

## Article

# Estimation of the Bio-Parameters of Winter Wheat by Combining Feature Selection with Machine Learning Using Multi-Temporal Unmanned Aerial Vehicle Multispectral Images

Changsai Zhang <sup>1</sup>, Yuan Yi <sup>2</sup>, Lijuan Wang <sup>3</sup>, Xuewei Zhang <sup>1</sup>, Shuo Chen <sup>1</sup>, Zaixing Su <sup>2</sup>, Shuxia Zhang <sup>4</sup> and Yong Xue <sup>5,\*</sup>

<sup>1</sup> School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China; tbh261@cumt.edu.cn (C.Z.)

<sup>2</sup> Jiangsu Xuhuai Regional Institute of Agricultural Science, Xuzhou 221131, China

<sup>3</sup> School of Geography, Geomatics and Planning, Jiangsu Normal University, Xuzhou 221116, China

<sup>4</sup> School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou 221116, China

<sup>5</sup> School of Computing and Mathematics, College of Science and Engineering, University of Derby, Kedleston Road, Derby DE22 1GB, UK

\* Correspondence: y.xue@derby.ac.uk

**Abstract:** Accurate and timely monitoring of biochemical and biophysical traits associated with crop growth is essential for indicating crop growth status and yield prediction for precise field management. This study evaluated the application of three combinations of feature selection and machine learning regression techniques based on unmanned aerial vehicle (UAV) multispectral images for estimating the bio-parameters, including leaf area index (LAI), leaf chlorophyll content (LCC), and canopy chlorophyll content (CCC), at key growth stages of winter wheat. The performance of Support Vector Regression (SVR) in combination with Sequential Forward Selection (SFS) for the bio-parameters estimation was compared with that of Least Absolute Shrinkage and Selection Operator (LASSO) regression and Random Forest (RF) regression with internal feature selectors. A consumer-grade multispectral UAV was used to conduct four flight campaigns over a split-plot experimental field with various nitrogen fertilizer treatments during a growing season of winter wheat. Eighteen spectral variables were used as the input candidates for analyses against the three bio-parameters at four growth stages. Compared to LASSO and RF internal feature selectors, the SFS algorithm selects the least input variables for each crop bio-parameter model, which can reduce data redundancy while improving model efficiency. The results of the SFS-SVR method show better accuracy and robustness in predicting winter wheat bio-parameter traits during the four growth stages. The regression model developed based on SFS-SVR for LAI, LCC, and CCC, had the best predictive accuracy in terms of coefficients of determination ( $R^2$ ), root mean square error (RMSE) and relative predictive deviation (RPD) of 0.967, 0.225 and 4.905 at the early filling stage, 0.912, 2.711  $\mu\text{g}/\text{cm}^2$  and 2.872 at the heading stage, and 0.968, 0.147  $\text{g}/\text{m}^2$  and 5.279 at the booting stage, respectively. Furthermore, the spatial distributions in the retrieved winter wheat bio-parameter maps accurately depicted the application of the fertilization treatments across the experimental field, and further statistical analysis revealed the variations in the bio-parameters and yield under different nitrogen fertilization treatments. This study provides a reference for monitoring and estimating winter wheat bio-parameters based on UAV multispectral imagery during specific crop phenology periods.



**Citation:** Zhang, C.; Yi, Y.; Wang, L.; Zhang, X.; Chen, S.; Su, Z.; Zhang, S.; Xue, Y. Estimation of the Bio-Parameters of Winter Wheat by Combining Feature Selection with Machine Learning Using Multi-Temporal Unmanned Aerial Vehicle Multispectral Images. *Remote Sens.* **2024**, *16*, 469. <https://doi.org/10.3390/rs16030469>

Academic Editors: Murilo M. Maeda, Mahendra Bhandari and Jinha Jung

Received: 16 November 2023

Revised: 18 January 2024

Accepted: 22 January 2024

Published: 25 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** unmanned aerial vehicle (UAV); multispectral; leaf area index; canopy chlorophyll; machine learning

## 1. Introduction

Winter wheat is a crucial grain crop that plays a pivotal role in global food security and agricultural sustainability. In recent years, the significance of winter wheat research

has been underscored by the growing challenges posed by climate change, population growth, and the need for sustainable agricultural practices [1,2]. Crop biophysical and biochemical parameters provide important information about various aspects of crop conditions that have direct implications for productivity. Leaf area index (LAI) is a key biophysical parameter for quantifying crop canopy structure and function. Previous studies have highlighted the significance of LAI data in enhancing estimates of crop yield and land–atmosphere carbon dioxide exchanges by updating state variables in process-based agroecosystem models [3–5]. Canopy chlorophyll content (CCC) is defined as the total chlorophyll content per unit ground area in a contiguous group of plants, serving as a valuable metric for estimating canopy nitrogen content, vegetation physiological status and gross primary production [6,7]. Different from other crop physiological and biochemical traits, leaf chlorophyll content (LCC) directly reflects the nutrition status of individual crop plants. LAI, LCC, and CCC serve as crucial phenotypic traits in crops, offering effective insights into crop growth, plant health and yield prediction [8,9]. Timely monitoring and accurate estimation of the bio-parameters are necessary for grasping wheat growth dynamics and offer guidance for field management.

Remote sensing techniques have found widespread application in estimating growth bio-parameters and grain yield across various experimental environments for precision agriculture practices [10]. The collection of bio-parameter data at the ground level typically involved labor-intensive and time-consuming manual processes conducted via pointwise sampling. Moreover, ground-measured data were often limited to a few sampling points, posing challenges in representing the traits of the entire crop field area, and thereby restricting the scope of traditional ground bio-parameter data [11]. Large-scale and high throughput data can be acquired by the satellite-based remote sensing technology; however, it is difficult to reveal detailed local features due to coarse spatial resolution. In this context, unmanned aerial vehicle (UAV) remote sensing technologies have emerged as a capable tool for mapping crop bio-parameters traits with fine spatial and temporal resolution.

Over the past few decades, various vegetation indices (VIs) have been proposed for spectral remotely sensed bio-parameters estimation to simplify predictive modeling [12–14]. However, the optimal VI relevant to bio-parameters varies depending on the crop growth stage, the range of variation in bio parameter or crop phenotype [15–20]. Consequently, using a single VI to calibrate a general-purpose model for the entire growing season may not accurately capture the variation in bio-parameters for crucial individual stages. Many studies emerged using the combinations of spectral bands or spectral indices as input variables, combined with multiple linear regression or machine learning (ML) for predictive modeling [21–25]. However, it should be noted that the highly correlated VIs will be generated in regression modeling due to the similarity of spectral index calculation formulas and spectral information when using VIs calculated by several spectral bands, especially for multispectral data with board-band [11]. The presence of data redundancy and multicollinearity among spectral variables will significantly diminish the stability and efficiency of model prediction.

Feature variable selection can adaptively select the optimal combination of variable candidates to match the ML model, reduce data dimensionality, and improve modeling accuracy and efficiency [26]. Therefore, many studies utilized feature selection to improve predictive modeling performance, and they can be divided into three categories: filter, wrapper, and embedded [27,28]. For embedding algorithms, variable selection is embedded into the model training process, and is achieved by determining high-importance score contributed to the model, such as LASSO [29], variable importance in projection based on partial least squares (PLS-VIP) [30] and various regression trees [31,32]. The filter-based algorithm, such as Pearson correlation coefficient thresholding, is most commonly used due to its simplicity. The selected variables by the filter-based algorithm can be explained easily for the dependent variable. The disadvantage is that it does not take into account the characteristics of the ML model and is more suitable for simple empirical regression algorithms. The wrapper algorithm treats the feature selection as a search problem and

evaluates the merits of feature variables through the evaluation function of the induction learner, which can select “tailor-made” variables for each model. The generation procedure for finding the optimal variable combination based on the wrapper includes forward or backward search, recursive feature selection (RFE), and bionic algorithms [33]. The wrapper algorithms are computationally more expensive than the filtering algorithm, due to repetitive training steps and cross-validation. However, the wrapper algorithms are more accurate than filtering algorithms. Wang [34] and Wang [35] estimated the wheat LCC using multiple ML models combined with the important ranking of the random forest model. Zhu [15] and Yin [36] used multiple MLs combined with filter-based and RFE feature selection to estimate wheat LCC at different growth stages, respectively. These studies indicated that using the ML model alone cannot achieve optimal model accuracy, and the combination of feature selection and ML model can more accurately estimate the LCC of winter wheat. However, the results in the above studies indicated that the number of feature variables selected by the RF or the RFE is still relatively large, which has not effectively achieved the goal of data dimensionality reduction, and there is still a problem of data redundancy for regression modeling.

Therefore, the primary objective of this study was to develop a machine learning regression modeling combined with an adapted variable selection scheme for estimating the bio-parameters of winter wheat at various growth stages. The specific objectives were (i) to examine changes in crop bio-parameters during the growth stages and the correlation with spectral variables; (ii) to compare and evaluate the combination of variable selection and machine learning estimation performance in monitoring bio-parameters traits during the growth stages; and (iii) to explore the variations in crop bio-parameters and grain yield under multi-fertilization treatments, with an aim to provide a reference and technical support for UAV remote sensing monitoring of crop bio-parameters with fertilizer management, thus boosting the applications of UAV multispectral remote sensing technologies in precision agriculture.

## 2. Materials and Methods

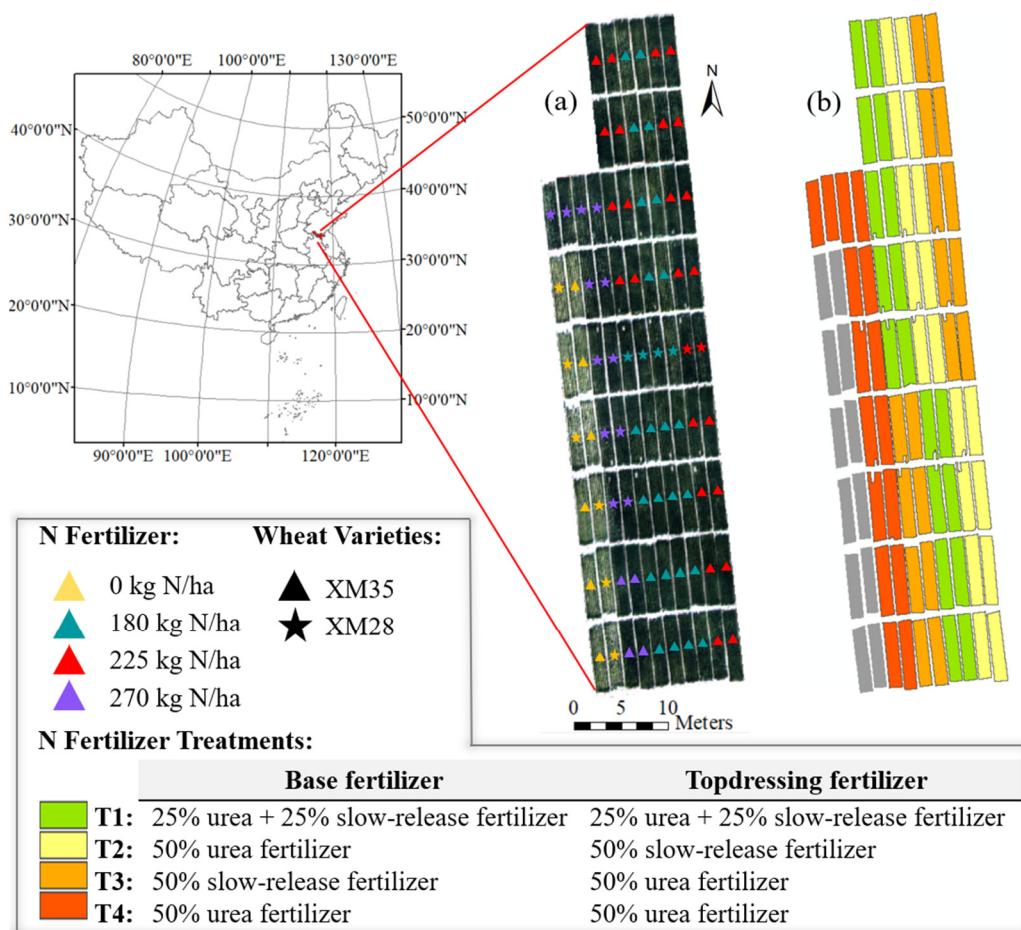
### 2.1. Study Site and Experimental Design

