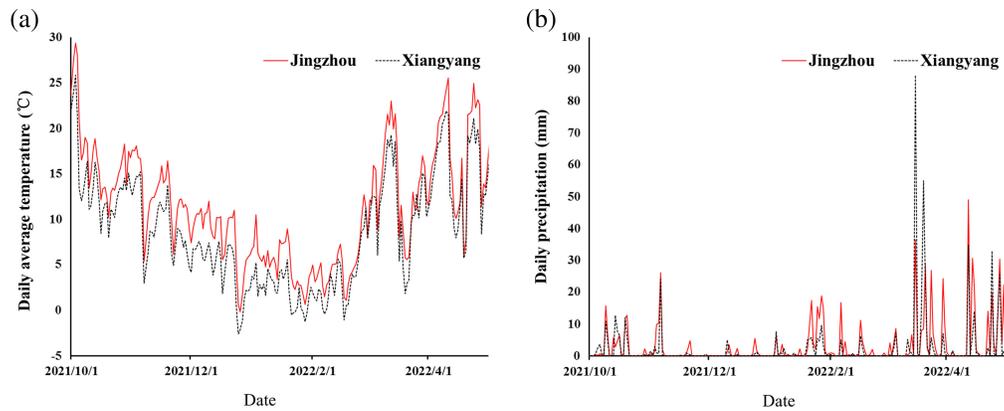


Fig. 2 Profile of (a) temperature and (b) precipitation during the entire growth period of rapeseed in two planting sites.

Table 1 Experimental design and UAV image collection date.

Planting site	Cultivars	Planting density (plants/m ²)	Growth stage	Image collection date (year/month/date)
Xiangyang	Zhongyouza 19	15.0	Leaf stage	2021/12/09
		45.0		2021/12/23
	Flower type	75.0	Flower stage	2022/01/19
				2022/03/13
				2022/03/27
Jingzhou	Zhongyouza 19	22.5	Leaf stage	2021/12/02
		45.0		2021/12/21
	Dadi 199	67.5	Flower stage	2022/01/18
				2022/03/10
				2022/03/10

Similarly, the experiment was conducted three times in Xiangyang with sowing dates of September 30 2021, October 15 2021, and October 30 2021, respectively. In general, there were 90 field plots in this study, including 36 plots in Jingzhou and 54 plots in Xiangyang. Except for the rapeseed cultivar and planting density, the field management of these experimental plots was the same. Given the uneven distribution of planting density, the planting densities of 15.0 and 22.5 plants/m² were considered as low, 45.0 plants/m² as medium, and 67.5 and 75.0 plants/m² as high. Therefore, this study involved two planting sites (Jingzhou and Xiangyang), three rapeseed cultivars [Zhongyouza 19 (ZY 19), Dadi 199 (DD 199) and FT cultivar], and three planting densities (low density, medium density, and high density). The profile of meteorological parameters during the rapeseed growth period in two planting sites is shown in Fig. 2.

UAV flight campaign was first carried out, and field measurement of rapeseed plants was immediately performed after the UAV flight. Planting site, rapeseed cultivar, planting density, and UAV image collection date of this study are summarized in Table 1. The workflow is presented in Fig. 3.

2.2 UAV Image Collection

This study employed a six-band sensor (MS600 Pro, Ysense, Qingdao City, China) to obtain the multispectral image of field plots as shown in Fig. 4(a). The multi-band sensor consists of

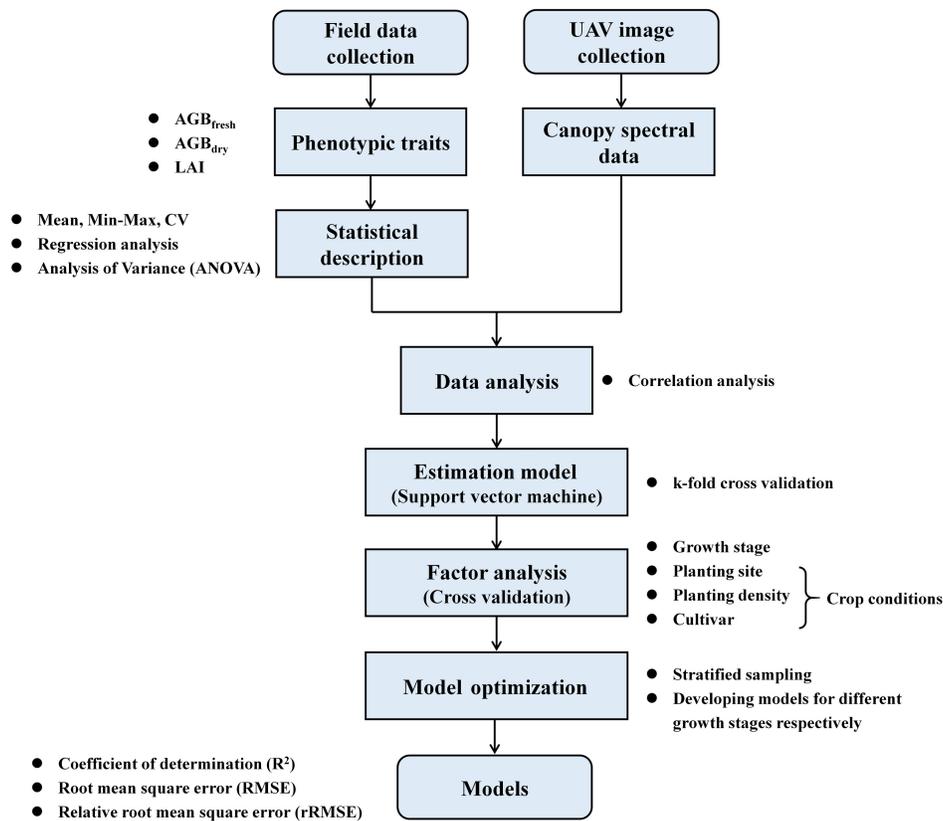


Fig. 3 Workflow of this study.

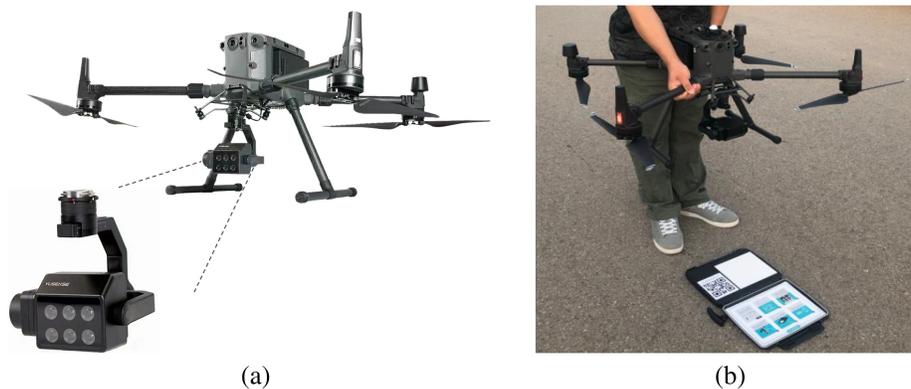


Fig. 4 (a) UAV and multispectral sensor used in this study and (b) the radiometric correction of multispectral images.

six individual miniature digital cameras with each camera equipped with a customer-specified band pass filter centered at the wavelength of 450, 555, 660, 720, 750, and 840 nm, respectively, which are sensitive bands of vegetation.²⁵ The sensor was attached to the UAV (Matrice 300, DJI Technology, Co., Ltd., Shenzhen City, China) by a gimbal to help compensate for UAV movement (pitch and roll) and guarantee close-to-nadir image collection during flight.³⁰ UAV flight campaigns were carried out between 10:00 and 14:00 local time when the changes of solar zenith angle were minimal and under a sunny sky to avoid the influence of cloud cover shadow. UAV flight altitude was set at 100 m, in which one exposure can cover all field plots of each experiment.



