agronomy

MDPI

Article

Analysis of Growth Variation in Maize Leaf Area Index Based on Time-Series Multispectral Images and Random Forest Models

Xuyang Wang, Jiaojiao Ren and Penghao Wu *

College of Agriculture, Xinjiang Agricultural University, Urumqi 830052, China; wonderfulxy@outlook.com (X.W.); renjiaojiao789@xjau.edu.cn (J.R.)
* Correspondence: wupenghao@xjau.edu.cn

Abstract: The leaf area index (LAI) is a direct indicator of crop canopy growth and serves as an indirect measure of crop yield. Unmanned aerial vehicles (UAVs) offer rapid collection of crop phenotypic data across multiple time points, providing crucial insights into the evolving dynamics of the LAI essential for crop breeding. In this study, the variation process of the maize LAI was investigated across two locations (XD and KZ) using a multispectral sensor mounted on a UAV. During a field trial involving 399 maize inbred lines, LAI measurements were obtained at both locations using a random forest model based on 28 variables extracted from multispectral imagery. These findings indicate that the vegetation index computed by the near-infrared band and red edge significantly influences the accuracy of the LAI prediction. However, a prediction model relying solely on data from a single observation period exhibits instability (R^2 = 0.34–0.94, RMSE = 0.02–0.25). When applied to the entire growth period, the models trained using all data achieved a robust prediction of the LAI (R^2 = 0.79–0.86, RMSE = 0.12–0.18). Although the primary variation patterns of the maize LAI were similar across the two fields, environmental disparities changed the variation categories of the maize LAI. The primary factor contributing to the difference in the LAI between KZ and XD lies in soil nutrients associated with carbon and nitrogen in the upper soil. Overall, this study demonstrated that UAV-based time-series phenotypic data offers valuable insight into phenotypic variation, thereby enhancing the application of UAVs in crop breeding.

Keywords: unmanned aerial vehicle; leaf area index; time-series data; maize

Citation: Wang, X.; Ren, J.; Wu, P. Analysis of Growth Variation in Maize Leaf Area Index Based on Time-Series Multispectral Images and Random Forest Models. *Agronomy* **2024**, *14*, 2688. https://doi.org/10.3390/agronomy14112688

Academic Editor: Gniewko Niedbała

Received: 11 October 2024
Revised: 1 November 2024
Accepted: 12 November 2024
Published: 14 November 2024

1. Introduction

Maize, with an annual production exceeding 1 billion tons, has become a crucial global resource for food, fodder, and fuel [1,2]. Although the process of developing, testing, and releasing improved cultivars in crops such as maize is time-consuming, advancements in modern crop breeding technologies have substantially enhanced the selection efficiency for maize and numerous other crops [3,4]. The effectiveness of a crop breeding technique depends on the accurate and rapid collection of crop phenotype data [4,5]. Conventional methods for collecting quantitative plant traits have been relatively slow and costly. Therefore, there is a need for a cost-effective high-throughput approach for collecting quantitative maize traits.

The leaf area index (LAI) is a crucial phenotypic trait that significantly influences the photosynthesis, respiration, and evapotranspiration of a canopy [6,7]. The LAI is a primary descriptor of vegetation function and structure, and it has been utilized to assess biological changes within ecosystems [7,8]. Changes in the LAI not only indicate variations in leaf quantity within a canopy but also respond to plant physiological and phenology characteristics [8]. For example, the LAI offers estimations regarding crop yield and dry matter content, which indirectly indicate soil moisture and fertility conditions [9,10]. Early reports demonstrated that time-series data provided additional information on phenotypic changes and gene expression during crop growth for breeding work [11].

Furthermore, exploring changes in time-series phenotypes can assist in predicting potential yield fluctuations and evaluating the adaptability of crop phenotypes to environmental changes [11,12]. The accurate quantification of temporal changes in the LAI is indispensable for comprehending the physiological and ecological dynamics of plants at the canopy level. However, conventional approaches are constrained by factors such as destructive sampling, capacity, accuracy, and subjectivity, rendering them unsuitable for continuous observation [6,7]. Obtaining time-series LAI data on thousands of maize plots at multiple time points throughout the growing season presents a considerable challenge. Therefore, variability in the growth dynamics of the maize LAI remains unclear.