During the winter wheat growing season of 2022/2023 in Xuzhou, Jiangsu Province, China, the experimental study was conducted at the Jiangsu Xuhuai Regional Institute of Agricultural Science ( $33^{\circ}16'58''N$ ;  $117^{\circ}17'23''E$ , elevation 35 m a.s.l.). The experiment involved two local wheat varieties (XM35 and XM28) and four nitrogen fertilizer rates (0, 180, 225, 270 kg N/ha). The field experiment used a split-plot design; a total of 82 plots with  $7.5 \times 1.5 \text{ m}^2$  each (Figure 1). The treatment of nitrogen fertilizer in each plot was split into base and topdressing fertilizer in a proportion of 1:1. Four treatments of nitrogen fertilizer were applied before sowing and at the jointing stage. Irrigation applied natural rainfed field conditions and weed control followed local field management practices. Winter wheat was sown on 10 October 2022, and harvested on 11 June 2023, completing a 243-day life span. Measurements were conducted at four growth stages of winter wheat: late jointing (DAS 172), booting (DAS 185), heading (DAS 198), and early filling stage (DAS 214) (Table 1).

**Table 1.** Measurement date and corresponding growth stage of winter wheat.

Ground and UAV Measurement Data (2023)	Growth Stage Description	Abbreviation
31 March	Late Jointing Stage (DAS 172)	LJ
12 April	Booting Stage (DAS 184)	BS
26 April	Heading Stage (DAS 198)	HS
12 May	Early Filling Stage (DAS 214)	FS

Note: Days after sowing (DAS). Winter wheat was sown on 10 October 2022, and harvested on 11 June 2023, completing a 243-day life span.



**Figure 1.** Diagram of the winter wheat experimental site. (a) RGB image marked with multiple nitrogen fertilizer levels application, and (b) experimental design with multiple fertilizer treatments.

## 2.2. Data Collection

### 2.2.1. In Situ Measurements and Laboratory Processes

Growth-related bio-parameters, including the LAI, LCC, and CCC, were collected during the growing season (Table 1). Examples of ground photos reflecting crop growth status are shown in Figure 2, the photos were taken at a height of approximately one meter above the wheat canopy. The LAI and LCC measurements were conducted within a 1 m<sup>2</sup> area in each field plot. Each ground sample area's center position was recorded using a standard portable navigational equipment combining the Network Real-Time Kinematic technology (Network RTK). The LAI value was obtained using an LAI-2200C plant canopy analyzer (Plant Canopy Analyzer, LI-COR, Lincoln, NE, USA). For each sampling area, one sky value and five target values recorded by LAI-2200C were utilized in the LAI calculation of crop canopy. The sky value served as the calibration reference, while the average of the five target values was taken as the ground truth LAI for the corresponding site. Measurements were executed between 16:00 and 18:00 local time, specifically avoiding direct sunlight whenever possible.



**Figure 2.** Examples of ground photos reflecting winter wheat growth status.

For the measurement of LCC, the “five-point sampling” method was applied to select five flag leaves from wheat plants within one-meter square in each sampling area. Subsequently, these selected leaves were promptly placed in an insulated box with ice for transportation to the laboratory. In total, 0.1 g fresh leaf disks were collected from wheat leaf blades of each sampling area using a leaf puncher (diameter = 8 mm), and their pigments were extracted using 10 mL 95% analytical reagent alcohol. Extract absorbance at 649 nm and 665 nm was measured using an ultraviolet-visible spectrophotometer (MAPADA, Shanghai, China) after 24 h of dark storage. The determination of total chlorophyll concentrations in mg/mL involved the utilization of extinction coefficients to the absorbance values. These concentrations were subsequently converted to  $\mu\text{g}/\text{cm}^2$ , taking into account the specific area of the leaf disks and the solution volume, as detailed in the work [13]. The Canopy Chlorophyll Content (CCC), expressed per unit of leaf area, was determined by multiplying the LAI and LCC [37]. A total of 194 valid measurements for each crop bio-parameter were obtained from the four sampling campaigns during the growth season.

### 2.2.2. UAV Platform and Flight Configuration

Multispectral data were simultaneously acquired by a DJI Phantom 4 multispectral UAV (Da-Jiang Innovations, Shenzhen, China). The equipment integrates five optical filter sensors with different central wavelengths (blue:  $450 \pm 16$  nm, green:  $560 \pm 16$  nm, red:  $650 \pm 16$  nm, red edge:  $730 \pm 16$  nm, near infrared:  $840 \pm 26$  nm). The UAV campaigns were conducted between 10:00 and 14:00 local time, under clear sky and low wind speed conditions. The 90% reflectance calibration board took a photo using the UAV camera before takeoff and landing. The flight path was automatically generated by DJI GS Pro and the flight parameter settings were kept consistent each time (Table 2). Network RTK technology was utilized to enhance UAV positioning accuracy. Following image acquisition, band registration and image stitching were performed using DJI Terra, followed by a radiometric correction to obtain multispectral reflectivity orthophoto.

**Table 2.** Flight parameters of Unmanned aerial vehicle (UAV).

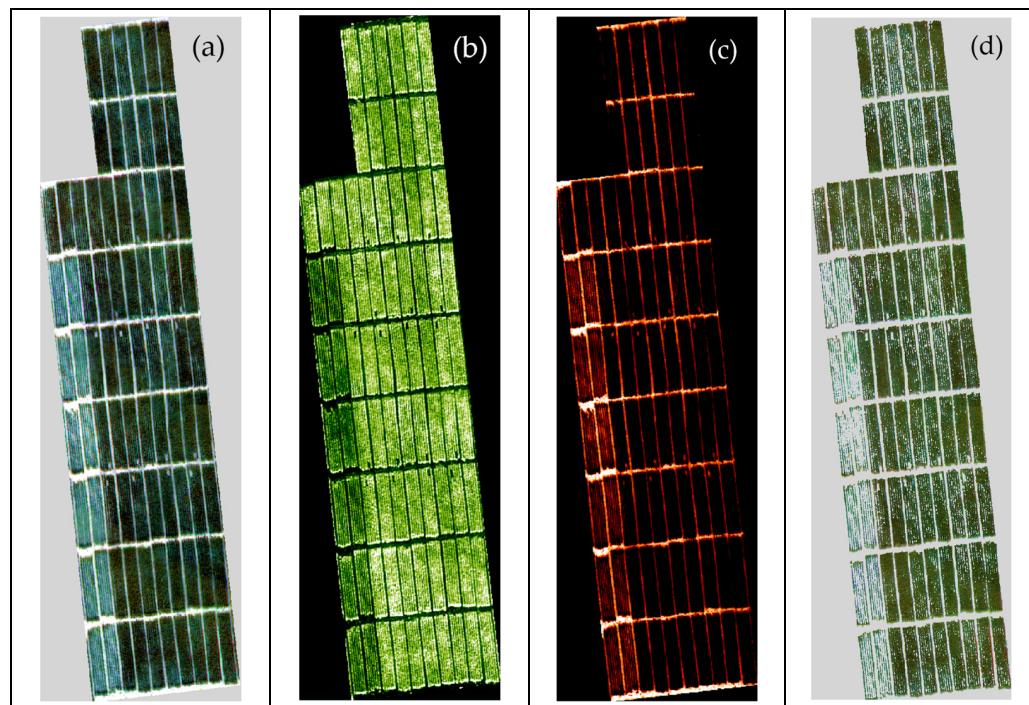
Parameters	Parameter Value
Flight altitude	50 m
Flight Speed	3.8 m/s
Heading overlap ratio	75%
Collateral overlap ratio	80%
Ground Sampling Distance	3 cm

### 2.3. Image Pre-Processing and Data Extraction

#### 2.3.1. Soil Background Removal

To reduce the influence of additional factors such as soil, we performed an image segment using the Sequential Maximum Angle Convex Cone (SMACC) tool [38] within the ENVI 5.3 software (Harris Geospatial Solutions Inc., Boulder, CO, USA). Linear spectral

unmixing is a commonly used method in spectral image classification for mixed pixels. The method consists of two steps: first, extracting the spectra of “pure” ground objects (end-member extraction); and second, representing mixed pixels through linear combinations of end elements (mixed pixel decomposition). The abundance image is the visualization result of mixed pixel decomposition, revealing the relative contributions of each endmember within each pixel. The SMACC tool integrates linear spectral unmixing to simplify the endmember extraction process. It enables the rapid and automated extraction of endmember spectra and abundance images from raw spectral images with a streamlined process. The experimental field predominantly contained wheat and soil. Thus, the number of endmembers considered included wheat, bare soil, and shadow. According to the statistical histogram watershed of wheat abundance image at each growth stage, a threshold was set to remove the soil background. As shown in Figure 3, we could identify the endmembers (wheat and soil) from the images with a spatial resolution of 3 cm.



**Figure 3.** The results of wheat extraction based on the sequential maximum angle convex cone (SMACC) method. (a) RGB Image, (b) wheat abundance, (c) soil abundance, (d) wheat image after removing soil background.

### 2.3.2. Calculation of Vegetation Index

The reflectance values of five bands served as the basis for calculating the vegetation indices, which are commonly utilized in the works to estimate growth bio-parameters and monitor crop growth status. In this study, the initial variables set was established by the combination of five spectral bands and the 13 vegetation indices, which are used to develop a feature selection and machine learning-based model for estimating bio-parameters values of winter wheat. The spectral variables formulas applied are presented in Table 3.

**Table 3.** Spectral variables used in this study.

Variable	Abbreviation	Formulation	Reference
Blue band	B	—	
Green band	G	—	
Red band	R	—	
Red edge band	RE	—	

**Table 3.** Cont.

Variable	Abbreviation	Formulation	Reference
NIR band	NIR	—	
Agriculture Chlorophyll Index	ACI	Green/NIR	[39]
Canopy Chlorophyll Content Index	CCCI	NDRE/NDVI	[40]
Chlorophyll Index using Red Edge Reflectance	CI <sub>red-edge</sub>	(NIR/Edge) – 1	[41]
Chlorophyll Vegetation Index	CVI	NIR × (Red/Blue <sup>2</sup> )	[42]
Green Normalized Difference Vegetation Index	GNDVI	(NIR – Green)/(NIR + Green)	[43]
Leaf Chlorophyll Index	LCI	(NIR – Edge)/(NIR + Red)	[44]
Modified Soil Adjusted Vegetation Index	MSAVI	$\frac{(2 \times \text{NIR}+1)-\sqrt{(2 \times \text{NIR}+1)^2-8 \times (\text{NIR}-\text{Red})}}{2}$	[45]
MERIS Terrestrial Chlorophyll Index	MTCI	(NIR – Edge)/(Edge – Red)	[46]
Modified Triangular Vegetation Index 2	MTVI2	$\frac{1.5 \times}{\sqrt{(2 \times \text{NIR}+1)^2-(6 \times \text{NIR}-5 \times \sqrt{\text{Red}})-0.5}}$	[47]
Normalized Difference Red Edge Index	NDRE	(NIR – Edge)/(NIR + Edge)	[40]
Normalized Difference Vegetation Index	NDVI	(NIR – Red)/(NIR + Red)	[48]
Green NDVI	NDVIG	(Edge – Green)/(Edge + Green)	[43]
Structure Insensitive Pigment Index	SICI	(NIR – Blue)/(NIR – Red)	[49]

Note: In the formulations, B, G, R, RE, and NIR represent the reflectance values corresponding to the blue (450 nm), green (560 nm), red (650 nm), red edge (730 nm), and near-infrared (840 nm) bands, respectively. ‘—’ refers to the reflectance value of the corresponding band.