Fig. 5 Measurement of leaf area. (a) RGB image of rapeseed leaves and (b) leaf segmentation results.

2.3 Field Data Collection

A destructive sampling method was used to obtain the data of phenotypic traits of each field plot. Five rapeseed plants were randomly harvested with roots together in each field plot, placed into a polythene bag, and taken to the laboratory. The measurement was conducted as soon as possible after the plant samples arrived at the laboratory. First, the roots were cut off in the cotyledonary node. The plant tissues without roots were referred to as above-ground components, which were weighed immediately to obtain the above-ground fresh biomass. Then, the green leaves were separated from other components and spread on a horizontal table with a white background as shown in Fig. 5(a). A Canon camera (EOS 5D Mark II, Canon Inc., Tokyo, Japan) was utilized to take photos for rapeseed leaves and then an image segmentation algorithm was performed in MATLAB (MATLAB 2016a, MathWorks Inc., Natick, Massachusetts, USA) to calculate the area of leaves [Fig. 5(b)]. Next, the above-ground components were dried in an oven for 30 min at 105°C to deactivate enzymes and dried again at 80°C until constant weight. The oven-dried above-ground components were weighed to obtain the above-ground dry biomass. The average value of five plants was used as the plant-level phenotypic trait of each plot, and the plot-level trait was derived by the product of plant-level trait and planting density. Therefore, three phenotypic traits were obtained for each field plot, including above-ground fresh biomass (AGB_{fresh}), above-ground dry biomass (AGB_{dry}), and LAI

$$LAI = LA \times d, \quad (1)$$

$$AGB = x \times d, \quad (2)$$

where LA and x are the plant-level leaf area and above-ground biomass (fresh or dry biomass) in each plot, respectively, d is planting density with the unit of plants per square meter, and the unit of AGB is weight per unit area (g/m^2).

2.4 UAV Image Processing

The process to obtain multispectral images by MS600 Pro included band-to-band registration, radiometric correction, VI calculation, and determination of the region of interest (ROI). MS600 Pro consisted of six individual cameras that could take six images simultaneously with one single exposure. Due to the spatial distribution difference of the six cameras, there was significant camera misregistration among the images of different bands. To remove the effect of misregistration, band-to-band registration was performed in the built-in software (Yusense Map, Yusense, Qingdao City, China) of MS600 pro and corresponding pixels of each band were spatially overlapped in the same focal plane. In addition, the pixel value of the raw image obtained by MS600 Pro is a digital number (DN) and the radiometric correction is necessary to transform DN into reflectance. A calibration target with a constant reflectance of 0.60 was used for radiometric correction. Before each UAV flight, the calibration target was posed on the ground for the sensor to take radiometric reference images [Fig. 4(b)]. These images were imported into Yusense Map for radiometric correction and the reflectance images of different bands were exported to calculate VIs (Table 2). To extract VI value of each field plot, a maximum square was defined for each plot as ROI in the UAV image, and the plot-level VI was retrieved by averaging all the

Table 2 Vegetation indices used in this study.

Vegetation index	Formula	Reference
Red edge chlorophyll index ($CI_{rededge}$)	$R_{840\text{ nm}}/R_{750\text{ nm}} - 1$	Gitelson et al. ¹⁷
Green chlorophyll index (CI_{green})	$R_{840\text{ nm}}/R_{555\text{ nm}} - 1$	Gitelson et al. ¹⁷
Normalized difference vegetation index (NDVI)	$(R_{840\text{ nm}} - R_{660\text{ nm}})/(R_{840\text{ nm}} + R_{660\text{ nm}})$	Rouse et al. ³¹
Normalized difference red edge vegetation index (NDRE)	$(R_{840\text{ nm}} - R_{750\text{ nm}})/(R_{840\text{ nm}} + R_{750\text{ nm}})$	Glenn et al. ³²
Enhanced vegetation index (EVI)	$2.5(R_{840\text{ nm}} - R_{660\text{ nm}})/(R_{840\text{ nm}} + 6R_{660\text{ nm}} - 7.5R_{450\text{ nm}} + 1)$	Liu et al. ³³
Two-band enhanced vegetation index (EVI2)	$2.5(R_{840\text{ nm}} - R_{660\text{ nm}})/(R_{840\text{ nm}} + 2.4R_{660\text{ nm}} + 1)$	Jiang et al. ³⁴

per-pixel values within ROI. VI calculation was performed by band math function of ENVI 5.3 software (EXELIS; Boulder, Colorado, USA) with the reflectance image of different bands and ROI determination was performed by ROI tool.

2.5 Data Analysis and Model Development

Data analysis and model development of this study included statistical analysis of phenotypic traits, model generation and evaluation, and model optimization, which were all performed by MATLAB (MATLAB 2016a, MathWorks Inc., Natick, Massachusetts, USA).

2.5.1 Statistical analysis

The mean, minimum, maximum, and coefficient of variation (CV) of phenotypic traits were first calculated. Anderson–Darling test was used to determine whether the data of phenotypic traits follow normal distribution. Linear regression was used to analyze the relationship between different phenotypic traits, and analysis of variance (ANOVA) was used to analyze the differences in phenotypic traits under different crop conditions, including planting site, planting density, and cultivar.

2.5.2 Model construction and evaluation

The correlation of rapeseed phenotypic traits and UAV image features (canopy reflectance and VIs) was evaluated using Pearson correlation coefficient (r) before model construction. Generally, the image features with higher correlations with phenotypic traits were selected to develop estimation models. Four machine learning algorithms were used to develop the estimation models of rapeseed phenotypic traits from the selected image features, including PLSR,³⁵ SVR,³⁶ RF,³⁷ and ANN,³⁸ which have been employed by some studies to estimate crop phenotypic traits based on RS data. For example, Chen et al.³⁹ used PLSR, SVR, and RF to estimate plant nitrogen concentration of winter wheat with UAV hyperspectral data and found that SVR and RF performed better than PLSR with an R^2 above 0.8. Zhang et al.⁴⁰ used PLSR, SVR, RF, and ANN to estimate anthocyanins of apple tree leaves with ground hyperspectral data and found that the estimation accuracy varied significantly among these algorithms, but all of them showed satisfactory performance with R^2 from 0.85 to 0.95 in training data, and RF was relatively more accurate and stable. The performance of these four algorithms in the estimation of rapeseed phenotypic traits was evaluated by k-fold cross-validation. A 10-fold cross-validation methodology was used to randomly divide all the samples into ten groups with the same number of samples in each group. Nine groups were used as training data, while the remaining one group was used as test data. The process was repeated 10 times until each group was used as test data exactly one time to ensure the reliability of the tested models. After 10 iterations, the coefficient of

determination (R^2), RMSE, and relative root mean square error (rRMSE) were used to quantify the model accuracy

$$R^2 = \frac{\sum_{i=1}^K R_i^2}{K}, \quad (3)$$

where R_i^2 is the coefficient of determination in each tested group, and K is the iteration number of the cross-validation ($K = 10$ in this step)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n E_i^2}{n}}, \quad (4)$$

where E_i^2 is the estimation error of each sample, and n is the number of all samples involved in the cross-validation process

$$\text{rRMSE} = \frac{\text{RMSE}}{\text{mean}(\tilde{y})}, \quad (5)$$

where \tilde{y} is the ground measured value (true value) of all the samples involved in the cross-validation process.