Optical sensors combined with deep learning algorithms have demonstrated their effectiveness as reliable high-throughput phenotyping tools for accurately estimating phenotype data with a high temporal resolution without the need for destructive methods [13]. The unmanned aerial system (UAS) is one of the most extensively employed technology platforms for high-throughput phenotyping. As technology advances, unmanned aerial vehicles (UAVs) are equipped with various sensors to create a UAS. These sensors include red–green–blue (RGB) sensors [14], multispectral sensors [15], hyperspectral sensors [16], and 3D point cloud sensors [17,18]. Three-dimensional point cloud sensors are primarily employed for measuring crop sizes [17]. Among them, multispectral sensors provide a quick and accurate reflection of sufficient spectral information for analysis, rendering multispectral imaging a promising technology for crop phenotype measurement [19]. In addition, machine learning techniques have been employed to establish a correlation between target traits and derived vegetation indices (VIs). Zhou et al. [20] compared the random forest (RF) model, support vector machine (SVM) model, and multivariate linear regression model (MLR) for estimating multiple traits of the maize canopy. Their findings revealed that the RF model demonstrated the most superior performance for trait prediction (R^2 values were 0.91–0.96). In the RF models, many classification and regression trees are constructed using randomly selected training datasets and random subsets of predictor variables for modeling outcomes. Therefore, RF often provides higher accuracy while maintaining some of the beneficial qualities of tree models [21]. Numerous studies have employed the RF model in different environments to estimate phenotypic traits in diverse crops, including wheat [22], quinoa [19], and sugarcane [23]. These findings demonstrate the adaptability of RF models to variable conditions, thereby enhancing their suitability for the development of predictive models applicable to diverse environments. Nevertheless, the VI used in prediction models is typically specific to particular observation periods, and the results of previous studies have not evaluated VIs applied in time-series models from field locations with varying environmental conditions. The incorporation of multiple types of VIs in prediction models at different field locations enables the analysis of the sensitivity of target traits to the VI type while minimizing confounding factors such as soil background reflectance, directional viewing angles, and atmospheric effects [19]. This will further optimize VI selection in predictive model development.

To do this, our main objectives were (i) to assess the sensitivity of multiple types of variables in LAI estimation across diverse environments, (ii) to assess the feasibility of using multispectral sensors and the RF model for estimating the LAI in large-scale maize breeding trials across various growth stages, and (iii) to investigate the dynamic changes and primary variations in the maize LAI. Acquiring time-series phenotypic data on crop performance can offer valuable insights into significant phenotypic changes and environmental effects during the breeding process. Such data have potential application value in modern crop breeding.

2. Materials and Methods

2.1. Experimental Site and Design

The maize experiment was performed from May to September 2023 at two locations in Qitai County, Xinjiang Uygur Autonomous Region, China: Wheat Experiment Station of Xinjiang Academy of Agricultural Sciences (89°12′ E, 44°13′ N) in Kanerz Township (KZ)

and Join Hope Seed Company Experiment Station (89°76′ E, 44°08′ N) in Xidi Township (XD) (Figure 1a). These sites experience a temperate continental climate, with an annual rainfall of 270 mm and predominantly gray desert soils. The maize experiment conducted at XD and KZ employed different fertilizer management strategies: (i) At XD, fertilizer management was the same as that in local practices. Nitrogen (N) fertilization was applied once, whereas Zinc (Zn) and potassium (K) fertilizers were applied four times during maize growth. The fertilizers used at XD were urea (800 kg ha^{-1}), zinc sulfate (133 kg ha^{-1}), and potassium sulfate (133 kg ha^{-1}). (ii) In KZ, N fertilizer was applied several times. The fertilizer used in KZ included urea (583 kg ha^{-1}), monopotassium phosphate (83 kg ha^{-1}), monoammonium phosphate (83 kg ha^{-1}), diammonium phosphate (83 kg ha^{-1}), and zinc sulfate (150 kg ha^{-1}).

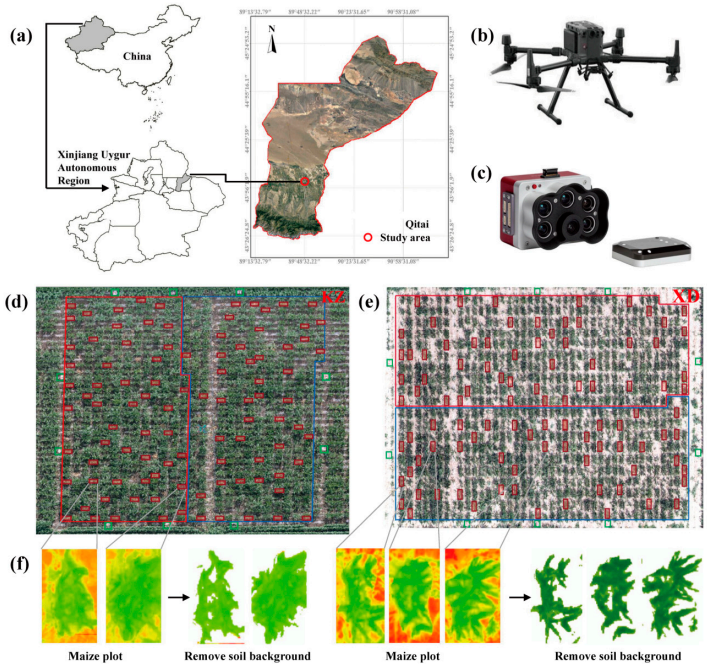

Figure 1. Location and design of study area. (**a**) Location of the study area; (**b**) DJI Matrice 300 RTK quadcopter; (**c**) MicaSense RedEdge-P multispectral sensor; (**d**) RGB map and design of KZ shows the 100 selected plots (red squares) for situ measurements, nine ground control points (green squares), and two replications (red and blue borders); (**e**) RGB map and design of XD, display content is the same as (**c**); (**f**) removed soil background from maize plot.