## 2.4. Modeling Methods

### 2.4.1. Least Absolute Shrinkage and Selection Operator Regression (LASSO)

LASSO is a statistical method used for variable selection and regularization in linear regression models. In LASSO regression, a shrinkage (or regularization) process is incorporated into the traditional linear regression model, which helps prevent overfitting and select the most relevant predictor variables. It achieves this by introducing a penalty term based on the absolute values of the regression coefficients. The key feature of LASSO is its ability to shrink some coefficients to exactly zero, effectively performing automatic variable selection. The final objective of the process is to minimize the prediction error. The parameter for regularization amount control was tuned in this study through 5-fold cross-validation. The range of the parameter was set between 0.01 and 100 values along the regularization pass to identify the parameter value with the minimal mean squared error. The regression modeling was performed with the R package ‘glmnet’.

### 2.4.2. Random Forest Regression (RFR)

RFR is an ensemble learning method in machine learning that leverages the power of multiple decision trees using the “Bagging” idea [50]. RFR regression works by constructing a multitude of decision trees during training and outputs the average prediction of the individual trees for the regression task. RF also provides insights into feature importance, aiding in variable selection and understanding the factors influencing the regression model. In this study, three parameters were tuned with grid search, namely the number of rounds ranged from 50 to 150 with step 20 and the max tree depth ranged from 3 to 20 with step 5, the mode of max branching features number was set ‘None’, ‘log2’ and ‘sqrt’. The regression modeling was performed with Python’s ‘sklearn’ library.

### 2.4.3. Support Vector Machine Based Sequential Forward Selection Regression (SFS-SVR)

SFS-SVR is the combination of the sequential forward selection algorithm and support vector machine. The variable selection is firstly performed by SFS by wrapper algorithm idea, which extends the variables subset from an initial set of variables in each iteration with the variable that increases the induction learner performance the most [28,51]. SFS starts with an empty subset and iteratively adds a variable to the subset to select the input variable combination that has the best merit value based on the evaluation function. In this study, the support vector machine with radial basis kernel was utilized as induction

learner, and its root mean square error was used as the criterion to be minimized, using the resampling technique at each iteration to stabilize the feature rankings. The variables selected by SFS are taken as input variables for the next step regression modeling. The SVR with radial basis function kernel was utilized as regression model, and a grid search was employed to optimize model parameters  $C$  and  $\gamma$ . To avoid overfitting,  $C$  was varied from 0.1 to 10 and combined with  $\gamma$  from 0.005 to 5 in the grid search. The variable selection and regression modeling were performed with R package ‘mlr3’ and ‘caret’, respectively.

#### 2.4.4. Accuracy Assessment

To test how accurately the models predict the value of bio-parameters values, including LAI, LCC, and CCC, the coefficients of determination ( $R^2$ ), root mean square error (RMSE), and relative predictive deviation (RPD) were selected to evaluate the accuracy of model training and model validation. Hold-out validation was utilized to obtain the merits of this study. With regard to dataset partitioning, 70% of the samples were used for training and 30% were for validation.

### 3. Results

#### 3.1. Descriptive Statistics

##### 3.1.1. Distribution of Biochemical Parameters in the Winter Wheat

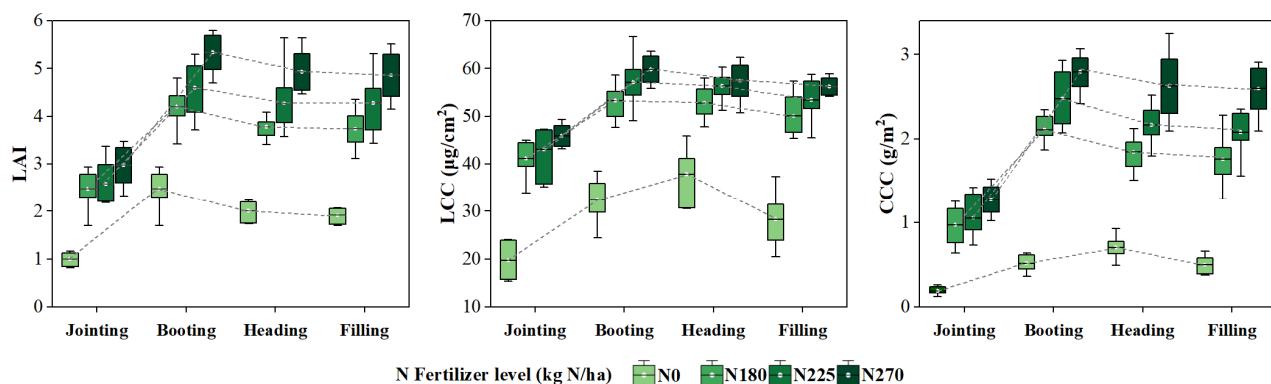
Table 4 displays the variations in ground-measured LAI, LCC, and CCC values at four growth stages of winter wheat. Across all stages, the LAI varies from 0.70 to 5.82, with the standard deviation (SD) of 1.23, the LCC varies from  $15.40 \mu\text{g}/\text{cm}^2$  to  $70.08 \mu\text{g}/\text{cm}^2$  with SD of 12.98, the CCC varies from  $0.12 \text{ g}/\text{m}^2$  to  $3.25 \text{ g}/\text{m}^2$  with SD of 0.82. The mean values of three biochemical parameters showed a trend of increasing first and then decreasing.

**Table 4.** Descriptive statistics of the values of bio-parameters for different growth stages.

Growth Stage	Parameter	Samples	Min	Mean	Max	S-D
Jointing	LAI		0.70	1.88	3.37	0.81
	LCC ( $\mu\text{g}/\text{cm}^2$ )	40	15.40	35.83	51.68	11.42
	CCC ( $\text{g}/\text{m}^2$ )		0.12	0.70	1.42	0.44
Booting	LAI		1.47	3.55	5.30	1.24
	LCC ( $\mu\text{g}/\text{cm}^2$ )	46	24.47	51.70	66.65	11.48
	CCC ( $\text{g}/\text{m}^2$ )		0.36	1.84	3.05	0.84
Heading	LAI		1.57	3.46	5.65	1.04
	LCC ( $\mu\text{g}/\text{cm}^2$ )	52	28.20	52.45	66.72	10.01
	CCC ( $\text{g}/\text{m}^2$ )		0.42	1.74	3.25	0.70
Filling	LAI		1.45	3.57	5.82	1.11
	LCC ( $\mu\text{g}/\text{cm}^2$ )	56	20.44	50.41	70.08	13.32
	CCC ( $\text{g}/\text{m}^2$ )		0.30	1.75	2.91	0.76
All Stages	LAI		0.70	3.26	5.82	1.23
	LCC ( $\mu\text{g}/\text{cm}^2$ )	194	15.40	48.93	70.08	12.98
	CCC ( $\text{g}/\text{m}^2$ )		0.12	1.60	3.25	0.82

Figure 4 shows the LAI, LCC, and CCC values of winter wheat at four growth stages under different nitrogen fertilizer levels. The results showed that LAI values presented a consistent changing trend under different N fertilizer levels, and reached their maximum at the booting stages. Similar to the changing trend of LAI, LCC, and CCC values of winter wheat under N180, N225, and N270 treatments reached their maximum at the booting stage, and started to decrease afterward. The results indicated that the bio-parameter values of wheat are at their peak during the booting stage, and then begin to decrease. This is because nutrients and water are primarily utilized for the growth of roots, stems, and leaves of winter wheat before the jointing state, and then they are allocated to the growth of wheat spikes. Besides, as the leaves in the lower layer within the canopy senesce, this leads to changes in bio-parameters traits of the canopy. Note that the LCC and CCC values of winter wheat under N0 treatment reached their maximum at the heading stage. This

relative delay phenomenon may be caused by the slow growth of wheat leaves due to nitrogen deficiency. The standard deviations of LCC at high N levels were lower than those at low N levels, indicating that the crop canopies with sufficient N fertilizer were more homogeneous, while the standard deviation results of LCC were the opposite for LAI and CCC. Overall, the values of three bio-parameters were roughly correlated with N fertilizer and presented a similar changing trend across the four growth stages.



**Figure 4.** Variation in bio-parameters under various nitrogen (N) fertilizer levels.

### 3.1.2. Correlation Analysis

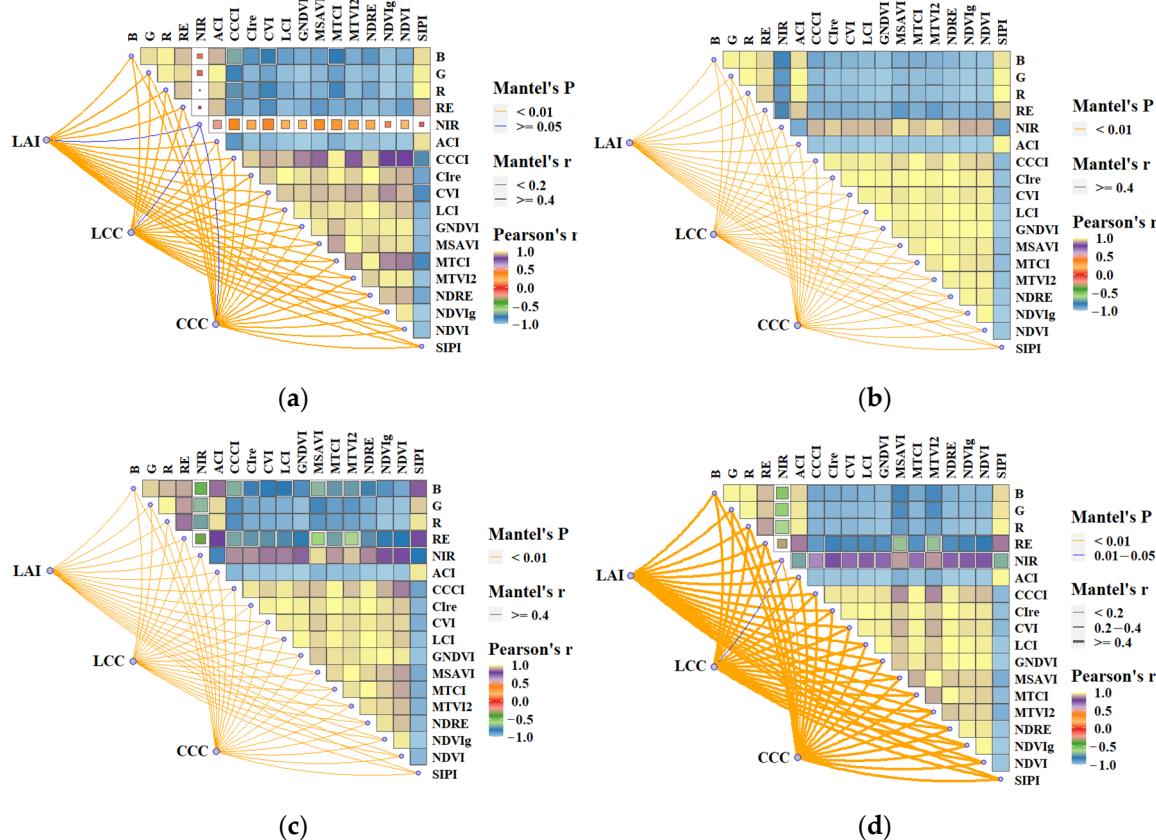
Figure 5 shows a mantel test heat map of correlation analysis between bio parameters and different spectral variables of four growth stages of winter wheat. The correlation between bio-parameters and variables is represented by line color and thickness, while the correlation between variables is represented by color and rectangular area. Mantel's P refers to  $p$ -value; the larger the Mantel test's  $r$  and the smaller the  $p$ -value, the greater the impact of the variable on the bio-parameter. The results indicated significant correlation between bio-parameters and most of the spectral variables across different growth stages. Meanwhile, a high correlation was observed among spectral variables, posing potential multicollinearity challenges for regression modeling. By using spectral variable selection, some redundant variables will be removed to obtain a more simplified model.