Ten-fold cross-validation allows the estimation model to be trained on 90% of data (9/10) and tested on 10% of data (1/10), which can ensure the representativeness of the estimation model on the whole data. The algorithm with the highest estimation accuracy was selected as the optimal algorithm.

To determine the influence of growth stage and crop conditions (planting site, planting density, and cultivar) on the estimation of phenotypic traits, all samples were divided into different groups by these influencing factors, respectively, including three groups divided by growth stage (leaf stage, flower stage, and pod stage), two groups by planting site (Jingzhou and Xiangyang), three groups by planting density (low density, medium density, and high density), and three groups by cultivar (ZY 19, DD199, and FT).

A five-fold cross-validation method was used to randomly divide each of the Leaf stage, Flower stage, and Pod stage groups into five sub-groups, respectively. The sub-groups were used to train and test the SVR estimation models, and R^2 , RMSE, and rRMSE were calculated to evaluate the estimation accuracy of phenotypic traits. In addition, a cross-validation methodology was used to quantify the influence of cultivar, planting site, and planting density on the estimation accuracy. For each factor, each group was used as the test dataset exactly one time with other groups as the training dataset. Taking the factor of cultivar for example, the cross-validation included three repeats, in which one of ZY 19, DD199, and FT was used as the test dataset, and the remaining two groups were used as the training dataset. Similarly, the SVR estimation models based on canopy reflectance were trained and tested with different groups. R^2 , RMSE, and rRMSE of each tested group were calculated, respectively.

2.5.3 Model optimization

Stratified sampling and developing models for different growth stages respectively were conducted to improve the estimation accuracy of phenotypic traits. To reduce the influence of crop conditions, stratified sampling was used to create the training and test data. For each factor, 70% of the data in each group were randomly selected as the training data, while the other 30% of the data were used as the test data. The SVR estimation models were trained and tested to evaluate the estimation accuracy. Taking cultivar for example, 70% of the data in ZY19, DD199, or FT were selected respectively to create the training data [denoted by ZY19 (70%), DD 199 (70%), and FT (70%)], and other 30% of the data in these groups were used as the test data [denoted by ZY19 (30%), DD199 (30%), and FT (30%)]. R^2 , RMSE, and rRMSE of the test data were calculated.

K means clustering was conducted in all samples based on canopy reflectance and phenotypic traits respectively to explore the difference between the leaf and flower stage. The results were compared with the sample label (labeled by stage) to obtain the identification accuracy of the growth stage. To reduce the influence of growth stage, the estimation models were developed

Table 3 Statistical descriptions and Anderson–Darling test results of rapeseed phenotypic traits.

	Plots	Minimum value	Maximum value	Mean value	Coefficient of variation (%)	<i>p</i> value
AGB_{fresh} (g/m²)	342	780.01	33340.35	8043.84	68.3	< 0.05
AGB_{dry} (g/m²)	342	90.71	4613.21	945.30	76.5	< 0.05
LAI	288	0.80	21.38	5.22	61.6	< 0.05

for different growth stages, respectively, on the basis of stratified sampling. First, the SVR estimation model was constructed for different growth stages respectively with the training dataset. Then, a growth stage identification model was developed by support vector machine (SVM) from canopy reflectance using the same training data. Next, the stage identification model was applied in the test dataset to identify which stage the sample belongs to. Finally, the estimation models for different growth stages were applied respectively according to the stage identification. R^2 , RMSE, and rRMSE of the test data were calculated to evaluate the estimation accuracy of phenotypic traits.

3 Results

3.1 Statistical Analysis of Rapeseed Phenotypic Traits

The statistical descriptions of AGB_{fresh}, AGB_{dry}, and LAI are shown in Table 3. Generally, the three phenotypic traits exhibited discrete distributions during the entire growth period of rapeseed with a coefficient of variation (CV) above 60%. In addition, Anderson–Darling test revealed that they all followed a normal distribution with *p* values below 0.05. Subsequently, regression

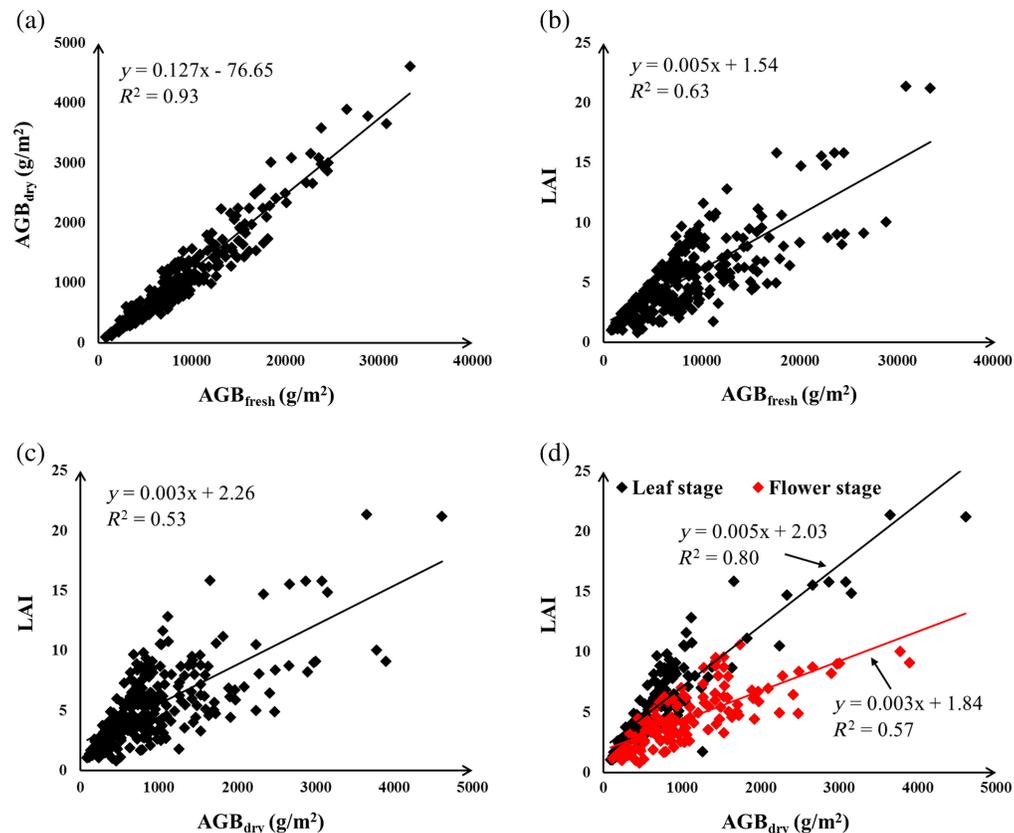


Fig. 6 Relationships of different phenotypic traits in rapeseed. The linear relationship of (a) AGB_{fresh} and AGB_{dry}, (b) AGB_{fresh} and LAI, (c) AGB_{dry} and LAI during the entire growth period; (d) The linear relationship of AGB_{dry} and LAI at the leaf stage and flower stage.