## 3.2. Estimation Models of Winter Wheat Bio-Parameters

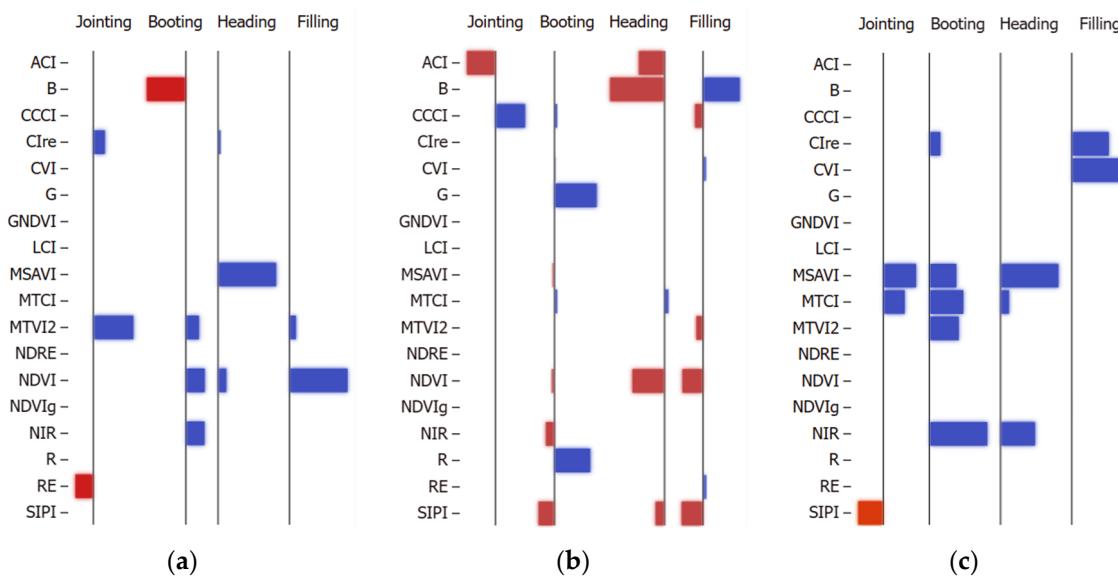
### 3.2.1. Feature Variable Selection

Eighteen variable candidates, including five spectral bands and 13 vegetation indices, were used to select the optimal variables suitable for modeling to estimate each of the three bio-parameters. LASSO, RF important measurement, and SFS were implemented on the 18 spectral variables to select the optimal variable combination at four growth stages.

Figure 6 shows the selected feature variables by the LASSO model for each bio-parameter and growth stage. The optimal combination of variables can be selected by the LASSO internal selector to simplify the model. LASSO reduces the coefficients of unimportant features to zero, and the size of the coefficient reflects the impact of the feature on the target variable. A coefficient with a larger absolute value indicates that the feature has a significant impact on the target variable, while a coefficient with a smaller or zero value indicates that the feature has a smaller or non-existent impact on the dependent variable [29,52]. The results showed that the selected feature variables at each growth stage were quite different; thus, this also demonstrated that it was not appropriate to choose a unified variable for further research. It was noted that MTVI2 and NDVI were the top selected feature variables for LAI across all growth stages, CCCI, NDVI, and SIPI were the preferred feature variables for LCC across all growth stages, while MSAVI, and MTCI were the favored feature variables for CCC across all growth stages.



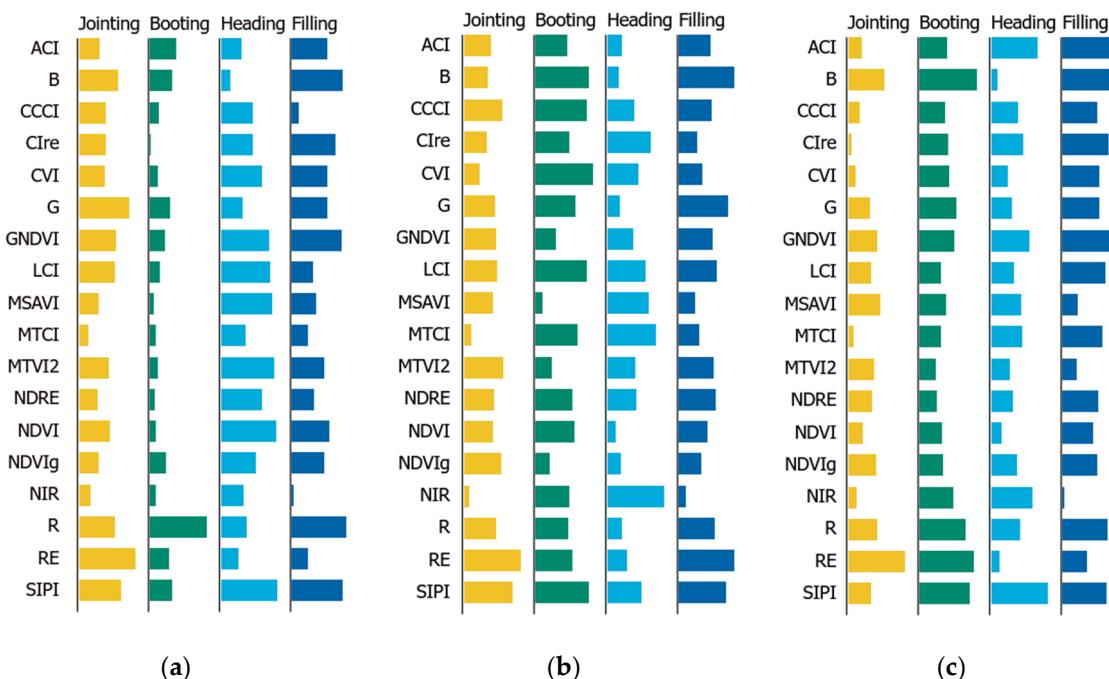
**Figure 5.** The results of correlation analysis between three bio-parameters and spectral variables at four stages of winter wheat. (a) Late jointing, (b) booting, (c) heading, and (d) early filling stage.



**Figure 6.** Selected feature variables by LASSO and their importance scores at four growth stages, displayed as regression coefficients with bio parameters. blue bar refers to positive correlation, red bar refers to negative correlation. (a) LAI, (b) LCC, and (c) CCC.

The feature importance score of all variable candidates derived from the RF model for each bio-parameter and growth stage are shown in Figure 7. RF considers the importance of each feature variable and assigns greater weights to more important features on the model. This means that the variable with a high importance score contributed a larger

share in the model's prediction and had a more significant impact on the final regression results. When the variable importance score is very low, it either means the variable is not important or it is highly collinear with one or more other variables [50]. It should be noted that although RF can estimate the importance of the feature variable, it cannot provide specific information on how the feature variable explains the target variable. GNDVI and SIPI made a relatively high contribution to the RF model for LAI at all growth stages. Rededge, CCCI, NDRE, and SIPI had relatively high and stable contributions to the RF model for LCC at all growth stages. Red, GNDVI, and SIPI had relatively high and stable contributions to the RF model for CCC at all growth stages.



**Figure 7.** The feature importance ranking from RF of winter wheat at four growth stages. (a) LAI, (b) LCC, and (c) CCC.

Different LASSOs and RFs embed internal selectors; SFS is a wrapper variable selection algorithm that separates from the regression modeling. SFS can form a feature subset from all feature candidates for the following regression modeling. The feature subset is the optimal combination of feature variables determined by the induction learner during the forward search process. This reflects which features are considered to have a positive impact on model performance, while which features are ignored or excluded.

The optimal variables selected by the three methods for the bio-parameters modeling are listed in Tables 5–7, respectively. Among them, the variables with the importance score are more than 0.7 derived from RF were presented in the tables. The results showed that the optimal variables were quite different among the three methods for each bio-parameter and growth stage. We marked the variable selected more than two in each growth stage for the three methods in the tables. For LAI, the selected frequency of rededge and MTVI2 were highest at the late jointing stage, the selected frequency of blue, red, rededge, ACI and MTVI2 were highest at the booting stage, the selected frequency of LCI, MSAVI, MTVI2 and NDVI were highest at the heading stage, NDVI was the only variable that was commonly selected by all three models at the early filling stage. For LCC, CCCI was the only variable that was commonly selected by all three models at the late jointing stage, the selected frequencies of CCCI, MTCI and SIPI were highest at the booting stage, the selected frequency of MTCI and NDVI were highest at the heading stage, the selected frequency of blue, rededge, CCCI and SIPI were highest at the early filling stage. For CCC, MSAVI and CIre were the variables commonly selected by all three models at the late jointing and

booting stage, respectively. The selected frequency of NIR and MTCI was highest at the heading stage, while the selected frequency of Clre and CVI was highest at the early filling stage. Overall, SFS selected fewer variables from all variable candidates than LASSO and RF, thus having better performance in reducing data redundancy.

**Table 5.** The results of variable selection for LAI modeling.

Variables	Late Jointing			Booting			Heading			Early Filling		
	LASSO	RF	SFS	LASSO	RF	SFS	LASSO	RF	SFS	LASSO	RF	SFS
B		✓		✓	✓	✓						✓
G		✓			✓							
R					✓		✓					✓
RE	✓	✓				✓	✓					✓
NIR				✓								
ACI					✓		✓			✓		
CCCI												
Clre	✓							✓				✓
CVI			✓									✓
GNDVI								✓		✓		✓
LCI									✓			
MSAVI									✓			
MTCI												
MTVI2	✓		✓	✓			✓	✓	✓	✓		✓
NDRE												
NDVI				✓				✓	✓		✓	
NDVIG			✓									✓
SIP1		✓	✓		✓				✓			✓

**Table 6.** The results of variable selection for LCC modeling.

Variables	Jointing			Booting			Heading			Filling		
	LASSO	RF	SFS	LASSO	RF	SFS	LASSO	RF	SFS	LASSO	RF	SFS
B					✓					✓	✓	
G				✓								✓
R				✓								
RE		✓								✓	✓	
NIR				✓						✓	✓	
ACI	✓											
CCCI	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
Clre				✓								
CVI			✓	✓								✓
GNDVI												
LCI					✓					✓	✓	
MSAVI					✓							✓
MTCI				✓			✓	✓				
MTVI2		✓										✓
NDRE					✓							
NDVI					✓					✓		
NDVIG						✓						
SIP1		✓			✓	✓				✓	✓	

**Table 7.** The results of variable selection for CCC modeling.

Variables	Jointing				Booting				Heading				Filling			
	LASSO	RF	SFS	LASSO	RF	SFS	LASSO	RF	SFS	LASSO	RF	SFS	LASSO	RF	SFS	
B		✓				✓										✓
G																
R		✓				✓										
RE		✓				✓										
NIR					✓					✓	✓	✓				
ACI										✓	✓					
CCCI																
Cle			✓		✓			✓								
CVI			✓													
GNDVI		✓														
LCI			✓													
MSAVI	✓		✓													
MTCI	✓															
MTVI2																
NDRE																
NDVI																
NDVIdg																
SPI	✓															

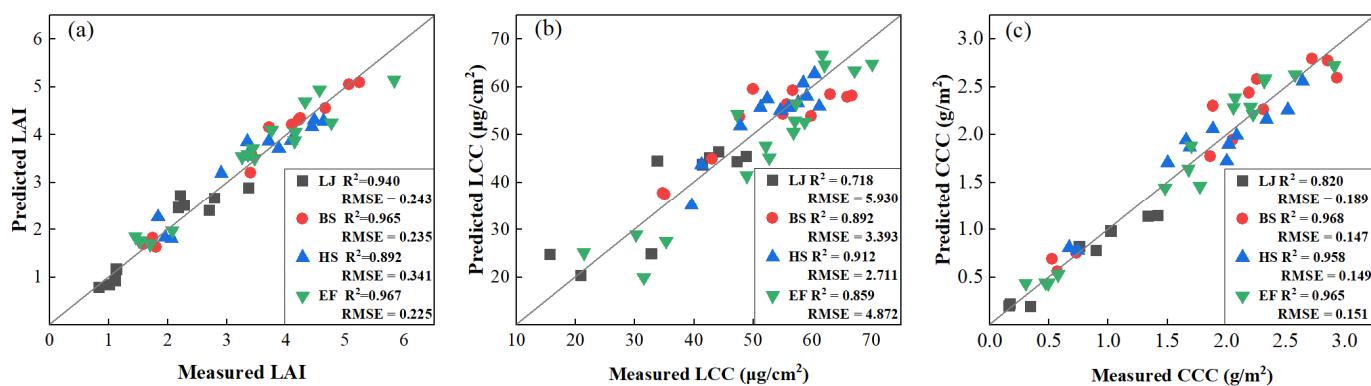
### 3.2.2. Model Accuracy Comparison

The regression results of bio-parameters at each growth stage conducted by LASSO, RFR, and SFS-SVR are presented in Table 8. The optimal variable combination selected by SFS was applied for SVR-based bio parameter prediction. The LASSO and RFR applied their internal selector to screen optimal variables for bio-parameter prediction.

**Table 8.** The results of the regression models for winter wheat bio-parameters prediction at the four growth stages.

Late Jointing	Model	LAI				LCC				CCC			
		Training set		Test set		Training set		Test set		Training set		Test set	
		RMSE	RPD	RMSE	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD
LASSO	0.204	4.417	0.318	2.502	5.224	2.022	5.993	1.771	0.132	3.432	0.114	3.503	
RFR	0.108	7.111	0.267	2.439	1.992	5.588	6.405	1.461	0.056	7.247	0.113	3.101	
SFS-SVR	0.201	4.129	0.243	3.184	5.032	1.876	5.93	1.786	0.076	5.738	0.189	2.249	
Booting	Model	LAI				LCC				CCC			
		Training set		Test set		Training set		Test set		Training set		Test set	
		RMSE	RPD	RMSE	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD
LASSO	0.25	4.650	0.301	4.392	3.34	3.396	5.201	1.501	0.141	5.701	0.218	3.837	
RFR	0.094	14.699	0.403	3.543	3.127	3.933	5.097	2.071	0.098	9.362	0.227	4.069	
SFS-SVR	0.255	4.702	0.235	4.856	3.477	2.831	3.393	2.530	0.140	5.531	0.147	5.279	
Heading	Model	LAI				LCC				CCC			
		Training set		Test set		Training set		Test set		Training set		Test set	
		RMSE	RPD	RMSE	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD
LASSO	0.261	3.904	0.314	2.795	4.349	1.562	3.893	1.818	0.162	3.349	0.151	3.383	
RFR	0.124	9.954	0.377	2.671	2.408	3.962	4.219	2.186	0.073	11.404	0.300	2.290	
SFS-SVR	0.250	3.760	0.341	2.449	4.277	1.804	2.711	2.872	0.125	5.190	0.149	4.460	
Early Filling	Model	LAI				LCC				CCC			
		Training set		Test set		Training set		Test set		Training set		Test set	
		RMSE	RPD	RMSE	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\mu\text{g}/\text{cm}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD	RMSE ( $\text{g}/\text{m}^2$ )	RPD
LASSO	0.255	3.840	0.404	2.852	4.336	2.773	4.671	2.565	0.166	4.226	0.184	3.603	
RFR	0.126	10.277	0.337	3.987	2.236	6.257	6.784	1.736	0.086	10.241	0.209	4.242	
SFS-SVR	0.229	4.956	0.225	4.905	5.022	2.524	4.872	2.222	0.160	4.906	0.151	4.884	

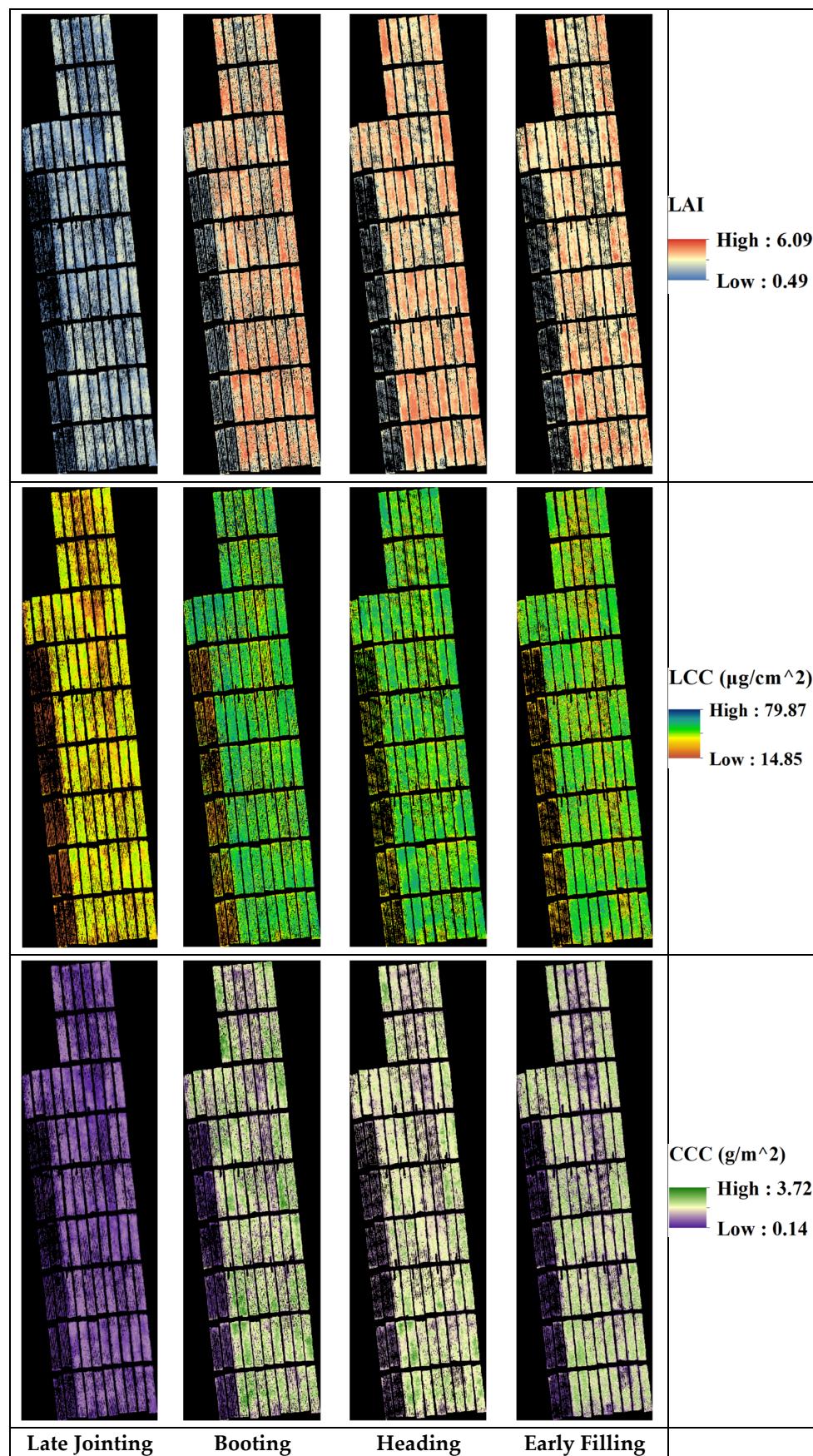
The results showed that, at most, one bio-parameter model built by LASSO across the four growth stages had the highest testing accuracy, namely the heading stage for LAI, the early filling stage for LCC, and the late joining stage for CCC. RFR had the highest training accuracy and the lowest testing accuracy. According to the comparison results, the SFS-SVR exhibited optimal accuracy in evaluating the LAI, LCC, and CCC. The LAI evaluation model based on the SFS-SVR resulted in the highest test performance having  $R^2$ , RMSE and RPD of 0.967, 0.225 and 4.905 at the early filling stage; the LCC evaluation model resulted in the highest test performance having  $R^2$ , RMSE and RPD of 0.912, 2.711  $\mu\text{g}/\text{cm}^2$  and 2.872 at the heading stage; and the CCC evaluation model resulted in the highest test performance having  $R^2$ , RMSE and RPD of 0.968, 0.147  $\text{g}/\text{m}^2$  and 5.279 at the booting stage, respectively. For estimating the three bio parameters levels of winter wheat, SFS-SVR shows better accuracy and robustness in predicting winter wheat bio-parameters during the four growth stages. Compared to LAI and CCC, the accuracy of LCC modeling using SFS-SVR was relatively poor. Scatter plots for the optimal regression model in evaluating each of the three bio-parameters are shown in Figure 8.



**Figure 8.** The validation results of the SFS-SVR model in evaluating the LAI (a), LCC (b) and CCC (c) status of winter wheat at four growth stages. (LJ: late jointing, BS: booting, HS: heading, EF: early filling stage).