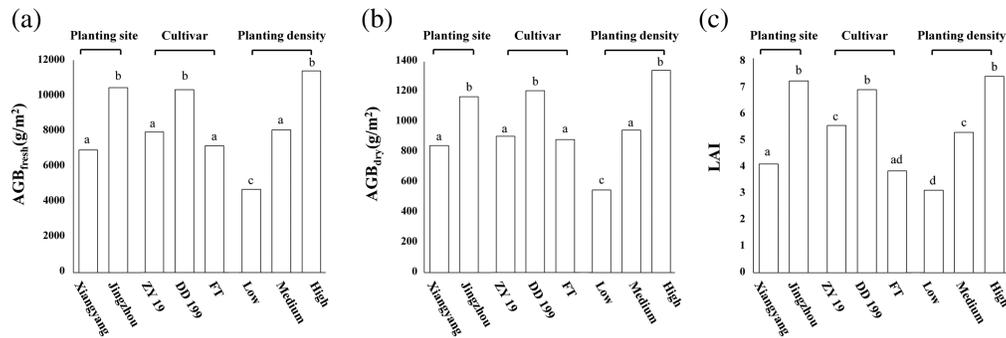


Fig. 7 ANOVA results of (a) AGB_{fresh} , (b) AGB_{dry} , and (c) LAI under different crop conditions. The same lower-case letter denotes no significant difference in the tested groups.

analysis was used to explore the relationship of different phenotypic traits. AGB_{fresh} showed a strong linear correlation with AGB_{dry} ($R^2 = 0.93$), but a relatively weak linear correlation with LAI ($R^2 = 0.63$) during the entire growth period of rapeseed [Figs. 6(a) and 6(b)]. Similarly, AGB_{dry} also showed a weak linear correlation with LAI ($R^2 = 0.53$) during the entire growth period [Fig. 6(c)]. AGB_{dry} and LAI showed different linear relationships at leaf and flower stages, and the linear relationship was stronger at the leaf stage ($R^2 = 0.80$) [Fig. 6(d)]. Therefore, the above-ground biomass and LAI showed different relationships at different growth stages of rapeseed.

The ANOVA results showed that AGB_{fresh} , AGB_{dry} , and LAI were significantly different for different planting sites, planting densities, and cultivars (Fig. 7). These three traits showed higher values in Jingzhou than in Xiangyang, and increased with planting density. As for different cultivars, ZY19 and FT had similar AGB_{fresh} and AGB_{dry} , both of which were lower than those of DD 199. In contrast, LAI was significantly different among the three cultivars, with DD 199 showing the highest value, followed by ZY 19 and then FT. Therefore, crop conditions, such as cultivar, planting site, and planting density, all had an important influence on rapeseed phenotypic traits.

3.2 Estimation Model Construction and Evaluation

The Pearson correlation coefficients (r) of AGB_{fresh} , AGB_{dry} , and LAI with canopy reflectance and VI are presented in Table 4. Generally, the above-ground biomass showed a stronger correlation with canopy reflectance, while LAI exhibited a higher correlation with VI. AGB_{fresh} and AGB_{dry} showed the highest correlation with the reflectance of 450 nm band ($R_{450\text{ nm}}$) ($r = 0.34$ and 0.31 , respectively). LAI displayed the highest correlation with $CI_{rededge}$ ($r = 0.42$). However, the phenotypic traits showed weak correlations with canopy reflectance and VI (r value below 0.50).

Four machine learning methods were employed to develop the estimation models of phenotypic traits from canopy reflectance (Table 5). The comparison of rRMSE showed that the highest estimation accuracy of AGB_{fresh} was $R^2 = 0.45$, $RMSE = 4095.78\text{ g/m}^2$, and

Table 4 Pearson correlation coefficients of phenotypic traits with canopy reflectance and vegetation index.

	Reflectance						Vegetation index					
	$R_{450\text{ nm}}$	$R_{555\text{ nm}}$	$R_{660\text{ nm}}$	$R_{720\text{ nm}}$	$R_{750\text{ nm}}$	$R_{840\text{ nm}}$	NDVI	NDRE	$CI_{rededge}$	CI_{green}	EVI	EVI2
AGB_{fresh}	-0.34**	0.10	0.15**	-0.15**	-0.17**	-0.11*	-0.16**	0.10	0.11*	-0.07	-0.25**	-0.18**
AGB_{dry}	-0.31**	0.15**	0.19**	-0.13**	-0.19**	-0.13*	0.22**	0.05	0.06	-0.15**	-0.30**	-0.23**
LAI	-0.06	-0.31**	-0.28**	-0.31**	0.04	0.07	0.30**	0.42**	0.42**	0.40**	0.24**	0.28**

*Correlation is significant at the 0.05 level (two-tailed).

**Correlation is significant at the 0.01 level (two-tailed).

Table 5 Estimation accuracy of phenotypic traits by different algorithms based on canopy reflectance.

Algorithm	AGB _{fresh} (g/m ²)			AGB _{dry} (g/m ²)			LAI		
	R ²	RMSE	rRMSE (%)	R ²	RMSE	rRMSE (%)	R ²	RMSE	rRMSE (%)
PLSR	0.19	4924.77	61.3	0.20	646.43	68.5	0.18	2.91	56.1
SVR	0.45	4095.78	50.9	0.49	517.70	54.8	0.40	2.52	48.2
RF	0.40	4258.58	52.9	0.44	540.53	57.2	0.27	2.74	52.6
ANN	0.44	4121.73	52.2	0.46	529.85	57.0	0.40	2.50	48.0

rRMSE = 50.9%; that of AGB_{dry} was $R^2 = 0.49$, RMSE = 517.70 g/m², and rRMSE = 54.8%; and that of LAI was $R^2 = 0.40$, RMSE = 2.52 g/m², and rRMSE = 48.2%. Generally, SVR could achieve the highest estimation accuracy for AGB_{fresh}, AGB_{dry}, and LAI, and therefore was the optimal algorithm in this study to construct estimation models for phenotypic traits based on canopy reflectance.

3.3 Factors Influencing the Estimation Accuracy

To determine the influence of growth stage on phenotypic trait estimation, the SVR estimation models were developed for different growth stages, respectively (Table 6). Compared with the models established for the whole growth period (Table 5), the models developed for different growth stages showed higher estimation accuracies of AGB_{fresh}, AGB_{dry}, and LAI. Generally, the estimation accuracy for these three phenotypic traits was the highest at the pod stage, followed by the flower stage, while the estimation accuracy at the leaf stage was relatively lower (with higher rRMSE). Therefore, the influence of growth stage should be considered in the estimation of phenotypic traits during the entire growth period.

To determine the influence of crop conditions on phenotypic trait estimation, the SVR estimation models were further trained and tested with different datasets (Table 7). The estimation accuracy of AGB_{fresh}, AGB_{dry}, and LAI was low for different planting sites with R^2 below 0.10 and rRMSE of 60.7% – 86.8%, while the estimation accuracy was relatively higher for different cultivars with R^2 of 0.20 – 0.54 and rRMSE of 46.1% – 65.3%. As for different planting densities, the estimation accuracy of low density and high density was low with rRMSE above 50%, while higher for medium density with R^2 of 0.43 – 0.49 and rRMSE of 36.9% – 45.2%. Notably, the three phenotypic traits were overestimated for low density and underestimated for high density (Fig. 8). In addition, the average estimation accuracy of different cultivars was

Table 6 Estimation accuracies of phenotypic traits at different rapeseed growth stages.