### 3.2.3. Winter Wheat Bio-Parameters Mapping

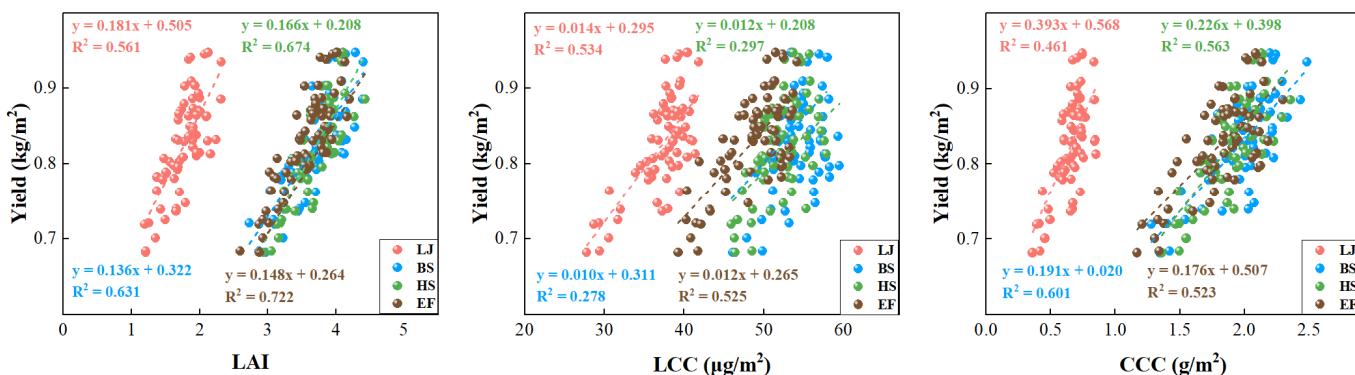
The models with the highest predictive capability among those developed in this study were used to construct pixel-level spatial mapping of the three bio-parameter values at each growth stage. Figure 9 displays the predicted maps of winter wheat bio-parameter values across four growth stages using the SFS-SVR model. The visualization results were highly consistent with the field experiment, as displayed in Figures 1 and 4, which indicates that the inversion results were reliable. The within-plot variance for each fertilization treatment was low, and the difference between plot treatments was obvious, particularly for the results of CCC. Consequently, the results were available for further field-level precision fertilization study.



**Figure 9.** The maps of winter wheat bio parameters retrieved by SFS-SVR at four growth stages.

### 3.3. The Relationship between Winter Wheat Grain Yield and Biochemical Parameters

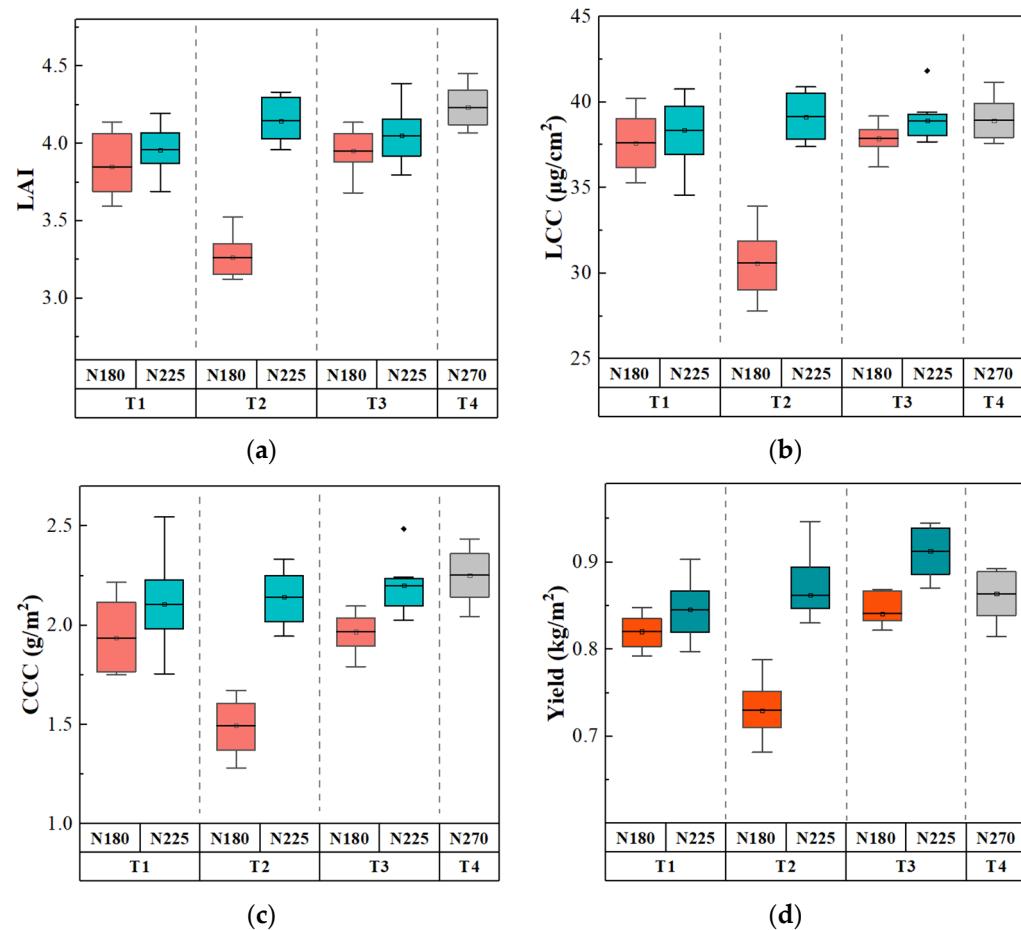
Remote sensing estimation of winter wheat bio-parameters is based on Vis, which can serve as indicators of grain yield [53]. It is necessary to verify whether the relationship between wheat LAI, LCC, CCC, and yield (measured) is significant. The average values of bio-parameters of each plot were extracted using the ArcGIS zoom statistics tool. Figure 10 shows the relationship between LAI, LCC, CCC, and yield under different stages. The results displayed that the LAI, LCC, and CCC were related to yield, and the relationship varied with the growth stage. For LAI, the goodness of fit ( $R^2$ ) values were 0.561 (late jointing), 0.631 (booting), 0.674 (heading) and 0.722 (early filling). For LCC, the  $R^2$  values were 0.534 (late jointing), 0.278 (booting), 0.297 (heading) and 0.525 (early filling). For CCC, the  $R^2$  values were 0.461 (late jointing), 0.601 (booting), 0.563 (heading) and 0.523 (early filling), respectively. The growth stages with the highest correlation between yield and LAI, LCC, and CCC, were early filling ( $R^2 = 0.722$ ), late jointing ( $R^2 = 0.534$ ), and booting stage ( $R^2 = 0.601$ ), respectively. The relevance ranking: LAI > CCC > LCC. The correlation analysis results showed that the LAI and CCC were crucial indicators for assessing the yield of winter wheat, but the assessing results can be affected by variations at the growth stage. This also provided a basis for the prediction of winter wheat yield using the crop bio-parameters.



**Figure 10.** The relationship between the winter wheat yield and LAI, LCC and CCC at four growth stages. (LJ: late jointing, BS: booting, HS: heading, EF: early filling stage).

To explore the variations in wheat biochemical parameters and yield under different nitrogen treatments, the bio-parameters values at the growth stage with the highest prediction accuracy were compared with yield under different nitrogen treatments (Figure 11). It can be seen that both bio-parameters and yield increased with N fertilizer level, except for the treatment of N270 under T4, the difference of bio-parameter values and yield under different treatments presented good consistency. For the four N fertilizer treatments, the average yield ranking was T3 > T4 > T1 > T2, and the average values of the three bio-parameters were T4 > T3 > T1 > T2. This indicated that wheat growth status and yield were not only related to the N fertilizer level, but also to the fertilization approach of base and topdressing fertilizers, and excessive fertilization cannot increase wheat yield. In this study, the optimal N treatment for increasing yield was N225 under T3, rather than N270 under T4, which had the highest fertilization rate. The reason could be that for the N270 under T4 treatment, its effective nutrients were not sufficient to supply the growth of wheat grain, but rather to the growth of other organs such as leaves and stems. Overall, the trend of changes in the three bio-parameters and yield presented consistency among different N treatments. With the increase in N fertilizer level except for the treatment of N270 under T4, both bio-parameters and yield increased. The results indicated that reasonable treatment with base fertilizer and topdressing fertilizer can promote the improvement of wheat growth and yield. This also indirectly demonstrated the accuracy of bio-parameter prediction results. In addition, these findings will provide a scientific basis for enhanced monitoring

and diagnosis of nitrogen nutrition, contributing to improved field management practices for winter wheat.



**Figure 11.** The variation in winter wheat LAI (a), LCC (b), CCC (c) and yield (d) under different N treatments. N represents the nitrogen fertilizer level (180, 225, 270 kg N/ha). Treatment T1 received fertilizer split a half mixing of urea and slow-release fertilizers in a proportion of 1:1 at sowing and a half mixing of urea and slow-release fertilizers in a proportion of 1:1 at the jointing state; T2 received fertilizer split a half urea fertilizer at sowing and a half slow-release fertilizer at the jointing state; T3 received fertilizer split a half slow-release fertilizer at sowing, and a half urea fertilizer at the jointing state; T4 received fertilizer split a half urea fertilizer at sowing and a half slow-release fertilizer at the jointing state.

#### 4. Discussion

##### 4.1. Uncertainty of Observed Data

Firstly, a commonly overlooked issue in UAV multispectral remote sensing applications is the presence of multi-source errors in multispectral data, which can affect the accuracy of wheat bio-parameters estimation. Due to the multiple independent sensor lenses with different spectral bands, band registration and image mosaic are necessary for data preprocessing. However, the surface texture of the field crop canopy is uniform, so it is easy for it to result in fewer matching feature points due to less distinctive texture features [54]. The existing approach to processing image data was performed by popularly used software such as DJI Terra. In addition, the growth of winter wheat leads to changes in canopy structure, the growth and senescence of wheat leaves and spikes, as well as the differences in changes of light radiation of different flight operation times, which brings a certain degree of uncertainty in data consistency. A general method alleviating this issue was achieved by two approaches: (i) collecting data during periods of relatively stable and

sufficient solar radiation conditions and without cloud coverage; and (ii) recording light radiation information using the built-in photometer of multispectral sensor and participating in subsequent calibration with a standard reflectance panel. Besides, this study did not consider the heterogeneity of bio-parameters vertical distribution due to the influence of light conditions, which have been reported in studies [55–57]. It was assumed that the foliar chlorophyll content in the vertical layer of wheat was constant in this study. Thus, using only the LCC of flag leaves may lack representativeness for the chlorophyll content of the sample area, which has to some extent affected the estimation of LCC. Thus, we used CCC to estimate canopy chlorophyll content. Finally, due to the similarity of the spectral index calculation formula, when using the five broadband multispectral bands to calculate VIs, a lot of highly correlated VIs will be generated during regression modeling. We attempt to use different algorithms to reduce collinearity effects and screen an optimal variable combination to reduce the redundancy of VIs. The results of this study showed that the variable candidates can be reduced from 18 spectral variables to only a few feature variables, further improving the accuracy and efficiency of modeling prediction.

#### 4.2. Comparison of Different Models

This study evaluated the effectiveness of three different regression methods combined with variable selection, including LASSO, RFR, and SFS-SVR, on estimating winter wheat growth bio-parameters traits. Previous studies have indicated that due to the saturation of a single VI and its low sensitivity during the growth stages, it is difficult to accurately estimate the multi-temporal changes in crop bio-parameters using traditional VI methods [16–20]. Non-parametric machine learning methods, such as RFR and SVR, are less sensitive to skewness in data distribution, and can therefore be used to handle non-normal data [21–25]. In this study, the results demonstrate that it is feasible to accurately predict the bio-parameters of winter wheat at variable stages based on the VIs and machine learning regression combined with variable selection. Moreover, the combination of machine learning with variable selection is suitable for solving data redundancy and multicollinearity problems. In addition, the relative importance of each input variable may vary depending on the crop growth stage and severity of crop stress. The utilization of different stages of data as input variables resulted in variations in model accuracy. Comparing SFS with LASSO and RF internal feature selector, the input variables selected by the three methods were different. Specifically, SFS selected the fewest variables for bio-parameters modeling at four growth stages. The training accuracy of the RFR model was the highest among all three models, but its training accuracy was much higher than the testing accuracy, and the testing accuracy was the lowest among the three models. It indicated that the constructed RFR model was overfitted, which may be due to limited modeling samples and the use of inappropriate optimization model parameters. For the regression results of the LASSO model at each growth stage, at most one bio-parameter prediction had the highest testing accuracy. By comparing the three models, we found that the performance of SFS-SVR was more robust, and it showed a better ability to predict wheat bio-parameters across different growth stages. In addition, compared to the other two methods, the SFS usually selected the least input variables, which can reduce data redundancy while greatly saving model prediction time and improving model efficiency.

#### 4.3. Effects of Crop Phenology on Bio-Parameters Estimation

Effects of growth stage, crop type and the range of variation in bio-parameters should be taken into account when applying remote sensing in precision agriculture [16,58,59]. The findings in this study confirmed varying accuracies of the models across different growth stages, and that the phenological factor can impact model accuracy within the experimental setup. Specifically, the results indicated that the LAI estimation at the early filling stage of winter wheat had the highest prediction accuracy, the LCC estimation at the heading stage had the highest prediction accuracy, and the CCC estimation at the booting stage had the highest prediction accuracy. For the prediction accuracy of winter wheat bio-parameters

values at the four growth stages, LAI: early filling > booting > late jointing > heading, LCC: heading > early filling > booting > late jointing, CCC: booting > early filling > heading > late jointing. The accuracy of bio-parameter estimation varied with different growth stages of winter wheat, which can be attributed to changes in various factors including crop canopy structure, leaf thickness and cell structure, leaf pigment content, and crop coverage [15,31,36,60]. In addition, the LAI, LCC, and CCC change with the increase in leaf size and number in the vertical distribution of wheat, and the growth and senescence of wheat spikes also affect the estimation of wheat canopy reflectance and bio-parameters. At the jointing stage, due to the small size of winter wheat, multiple scattering of leaves and soil background mixing significantly affect the canopy reflectance. At the peak booting stage of bio-parameters, the VIs may become saturated, reducing the prediction accuracy of three bio-parameters models. At the early filling stage, the senescent of wheat leaves and spikes may affect the canopy reflectance, which reduces the prediction accuracy at LCC at the early filling stage compared to that at the heading stage.

The study also demonstrated the importance of bio-parameters in evaluating comprehensive yield traits of winter wheat, which is consistent with previous research results [59,61]. Further correlation analysis between wheat yield and LAI, LCC, and CCC confirmed that the bio-parameters are important indicators for yield estimation, but their relationship varied depending on the growth stage (Figure 10). In addition, the correlation between yield and LCC was low ( $R^2 = 0.534$  at LJ), indicating that it is difficult to accurately evaluate wheat yield using LCC alone. As a product of the combination of LAI and LCC, the correlation between CCC and yield had improved ( $R^2 = 0.601$  at BS), but did not exceed the correlation between LAI and yield ( $R^2 = 0.722$  at EF). The high correlation between yield and LAI indicated that LAI can better characterize the winter wheat growth status, while achieving better yield evaluation. Therefore, further research should use LAI as an important factor for assimilating the wheat yield prediction model.

## 5. Conclusions

In this study, UAV-based spectral variables were adopted to estimate the growth-related bio-parameters of winter wheat at the four growth stages. We proposed three statistical methods with regard to feature selection and machine learning, including LASSO, RFR, and SFS-SVR, for winter wheat LAI, LCC and CCC estimation. The finding of this study revealed that: (1) the values of three bio-parameters were generally correlated with N fertilizer and presented similar changing trends across the four growth stages; (2) LAI estimates at the early filling stage, LCC estimates at the heading stage and CCC estimates at the booting of winter wheat were more suitable than estimates at other stages; (3) SFS-SVR was a robust method for the bio-parameters estimation of winter wheat at key growth stages based on UAV multispectral imagery, effectively reducing data redundancy and enhancing predicted accuracy; and (4) LAI was a more crucial indicator related to the yield of winter wheat than CCC and LCC, and the correlation can be impacted by the variations in the growth stage.

In summary, the results demonstrated the potential of using the SFS-SVR regression to estimate winter wheat bio-parameters traits in field scale. This study is valuable for monitoring and estimating winter wheat bio-parameter-based on UAV multispectral imaging of specific crop phenology periods, thereby offering guidance for field management and optimizing agricultural practices to enhance crop yield. Future studies should encompass multiple crops across diverse agricultural contexts, which would enhance the generalizability of findings. A more in-depth temporal analysis over multiple growing seasons can assess model robustness under varying environmental conditions. Exploring sensor fusion with different types of sensors, on-farm validation studies, and optimizing UAV flight parameters contribute to the scalability and practical applicability of the methodology. Additionally, incorporating advanced machine learning techniques and assessing crop yield is crucial for comprehensive advancements.

**Author Contributions:** Conceptualization, C.Z.; methodology, C.Z.; validation, C.Z. and S.Z.; formal analysis, C.Z.; investigation, C.Z., L.W., X.Z., S.C. and Y.Y.; resources, Y.Y. and Z.S.; data curation, C.Z., L.W., X.Z. and S.C.; writing—original draft preparation, C.Z.; writing—review and editing, C.Z. and Y.X.; visualization, C.Z.; supervision, C.Z. and Y.X.; funding acquisition, Y.X. and C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China (grant number 42275147) and Jiangsu Funding Program for Excellent Postdoctoral Talent (grant number 2022ZB507).

**Data Availability Statement:** The data are available from the authors upon reasonable request as the data need further use.

**Acknowledgments:** Special thanks to Weikai Lin and Junwei Ma, Jiangsu Normal University, for their kind help in field surveys. We thank the Editors and Reviewers for their service and insightful comments. All authors have read and agreed to the published version of the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Lesk, C.; Rowhani, P.; Ramankutty, N. Influence of Extreme Weather Disasters on Global Crop Production. *Nature* **2016**, *529*, 84–87. [[CrossRef](#)]
2. Vishwakarma, S.; Zhang, X.; Lyubchich, V. Wheat Trade Tends to Happen between Countries with Contrasting Extreme Weather Stress and Synchronous Yield Variation. *Commun. Earth Environ.* **2022**, *3*, 261. [[CrossRef](#)]
3. Huang, J.; Sedano, F.; Huang, Y.; Ma, H.; Li, X.; Liang, S.; Tian, L.; Zhang, X.; Fan, J.; Wu, W. Assimilating a Synthetic Kalman Filter Leaf Area Index Series into the WOFOST Model to Improve Regional Winter Wheat Yield Estimation. *Agric. For. Meteorol.* **2016**, *216*, 188–202. [[CrossRef](#)]
4. Huang, J.; Ma, H.; Su, W.; Zhang, X.; Huang, Y.; Fan, J.; Wu, W. Jointly Assimilating MODIS LAI and ET Products Into the SWAP Model for Winter Wheat Yield Estimation. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2015**, *8*, 4060–4071. [[CrossRef](#)]
5. Sus, O.; Heuer, M.W.; Meyers, T.P.; Williams, M. A Data Assimilation Framework for Constraining Upscaled Cropland Carbon Flux Seasonality and Biometry with MODIS. *Biogeosciences* **2013**, *10*, 2451–2466. [[CrossRef](#)]
6. Schlemmer, M.; Gitelson, A.; Schepers, J.; Ferguson, R.; Peng, Y.; Shanahan, J.; Rundquist, D. Remote Estimation of Nitrogen and Chlorophyll Contents in Maize at Leaf and Canopy Levels. *Int. J. Appl. Earth Obs. Geoinf. Int.* **2013**, *25*, 47–54. [[CrossRef](#)]
7. Gitelson, A.A.; Viña, A.; Verma, S.B.; Rundquist, D.C.; Arkebauer, T.J.; Keydan, G.; Leavitt, B.; Ciganda, V.; Burba, G.G.; Suyker, A.E. Relationship between Gross Primary Production and Chlorophyll Content in Crops: Implications for the Synoptic Monitoring of Vegetation Productivity. *J. Geophys. Res.* **2006**, *111*, 1–13. [[CrossRef](#)]
8. Dorigo, W.A.; Zurita-Milla, R.; De Wit, A.J.W.; Brazile, J.; Singh, R.; Schaepman, M.E. A Review on Reflective Remote Sensing and Data Assimilation Techniques for Enhanced Agroecosystem Modeling. *Int. J. Appl. Earth Obs. Geoinf. Int.* **2007**, *9*, 165–193. [[CrossRef](#)]
9. Wu, S.; Yang, P.; Ren, J.; Chen, Z.; Li, H. Regional Winter Wheat Yield Estimation Based on the WOFOST Model and a Novel VW-4DEnSRF Assimilation Algorithm. *Remote Sens. Environ.* **2021**, *255*, 112276. [[CrossRef](#)]
10. Sanaeifar, A.; Yang, C.; De La Guardia, M.; Zhang, W.; Li, X.; He, Y. Proximal Hyperspectral Sensing of Abiotic Stresses in Plants. *Sci. Total Environ.* **2023**, *861*, 160652. [[CrossRef](#)]
11. Wang, T.; Gao, M.; Cao, C.; You, J.; Zhang, X.; Shen, L. Winter Wheat Chlorophyll Content Retrieval Based on Machine Learning Using In Situ Hyperspectral Data. *Comput. Electron. Agric.* **2022**, *193*, 106728. [[CrossRef](#)]
12. Zhang, Y.; Hui, J.; Qin, Q.; Sun, Y.; Zhang, T.; Sun, H.; Li, M. Transfer-Learning-Based Approach for Leaf Chlorophyll Content Estimation of Winter Wheat from Hyperspectral Data. *Remote Sens. Environ.* **2021**, *267*, 112724. [[CrossRef](#)]
13. Sun, Q. Monitoring Maize Canopy Chlorophyll Density under Lodging Stress Based on UAV Hyperspectral Imagery. *Comput. Electron. Agric.* **2022**, *193*, 106671. [[CrossRef](#)]
14. Longmire, A.R.; Poblete, T.; Hunt, J.R.; Chen, D.; Zarco-Tejada, P.J. Assessment of Crop Traits Retrieved from Airborne Hyperspectral and Thermal Remote Sensing Imagery to Predict Wheat Grain Protein Content. *ISPRS-J. Photogramm. Remote Sens.* **2022**, *193*, 284–298. [[CrossRef](#)]
15. Zhu, W.; Sun, Z.; Yang, T.; Li, J.; Peng, J.; Zhu, K.; Li, S.; Gong, H.; Lyu, Y.; Li, B.; et al. Estimating Leaf Chlorophyll Content of Crops via Optimal Unmanned Aerial Vehicle Hyperspectral Data at Multi-Scales. *Comput. Electron. Agric.* **2020**, *178*, 105786. [[CrossRef](#)]
16. Yu, K.; Lenz-Wiedemann, V.; Chen, X.; Bareth, G. Estimating Leaf Chlorophyll of Barley at Different Growth Stages Using Spectral Indices to Reduce Soil Background and Canopy Structure Effects. *ISPRS-J. Photogramm. Remote Sens.* **2014**, *97*, 58–77. [[CrossRef](#)]
17. Zhang, S.; Zhao, G.; Lang, K.; Su, B.; Chen, X.; Xi, X.; Zhang, H. Integrated Satellite, Unmanned Aerial Vehicle (UAV) and Ground Inversion of the SPAD of Winter Wheat in the Reviving Stage. *Sensors* **2019**, *19*, 1485. [[CrossRef](#)]
18. Cui, B.; Zhao, Q.; Huang, W.; Song, X.; Ye, H.; Zhou, X. Leaf Chlorophyll Content Retrieval of Wheat by Simulated RapidEye, Sentinel-2 and EnMAP Data. *J. Integr. Agric.* **2019**, *18*, 1230–1245. [[CrossRef](#)]