Growth stage	Phenotyping trait	R ²	RMSE	rRMSE (%)
Leaf stage	AGB_{fresh} (g/m²)	0.62	3247.81	48.8
	AGB_{dry} (g/m²)	0.62	410.69	56.5
	LAI	0.57	2.42	42.5
Flower stage	AGB_{fresh} (g/m²)	0.51	3924.93	39.7
	AGB_{dry} (g/m²)	0.57	475.05	42.5
	LAI	0.43	1.78	38.6
Pod stage	AGB_{fresh} (g/m²)	0.69	2742.12	34.6
	AGB_{dry} (g/m²)	0.69	424.10	35.4
	LAI	—	—	—

Table 7 Estimation accuracies of phenotypic traits in different training and test datasets.

Controlled factor of sample division	Training data	Test data	Phenotypic trait	Estimation accuracy of test data			
				R^2	RMSE	rRMSE (%)	
Planting site	Xiangyang	Jingzhou	AGB _{fresh} (g/m ²)	0.02	8090.125	77.4	
			AGB _{dry} (g/m ²)	0.08	1014.61	86.8	
			LAI	0.06	4.34	60.7	
	Jingzhou	Xiangyang	AGB _{fresh} (g/m ²)	0.05	4657.78	67.2	
			AGB _{dry} (g/m ²)	0.08	619.02	73.5	
			LAI	0.10	2.56	63.2	
Cultivar	DD 199	ZY 19	AGB _{fresh} (g/m ²)	0.35	4245.27	53.5	
			AGB _{dry} (g/m ²)	0.39	503.04	55.6	
			LAI	0.35	2.79	50.7	
	ZY 19	DD 199	AGB _{fresh} (g/m ²)	0.32	5602.88	54.3	
			AGB _{dry} (g/m ²)	0.38	787.96	65.3	
			LAI	0.24	3.15	46.1	
	ZY 19	FT	AGB _{fresh} (g/m ²)	0.41	3785.97	52.9	
			AGB _{dry} (g/m ²)	0.54	481.37	54.5	
			LAI	0.20	1.90	49.9	
	Density	Medium density	Low density	AGB _{fresh} (g/m ²)	0.46	4470.15	95.0
				AGB _{dry} (g/m ²)	0.45	511.08	93.2
				LAI	0.50	2.70	87.5
Low density		Medium density	AGB _{fresh} (g/m ²)	0.43	3447.16	42.8	
			AGB _{dry} (g/m ²)	0.49	426.76	45.2	
			LAI	0.43	1.93	36.9	
Low density		High density	AGB _{fresh} (g/m ²)	0.44	6309.13	55.4	
			AGB _{dry} (g/m ²)	0.54	818.76	61.0	
			LAI	0.33	3.90	53.2	

$R^2 = 0.37$, RMSE = 4344.12 g/m², and rRMSE = 54.0% for AGB_{fresh}, $R^2 = 0.43$, RMSE = 551.21 g/m², rRMSE = 58.3% for AGB_{dry}, $R^2 = 0.34$, RMSE = 2.62, and rRMSE = 50.2% for LAI; and that of different planting densities was $R^2 = 0.22$, RMSE = 4887.75 g/m², and rRMSE = 60.8% for AGB_{fresh}, $R^2 = 0.29$, RMSE = 609.29 g/m², and rRMSE = 64.5% for AGB_{dry}, and $R^2 = 0.16$, RMSE = 2.96, rRMSE = 56.7% for LAI, which were lower than those of ten-fold cross-validation in Table 5. Generally, the estimation accuracy of rapeseed phenotypic traits depends on the training and test data, and the estimation model trained under a certain crop condition may have poor performance under other crop conditions, such as different planting sites, planting densities, and cultivars. In this study, the estimation of AGB_{fresh}, AGB_{dry}, and LAI suffered most from the influence of planting site and cultivar, followed by planting density.

3.4 Optimization of the Estimation Model

To improve the estimation accuracy of phenotypic traits under different crop conditions, stratified sampling was used to construct the training and test data (Table 8). The highest estimation

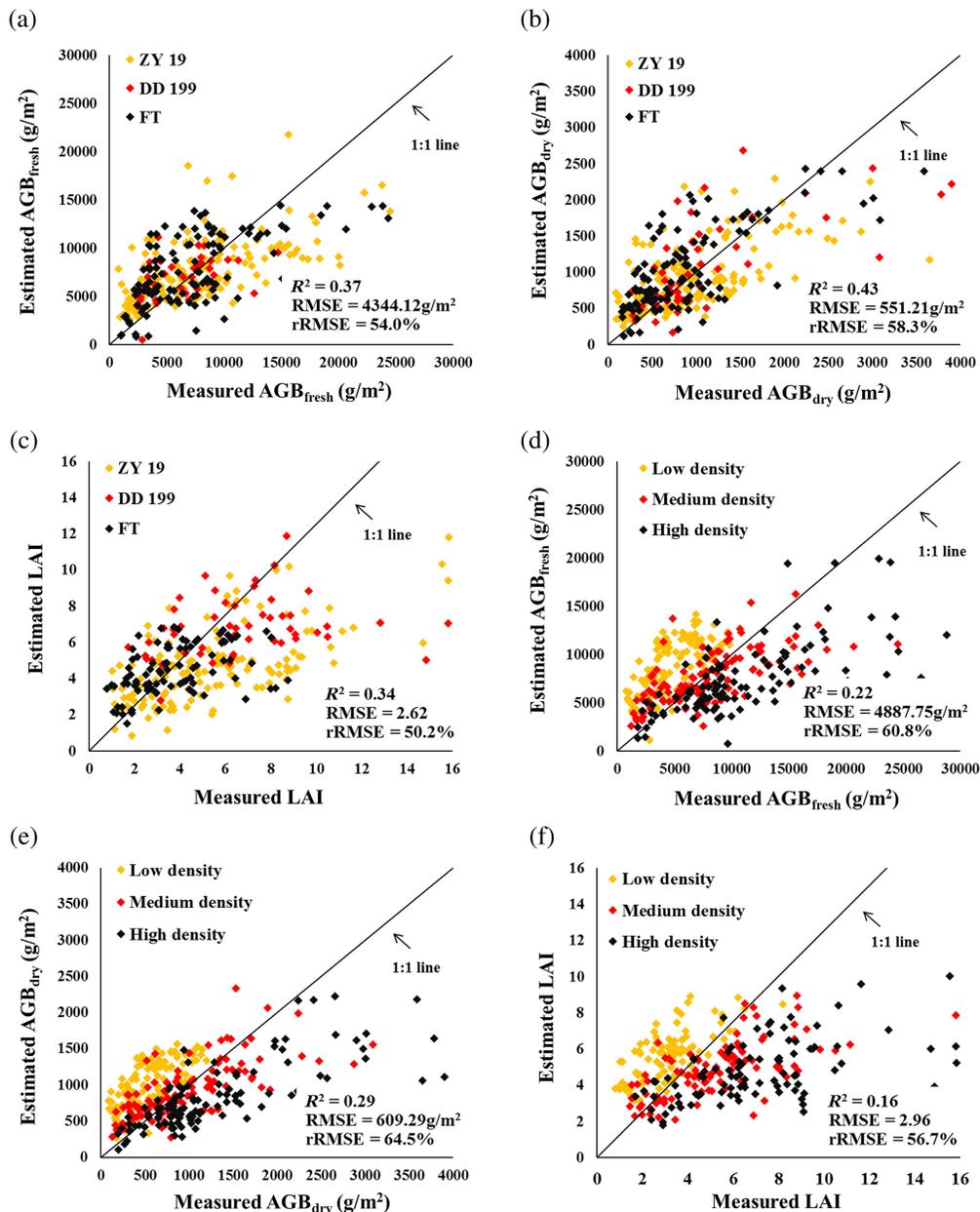


Fig. 8 Scatter plot between estimated values and ground-measured values of (a) AGB_{fresh} , (b) AGB_{dry} , and (c) LAI of different cultivars; (d) AGB_{fresh} , (e) AGB_{dry} and (f) LAI of different planting densities.

accuracy was $R^2 = 0.43$, RMSE = 3528.60 g/m², and rRMSE = 44.9% for AGB_{fresh} , $R^2 = 0.44$, RMSE = 425.37 g/m², and rRMSE = 47.2% for AGB_{dry} , and $R^2 = 0.48$, RMSE = 2.25, and rRMSE = 41.0% for LAI. Compared with random sampling and sampling by the influencing factors (Tables 5 and 7), stratified sampling improved the estimation accuracy of AGB_{fresh} , AGB_{dry} , and LAI for different planting sites and cultivars, but did not improve the estimation accuracy of medium density relative to that of low density and high density. Generally, stratified sampling can improve the estimation accuracy of rapeseed phenotypic traits for different planting sites and cultivars.

As shown in Table 9, *k*-means clustering indicated that the samples at the leaf and flower stage fell into two different clusters based on phenotypic traits and canopy reflectance with accuracies of 81.3% and 76.4%, respectively, while when all the leaf, flower, and pod stage were involved, the clustering accuracy was 59.4%. Therefore, the phenotypic traits and canopy

Table 8 Estimation accuracy of phenotypic traits by stratified sampling.

Controlled factor of sample division	Training data	Test data	Phenotypic trait	Estimation accuracy of test data		
				R^2	RMSE	rRMSE (%)
Planting site	Xiangyang (70%)	Xiangyang (30%)	AGB _{fresh} (g/m ²)	0.41	3958.90	49.4
	Jingzhou (70%)	Jingzhou (30%)	AGB _{dry} (g/m ²)	0.47	483.77	51.7
			LAI	0.43	2.66	48.7
Cultivar	ZY 19 (70%)	ZY 19 (30%)	AGB _{fresh} (g/m ²)	0.40	4307.84	52.1
	DD 199 (70%)	DD 199 (30%)	AGB _{dry} (g/m ²)	0.44	530.28	55.3
	FT (70%)	FT (30%)	LAI	0.41	2.64	46.1
Planting density	Low density (70%)	Low density (30%)	AGB _{fresh} (g/m ²)	0.43	3528.60	44.9
	Medium density (70%)	Medium density (30%)	AGB _{dry} (g/m ²)	0.44	425.37	47.2
	High density (70%)	High density (30%)	LAI	0.48	2.25	41.0

Table 9 Identification of rapeseed growth stages based on k-means clustering.

Input	Label	n	Accuracy (%)
AGB _{dry} , LAI	Leaf stage, Flower stage	288	81.3
$R_{450\text{ nm}}$, $R_{555\text{ nm}}$, $R_{660\text{ nm}}$, $R_{720\text{ nm}}$, $R_{750\text{ nm}}$, $R_{840\text{ nm}}$	Leaf stage, Flower stage	288	76.4
	Leaf stage, Flower stage, Pod stage	342	59.4

reflectance of rapeseed had significant differences between the leaf and flower stage, which can be identified by the canopy reflectance retrieved from UAV multispectral images.

Piecewise modeling at different stages was further employed to develop the estimation model based on stratified sampling (Table 10). The identification accuracy of rapeseed growth stages was high by using the SVM algorithm with accuracies above 85%. Compared with the models established for the whole growth period (Table 8), the models established for individual growth stages had higher estimation accuracies of AGB_{fresh}, AGB_{dry}, and LAI for different planting sites and cultivars, while almost the same performance for different planting densities. Therefore, integration of stratified sampling and developing models for different growth stages respectively could improve the estimation accuracy of rapeseed phenotypic traits for different planting sites and cultivars. As for different planting densities, the highest estimation accuracy was achieved with training data of low and high density and test data of medium density (Table 7).

4 Discussion

This study first determined the statistical characteristics of ground-measured phenotypic traits of rapeseed. AGB_{dry} showed a high linear correlation with AGB_{fresh}, but a weak correlation with LAI during the entire growth period of rapeseed (Fig. 6). Similarly, Duan et al.⁴¹ found that rice AGB has a weak correlation with LAI during the entire growth period and inferred that the weak correlation is caused by changes of structure in rice canopy after the heading stage, when the main element of canopy changes from leaves to both leaves and panicles. Compared with those in rice, the structural changes in rapeseed canopy are more significant during the entire growth period. Rondanini et al.⁴² suggested that rapeseed has a complex developmental phenotypic pattern as it evolves from an initial rosette to the main stem elongation and then indeterminate growth of floral raceme (Fig. 1). Due to the disturbance of flowers, the correlation between

Table 10 Estimation accuracy of phenotypic traits by integrating stratified sampling and developing models respectively for different growth stages.

Controlled factor of sample division	Training data	Test data	Identification accuracy of stage	Phenotyping traits estimation accuracy			
				Traits	R^2	RMSE	rRMSE (%)
Planting site	Xiangyang (70%)	Xiangyang (30%)	89.4%	AGB _{fresh} (g/m ²)	0.49	3251.51	43.9
	Jingzhou (70%)	Jingzhou (30%)		AGB _{dry} (g/m ²)	0.52	373.93	45.4
				LAI	0.29	2.29	48.4
Cultivar	ZY 19 (70%)	ZY 19 (30%)	94.0%	AGB _{fresh} (g/m ²)	0.54	3554.78	42.7
	DD 199 (70%)	DD 199 (30%)		AGB _{dry} (g/m ²)	0.59	402.95	44.5
	FT (70%)	FT (30%)		LAI	0.43	2.27	41.2
Density	Low density (70%)	Low density (30%)	98.8%	AGB _{fresh} (g/m ²)	0.40	3895.56	44.2
	Medium density (70%)	Medium density (30%)		AGB _{dry} (g/m ²)	0.60	377.57	49.0
	High density (70%)	High density (30%)		LAI	0.30	1.89	41.1

AGB_{dry} and LAI was weak at the flower stage and decreased during the entire growth period. In addition, the structure of the rapeseed canopy also depends on the plant growth characteristics and planting density. The growth characteristics of rapeseed plants are related to the inherent quality of plants (such as plant type and cultivar) and environment (such as temperature, precipitation, and sunshine). Therefore, AGB_{fresh}, AGB_{dry}, and LAI of rapeseed showed significant differences under different crop conditions (Fig. 7), which is consistent with the findings of previous studies. Li et al.⁴³ found that natural variations in silique number in rapeseed cultivars lead to differences in biomass accumulation. Wenyu et al.⁴⁴ reported that rapeseed AGB differs among different cultivars and planting densities. Based on Eqs. (1) and (2), these three population traits were positively correlated with the single plant traits and planting density. Rapeseed plants generally have stronger growth at higher temperature,⁴⁵ and thus AGB_{fresh}, AGB_{dry}, and LAI of Jingzhou were higher than those in Xiangyang (Fig. 2). Similarly, the three traits increased with planting density, and the increment was evenly coupled with the increase in planting density level. Therefore, rapeseed phenotypic traits are influenced by the growth stage, cultivar, planting density, and planting site. Since the performance of estimation models developed by machine learning depends on the representativeness of the training data, it can be inferred that the estimation accuracy is also influenced by these factors.