19. Verrelst, J.; Rivera, J.P.; Gitelson, A.; Delegido, J.; Moreno, J.; Camps-Valls, G. Spectral Band Selection for Vegetation Properties Retrieval Using Gaussian Processes Regression. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 554–567. [[CrossRef](#)]
20. Main, R.; Cho, M.A.; Mathieu, R.; O’Kennedy, M.M.; Ramoelo, A.; Koch, S. An Investigation into Robust Spectral Indices for Leaf Chlorophyll Estimation. *ISPRS-J. Photogramm. Remote Sens.* **2011**, *66*, 751–761. [[CrossRef](#)]
21. Yue, J.; Yang, G.; Tian, Q.; Feng, H.; Xu, K.; Zhou, C. Estimate of Winter-Wheat above-Ground Biomass Based on UAV Ultrahigh-Ground-Resolution Image Textures and Vegetation Indices. *ISPRS-J. Photogramm. Remote Sens.* **2019**, *150*, 226–244. [[CrossRef](#)]
22. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High Resolution Wheat Yield Mapping Using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410. [[CrossRef](#)]
23. Qi, H. Monitoring of Peanut Leaves Chlorophyll Content Based on Drone-Based Multispectral Image Feature Extraction. *Comput. Electron. Agric.* **2021**, *187*, 106292. [[CrossRef](#)]
24. Zhang, L. Evaluating the Sensitivity of Water Stressed Maize Chlorophyll and Structure Based on UAV Derived Vegetation Indices. *Comput. Electron. Agric.* **2021**, *187*, 106292. [[CrossRef](#)]
25. Wu, Q.; Zhang, Y.; Zhao, Z.; Xie, M.; Hou, D. Estimation of Relative Chlorophyll Content in Spring Wheat Based on Multi-Temporal UAV Remote Sensing. *Agronomy* **2023**, *13*, 211. [[CrossRef](#)]
26. Zou, X.; Zhao, J.; Povey, M.J.W.; Holmes, M.; Hanpin, M. Variables Selection Methods in Near-Infrared Spectroscopy. *Anal. Chim. Acta* **2010**, *667*, 14–32.
27. Chandrashekhar, G.; Sahin, F. A Survey on Feature Selection Methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [[CrossRef](#)]
28. Uncu, Ö.; Türkşen, I.B. A Novel Feature Selection Approach: Combining Feature Wrappers and Filters. *Inf. Sci.* **2007**, *177*, 449–466. [[CrossRef](#)]
29. Shafiee, S.; Lied, L.M.; Burud, I.; Dieseth, J.A.; Alsheikh, M.; Lillemo, M. Sequential Forward Selection and Support Vector Regression in Comparison to LASSO Regression for Spring Wheat Yield Prediction Based on UAV Imagery. *Comput. Electron. Agric.* **2021**, *183*, 106036. [[CrossRef](#)]
30. Huang, X.; Guan, H.; Bo, L.; Xu, Z.; Mao, X. Hyperspectral Proximal Sensing of Leaf Chlorophyll Content of Spring Maize Based on a Hybrid of Physically Based Modelling and Ensemble Stacking. *Comput. Electron. Agric.* **2023**, *208*, 107745. [[CrossRef](#)]
31. Han, S.; Zhao, Y.; Cheng, J.; Zhao, F.; Yang, H.; Feng, H.; Li, Z.; Ma, X.; Zhao, C.; Yang, G. Monitoring Key Wheat Growth Variables by Integrating Phenology and UAV Multispectral Imagery Data into Random Forest Model. *Remote Sens.* **2022**, *14*, 3723. [[CrossRef](#)]
32. Wang, W.; Gao, X.; Cheng, Y.; Ren, Y.; Zhang, Z.; Wang, R.; Cao, J.; Geng, H. QTL Mapping of Leaf Area Index and Chlorophyll Content Based on UAV Remote Sensing in Wheat. *Agriculture* **2022**, *12*, 595. [[CrossRef](#)]
33. Han, Y.; Tang, R.; Liao, Z.; Zhai, B.; Fan, J. A Novel Hybrid GOA-XGB Model for Estimating Wheat Aboveground Biomass Using UAV-Based Multispectral Vegetation Indices. *Remote Sens.* **2022**, *14*, 3506. [[CrossRef](#)]
34. Wang, J.; Zhou, Q.; Shang, J.; Liu, C.; Zhuang, T.; Ding, J.; Xian, Y.; Zhao, L.; Wang, W.; Zhou, G.; et al. UAV- and Machine Learning-Based Retrieval of Wheat SPAD Values at the Overwintering Stage for Variety Screening. *Remote Sens.* **2021**, *13*, 5166. [[CrossRef](#)]
35. Wang, W. Prediction of Chlorophyll Content in Multi-Temporal Winter Wheat Based on Multispectral and Machine Learning. *Front. Plant Sci.* **2022**, *13*, 896408. [[CrossRef](#)] [[PubMed](#)]
36. Yin, Q.; Zhang, Y.; Li, W.; Wang, J.; Wang, W.; Ahmad, I.; Zhou, G.; Huo, Z. Estimation of Winter Wheat SPAD Values Based on UAV Multispectral Remote Sensing. *Remote Sens.* **2023**, *15*, 3595. [[CrossRef](#)]
37. Zhang, C.; Xue, Y. Estimation of Biochemical Pigment Content in Poplar Leaves Using Proximal Multispectral Imaging and Regression Modeling Combined with Feature Selection. *Sensors* **2024**, *24*, 217. [[CrossRef](#)]
38. Gruninger, J.H.; Ratkowski, A.J.; Hoke, M.L. The Sequential Maximum Angle Convex Cone (SMACC) Endmember Model. *Proc. SPIE* **2004**, *5425*, 1–14.
39. Carter, G.A.; Cibula, W.G.; Miller, R.L. Narrow-Band Reflectance Imagery Compared with ThermalImagery for Early Detection of Plant Stress. *J. Plant Physiol.* **1996**, *148*, 515–522. [[CrossRef](#)]
40. Barnes, E.M.; Clarke, T.R.; Richards, S.E. Coincident Detection of Crop Water Stress, Nitrogen Status, and Canopy Density Using Ground Based Multispectral Data. In Proceedings of the 5th International Conference on Precision Agriculture and Other Resource Management, Bloomington, MN, USA, 16–19 July 2000.
41. Gitelson, A.A.; Viña, A.; Arkebauer, T.J.; Rundquist, D.C.; Keydan, G.; Leavitt, B. Remote Estimation of Leaf Area Index and Green Leaf Biomass in Maize Canopies. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
42. Datt, B.; McVicar, T.R.; Van Niel, T.G.; Jupp, D.L.B.; Pearlman, J.S. Preprocessing Eo-1 Hyperion Hyperspectral Data to Support the Application of Agricultural Indexes. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1246–1259. [[CrossRef](#)]
43. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
44. Datt, B. Remote Sensing of Water Content in Eucalyptus Leaves. *Aust. J. Bot.* **1999**, *47*, 909. [[CrossRef](#)]
45. Goel, N.S.; Qin, W. Influences of Canopy Architecture on Relationships between Various Vegetation Indices and LAI and Fpar: A Computer Simulation. *Remote Sens. Rev.* **1994**, *10*, 309–347. [[CrossRef](#)]
46. Dash, J.; Curran, P.J. The MERIS Terrestrial Chlorophyll Index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413. [[CrossRef](#)]
47. Haboudane, D. Hyperspectral Vegetation Indices and Novel Algorithms for Predicting Green LAI of Crop Canopies: Modeling and Validation in the Context of Precision Agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [[CrossRef](#)]

48. Tucker, C.J.; Elgin, J.H.; McMurtrey, J.E.; Fan, C.J. Monitoring Corn and Soybean Crop Development with Hand-Held Radiometer Spectral Data. *Remote Sens. Environ.* **1979**, *8*, 237–248. [[CrossRef](#)]
49. Blackburn, G.A. Spectral Indices for Estimating Photosynthetic Pigment Concentrations: A Test Using Senescent Tree Leaves. *Int. J. Remote Sens.* **1998**, *19*, 657–675. [[CrossRef](#)]
50. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
51. CRAN-Package ‘mlr3fselect’ Instruction. Available online: <http://ftp2.de.freebsd.org/pub/misc/cran/web/packages/mlr3/fselect/mlr3fselect.pdf> (accessed on 15 November 2023).
52. Fonti, V. *Feature Selection Using LASSO*; VU Amsterdam: Amsterdam, The Netherlands, 2017; pp. 1–26.
53. Wei, Z. Inversion of Winter Wheat Growth Parameters and Yield Under Different Water Treatments Based on UAV Multispectral Remote Sensing. *Front. Plant Sci.* **2021**, *12*, 609876.
54. Dandrifosse, S.; Carlier, A.; Dumont, B.; Mercatoris, B. Registration and Fusion of Close-Range Multimodal Wheat Images in Field Conditions. *Remote Sens.* **2021**, *13*, 1380. [[CrossRef](#)]
55. Li, H.; Zhao, C.; Huang, W.; Yang, G. Non-Uniform Vertical Nitrogen Distribution within Plant Canopy and Its Estimation by Remote Sensing: A Review. *Field Crop. Res.* **2013**, *142*, 75–84. [[CrossRef](#)]
56. He, L.; Song, X.; Feng, W.; Guo, B.-B.; Zhang, Y.-S.; Wang, Y.-H.; Wang, C.-Y.; Guo, T.-C. Improved Remote Sensing of Leaf Nitrogen Concentration in Winter Wheat Using Multi-Angular Hyperspectral Data. *Remote Sens. Environ.* **2016**, *174*, 122–133. [[CrossRef](#)]
57. Duan, D.; Zhao, C.; Li, Z.; Yang, G.; Yang, W. Estimating Total Leaf Nitrogen Concentration in Winter Wheat by Canopy Hyperspectral Data and Nitrogen Vertical Distribution. *J. Integr. Agric.* **2019**, *18*, 1562–1570. [[CrossRef](#)]
58. Hu, P.; Chapman, S.C.; Jin, H.; Guo, Y.; Zheng, B. Comparison of Modelling Strategies to Estimate Phenotypic Values from an Unmanned Aerial Vehicle with Spectral and Temporal Vegetation Indexes. *Remote Sens.* **2021**, *13*, 2827. [[CrossRef](#)]
59. Ganeva, D.; Roumenina, E.; Dimitrov, P.; Gikov, A.; Jelev, G.; Dyulgenova, B.; Valcheva, D.; Bozhanova, V. Remotely Sensed Phenotypic Traits for Heritability Estimates and Grain Yield Prediction of Barley Using Multispectral Imaging from UAVs. *Sensors* **2023**, *23*, 5008. [[CrossRef](#)] [[PubMed](#)]
60. Li, W.; Li, D.; Liu, S.; Baret, F.; Ma, Z.; He, C.; Warner, T.A.; Guo, C.; Cheng, T.; Zhu, Y.; et al. RSARE: A Physically-Based Vegetation Index for Estimating Wheat Green LAI to Mitigate the Impact of Leaf Chlorophyll Content and Residue-Soil Background. *ISPRS-J. Photogramm. Remote Sens.* **2023**, *200*, 138–152. [[CrossRef](#)]
61. Duan, B.; Fang, S.; Zhu, R.; Wu, X.; Wang, S.; Gong, Y.; Peng, Y. Remote Estimation of Rice Yield With Unmanned Aerial Vehicle (UAV) Data and Spectral Mixture Analysis. *Front. Plant Sci.* **2019**, *10*, 204. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.