Our results showed weak correlations of phenotypic traits with canopy reflectance and VI (r below 0.5) (Table 4), which is inconsistent with the findings of previous studies. In the studies of Peng et al.²⁵ and Liu et al.,⁴⁶ VI showed a high correlation with LAI with r above 0.5 with the same rapeseed cultivar under different nitrogen fertilizer treatments at the leaf stage. In contrast, this study involved different growth stages, cultivars, planting densities, and planting sites, and these multiple factors may together cause the weak correlation. Notably, VI showed no obviously stronger correlation with LAI than canopy reflectance, and AGB even had a weaker correlation with VI than with canopy reflectance. Therefore, the canopy reflectance obtained by UAV multispectral images was directly used to develop the estimation models of phenotypic traits by using machine learning. Wittenberghe et al.⁴⁷ suggested that the information to predict a leaf parameter of trees by machine learning should not be restricted to one or a few spectral bands, and more bands should be taken into account to reduce the influence of data noise. To make better use of the spectral bands, the canopy reflectance of all six bands was utilized to develop the estimation models. The results of ten-fold cross-validation showed that SVR was

the optimal algorithm to develop the estimation models, which is consistent with the reports of Teodoro et al.²⁰ and Teshome et al.²¹

The estimation accuracy of AGB_{fresh} , AGB_{dry} , and LAI during the entire growth period was low (Table 5), and development of estimation models for different stages respectively greatly improved the estimation accuracy (Table 6). Similarly, Fang et al.⁴⁸ found that the estimation accuracy of vegetation fraction in rapeseed with UAV data was low when the models were developed for the leaf stage and flower stage together, and the accuracy was improved when the model was developed respectively for the two stages. Besides, a low accuracy was found in the estimation of AGB_{fresh} , AGB_{dry} , and LAI for different cultivars, planting sites, and planting densities (Table 7). Liang et al.⁴⁹ revealed that the estimation accuracy of leaf nitrogen content decreased for different rice cultivars. They found that dividing the cultivars into early and late maturation type and then developing estimation models respectively for the two types could improve the accuracy. Similarly, stratified sampling, which constructs training data and test data with the same crop conditions (cultivar, planting site, or planting density), improved the estimation accuracy of rapeseed phenotypic traits in this study (Table 8). Their results indicated that developing estimation models for different growth stages respectively and construction of training data and test data by stratified sampling may improve the estimation accuracy of rapeseed phenotypic traits under different crop conditions during the entire growth period.

Fang et al.⁴⁸ proposed a threshold segmentation method of VI to identify the leaf stage and flower stage of rapeseed, and then the model was automatically selected for the leaf stage or flower stage to estimate vegetation fraction. In this study, *k*-means clustering indicated that both phenotypic traits and canopy reflectance could distinguish the leaf and flower stage into two clusters (Table 9), which means that the leaf and flower stage can be automatically identified by canopy reflectance. Thus, a support vector machine classifier was used to identify the growth stage of rapeseed. Integration of stratified sampling and developing estimation models for different growth stages respectively improved the estimation accuracy of AGB_{fresh} , AGB_{dry} , and LAI for different cultivars and planting sites during the entire growth period. However, for different planting densities, this method showed no better performance than construction of training data with low and high density and test data with medium density. According to Eqs. (1) and (2), there was a nearly linear relationship between these three phenotypic traits and planting density. The reason may be that the effect of planting site and cultivar on phenotypic traits is hard to quantify, while that of planting density is quantifiable. These results indicated that construction of the training data by considering the growth stage and crop conditions helps improve the estimation accuracy of phenotypic traits under varying crop conditions during the entire growth period.

This study developed an optimized strategy for training data to improve the estimation accuracy of rapeseed phenotypic traits. The results indicated that the influence of growth stage and crop conditions on phenotypic traits needs to be considered when estimating phenotypic traits by UAV data and machine learning methods. Instead of utilizing more powerful methods or more image features to develop the estimation model, it may be more effective and simpler to construct the training data by considering both the growth stage and crop conditions. The results of this work can provide a novel solution for the accurate estimation of crop phenotypic traits from the perspective of data optimization. This approach may introduce an agricultural background to the estimation model developed by machine learning methods and provide a new perspective for the cooperation of agriculture and RS. Our future work will apply this approach in other crop species and crop conditions, and explore the time information of multi-temporal UAV data for improving the estimation accuracy of phenotypic traits.

5 Conclusions

This study explored the influence of growth stage and crop conditions on the estimation of rapeseed phenotypic traits by using machine learning and UAV data, and proposed an optimized strategy for constructing training data by considering the influence of the growth stage and crop conditions on phenotypic traits to improve the estimation accuracy. The experiments were conducted at Jingzhou and Xiangyang, Hubei province of China, which included different rapeseed cultivars and planting densities. UAV images and data of three phenotypic traits of AGB_{fresh} , AGB_{dry} , and LAI were collected during the entire growth period of rapeseed. The results showed

that growth stage and crop conditions have great influence on the phenotypic traits. Four machine learning methods, PLSR, SVR, RF, and ANN, were used to develop estimation models of rapeseed phenotypic traits based on canopy reflectance obtained by UAV multispectral images, with SVR showing the best performance. The models established for a certain cultivar, planting site, or planting density had low estimation accuracies for other cultivars, planting sites, and planting densities during the entire growth period. Integration of stratified sampling and developing estimation model for different growth stages respectively could improve the estimation accuracy for different cultivars and planting sites, and construction of training data with samples of both low and high planting densities could improve the estimation accuracy for different planting densities. Therefore, construction of training data according to the growth stage and crop conditions is important when using machine learning to estimate crop phenotypic traits with UAV data.

Disclosures

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Code and Data Availability

The datasets used in this study are available on request from the corresponding author.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 31971855), the Open Project of Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, China (Grant No. KF2020006) and the National Key Research and Development Project (Grant No. 2021YFD1600503).

References

1. Y. Chen et al., "Effects of phytase/ethanol treatment on aroma characteristics of rapeseed protein isolates," *Food Chem.* **431**, 137119 (2024).
2. A. Chmielewska et al., "Canola/rapeseed protein—nutritional value, functionality and food application: a review," *Crit. Rev. Food Sci. Nutr.* **61**(22), 3836–3856 (2021).
3. M. Watt et al., "Phenotyping: new windows into the plant for breeders," *Annu. Rev. Plant Biol.* **71**, 689–712 (2020).
4. L. Han et al., "Clustering field-based maize phenotyping of plant-height growth and canopy spectral dynamics using a UAV remote-sensing approach," *Front. Plant Sci.* **9**, 18 (2018).
5. V. Singh et al., "Unmanned aircraft systems for precision weed detection and management: prospects and challenges," *Adv. Agron.* **159**, 93–134 (2020).
6. M. Reba and K. C. Seto, "A systematic review and assessment of algorithms to detect, characterize, and monitor urban land change," *Remote Sens. Environ.* **242**, 111739 (2020).
7. D. Mandal et al., "Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data," *Remote Sens. Environ.* **247**, 111954 (2020).
8. W. H. Maes and K. Steppe, "Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture," *Trends Plant Sci.* **24**(2), 152–164 (2019).
9. H. Zhang et al., "A review of unmanned aerial vehicle low-altitude remote sensing (UAV-LARS) use in agricultural monitoring in China," *Remote Sens.* **13**(6), 1221 (2021).
10. J. Ren et al., "Dynamic harvest index estimation of winter wheat based on UAV hyperspectral remote sensing considering crop aboveground biomass change and the grain filling process," *Remote Sens.* **14**(9), 1955 (2022).
11. L. Qiao et al., "Improving estimation of LAI dynamic by fusion of morphological and vegetation indices based on UAV imagery," *Comput. Electron. Agric.* **192**, 106603 (2022).
12. B. Koetz et al., "Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics," *Remote Sens. Environ.* **95**(1), 115–124 (2005).
13. H. Kimm et al., "Deriving high-spatiotemporal-resolution leaf area index for agroecosystems in the US Corn Belt using Planet Labs CubeSat and STAIR fusion data," *Remote Sens. Environ.* **239**, 111615 (2020).
14. X. Li et al., "Compared performances of SMOS-IC soil moisture and vegetation optical depth retrievals based on Tau-Omega and Two-Stream microwave emission models," *Remote Sens. Environ.* **236**, 111502 (2020).

15. M. Weiss, F. Jacob, and G. Duveiller, "Remote sensing for agricultural applications: a meta-review," *Remote Sens. Environ.* **236**, 111402 (2020).
16. A. Viña et al., "Comparison of different vegetation indices for the remote assessment of green leaf area index of crops," *Remote Sens. Environ.* **115**(12), 3468–3478 (2011).
17. A. A. Gitelson et al., "Remote estimation of canopy chlorophyll content in crops," *Geophys. Res. Lett.* **32**(8) (2005).
18. Y. Cai et al., "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches," *Agric. Forest Meteorol.* **274**, 144–159 (2019).
19. T. Kattenborn et al., "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.* **173**, 24–49 (2021).
20. P. E. Teodoro et al., "Predicting days to maturity, plant height, and grain yield in soybean: a machine and deep learning approach using multispectral data," *Remote Sens.* **13**(22), 4632 (2021).
21. F. T. Teshome et al., "Unmanned aerial vehicle (UAV) imaging and machine learning applications for plant phenotyping," *Comput. Electron. Agric.* **212**, 108064 (2023).
22. A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review," *Comput. Electron. Agric.* **151**, 61–69 (2018).
23. R. Colombo et al., "Retrieval of leaf area index in different vegetation types using high resolution satellite data," *Remote Sens. Environ.* **86**(1), 120–131 (2003).
24. A. A. Gitelson, "Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation," *J. Plant Physiol.* **161**(2), 165–173 (2004).
25. Y. Peng et al., "Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications," *Agric. Forest Meteorol.* **271**, 116–125 (2019).
26. H. Zha et al., "Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning," *Remote Sens.* **12**(2), 215 (2020).
27. J. Yue et al., "Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices," *ISPRS J. Photogramm. Remote Sens.* **150**, 226–244 (2019).
28. P. Sharma et al., "Above-ground biomass estimation in oats using UAV remote sensing and machine learning," *Sensors* **22**(2), 601 (2022).
29. P. Trivedi et al., "Plant–microbiome interactions: from community assembly to plant health," *Nat. Rev. Microbiol.* **18**(11), 607–621 (2020).
30. D. Turner et al., "Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over Antarctic moss beds," *Remote Sens.* **6**(5), 4003–4024 (2014).
31. J. Rouse et al., "Monitoring vegetation systems in the great plains with ERTS," *NASA Spec. Publ.* **351**, 309–317 (1973).
32. G. Fitzgerald, D. Rodriguez, and G. O'Leary, "Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—the canopy chlorophyll content index (CCCI)," *Field Crops Res.* **116**(3), 318–324 (2010).
33. H. Q. Liu and A. Huete, "A feedback based modification of the NDVI to minimize canopy background and atmospheric noise," *IEEE Trans. Geosci. Remote Sens.* **33**(2), 457–465 (1995).
34. Z. Jiang et al., "Development of a two-band enhanced vegetation index without a blue band," *Remote Sens. Environ.* **112**(10), 3833–3845 (2008).
35. H. Abdi, "Partial least squares regression and projection on latent structure regression (PLS Regression)," *Wiley Interdiscipl. Rev. Comput. Stat.* **2**(1), 97–106 (2010).
36. M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Apress, Berkeley, California (2015).
37. A. Liaw and M. Wiener, "Classification and regression by random forest," *R News* **23**(23), 18–22 (2002).
38. D. H. Wolpert, "Stacked generalization," *Neural Netw.* **5**(2), 241–259 (1992).
39. X. Chen et al., "Estimation of winter wheat plant nitrogen concentration from UAV hyperspectral remote sensing combined with machine learning methods," *Remote Sens.* **15**, 2831 (2023).
40. Z. Zhang et al., "Estimation of anthocyanins in leaves of trees with apple mosaic disease based on hyperspectral data," *Remote Sens.* **15**, 1732 (2023).
41. D. Bo et al., "Remote estimation of grain yield based on UAV data in different rice cultivars under contrasting climatic zone," *Field Crops Res.* **267**(1), 108148 (2021).
42. D. P. Rondanini et al., "Physiological responses of spring rapeseed (*Brassica napus*) to red/far-red ratios and irradiance during pre- and post-flowering stages," *Physiol. Plantarum* **152**(4), 784–794 (2014).
43. S. Li et al., "A systematic dissection of the mechanisms underlying the natural variation of silique number in rapeseed (*Brassica napus* L.) germplasm," *Plant Biotechnol. J.* **18**(2), 568–580 (2020).
44. Z. Wenyu et al., "An aboveground biomass partitioning coefficient model for rapeseed (*Brassica napus* L.)," *Field Crops Res.* **259**(1), 107966 (2020).
45. S. Peng et al., "Rice yields decline with higher night temperature from global warming," *Proc. Natl. Acad. Sci.* **101**(27), 9971–9975 (2004).

46. Y. Liu et al., "Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from UAV multispectral images," *Comput. Electron. Agric.* **166**, 105026 (2019).
47. S. Wittenberghe et al., "Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset," *J. Photochem. Photobiol. B Biol.* **134**, 37–48 (2014).
48. S. Fang et al., "Remote estimation of vegetation fraction and flower fraction in oilseed rape with unmanned aerial vehicle data," *Remote Sens.* **8**(5), 416 (2016).
49. T. Liang et al., "Identification of high nitrogen use efficiency phenotype in rice (*Oryza sativa* L.) through entire growth duration by unmanned aerial vehicle multispectral imagery," *Front. Plant Sci.* **12**, 740414 (2021).

Biographies of the authors are not available.