A natural population panel consisting of 399 diverse maize inbred lines was employed as the research material. This panel included several exceptional backbone elite inbred lines from China and some high-quality inbred lines imported from abroad. The trials were conducted using a randomized complete block design with two replications for each inbred line (Figure 1d,e). Each environmental treatment (XD and KZ) comprised 816 plots measuring 0.5 m × 1 m and spaced 0.5 m apart. In each plot, 10 seeds were planted with a spacing of 0.25 m between them.

2.2. Data Acquisition

Data collection occurred at eight time points starting in June. Two types of maize data were gathered: (i) manually collected LAI in situ and (ii) image data derived from UAS. In situ, LAI data were obtained by measuring the primary dimensions of all maize leaves. Representative plants of the two environmental treatments were randomly chosen for LAI measurements on the same date. To ensure the assessment of replicated accessions across various plots, sampling was conducted on 100 plots from each of the two fields. These measurements were performed five times at each survey plot, and the resulting average

LAI was considered the value for that particular plot (Figure 1d,e). For each maize plot and measurement date, LAI was computed as follows:

$$LAI = \frac{PLA(m^2)}{PS(m^2)}, \quad (1)$$

where *PLA* represents the total plant leaf area computed by summing the individual leaf area of all obtained plant leaves. *PS* is the crop spacing (estimated as the inverse of the plant density). The measured LAI values were employed as truth data for the RF model training and to evaluate the accuracy of the indirect LAI estimates.

The UAV platform used for collecting MicaSense RedEdge-P (MicaSense, Seattle, WA, USA) multispectral image data was a DJI Matrice 300 RTK quadcopter (DJI Innovations, Shenzhen, China) (Figure 1b,c). A multispectral sensor, with a resolution of 1456 × 1088 pixels, captured image data comprising blue (B; 459–491 nm), green (G; 546.5–573.5 nm), red (R; 660–676 nm), red edge (RE; 711–723 nm), and near-infrared (NIR; 813.5–870.5 nm) bands. Throughout the growing season, this study conducted eight UAV overflights. These overflights occurred at intervals of 7 days, corresponding to various growth stages of maize, including the six-leaf (V6), eight-leaf (V8), ten-leaf (V10), fourteen-leaf (V14), tasseling stage (VT), silking stage (R1), blister stage (R2), and milk stage (R3). To ensure accurate geometric correction, 10 ground control points (GCPs) were evenly distributed within the study area. Before and after the flight, the standardization of band values was achieved using a calibration board. The UAV followed predefined flight routes generated by the DJI software (version v10.01.08.01), ensuring at least 75% forward and side overlaps among images. Flights were consistently conducted at an altitude of 20 m with a speed of 3 m/s. Data collection occurred exclusively on sunny days between 11:00 AM and 13:00 PM.

2.3. Multispectral Image Data Processing

Figure 2 shows the LAI estimation framework outlined in this study. Pix4DMapper software (version 4.0, Pix4D S.A., Prilly, Switzerland) was used for image preprocessing. Subsequently, the preprocessed images were geo-calibrated using ArcMap software (version 10.8, EsriInc., Redlands, CA, USA), resulting in the generation of an orthographic image map of the field. For accurate extraction of maize spectral data, maize parts were segregated from the soil background (Figure 1f). An SVM classification performed using ENVI software (version 5.3, EXELIS., Boulder, CO, USA) was used to identify and extract maize pixels [19]. The SVM tool parameters were set to the default values (i.e., gamma in kernel function = 1, penalty parameter = 100, pyramid levels = 0, and classification probability threshold = 0). Based on ground truth data, the Kappa coefficients of the SVM classifications for all UAV imagery were >85%. Plot-level data extraction involved defining individual plot boundaries from orthographic images with an assigned plot ID corresponding to the maize inbred lines.

In this study, to evaluate the prediction performance of multiple types of VI in different environments, we extracted the pixel number (COUNT), 5 spectral bands, and 22 VIs (Table 1) from the time-series multispectral image for each maize plot. These data were used to develop prediction models of LAI at the plot level. To evaluate the predictive capacity of 22 VIs, they were divided into three classes: (i) VIs computed only in visible bands (R, G, and B); (ii) VIs incorporating the NIR band but excluding the RE band; and (iii) VIs computed including the RE band. In addition, in situ LAI measurements were obtained in a field with uneven emergence rates. To mitigate the potential overestimation of LAI in plots with higher pixels and to address scaling issues, the *COUNT* was normalized (*NCOUNT*) based on the emergence rate (ER) within each plot:

$$NCOUNT = COUNT \times \frac{ER_p}{ER_{max}}, \; ER_p = \frac{Gn}{Sn} \quad (2)$$

where *ERp* is the ER of each plot. *ERmax* denotes the maximum value of ER, with a default value of 1. *Gn* and *Sn* represent the number of germination and sowing in each plot, respectively. *NCOUNT* was employed for subsequent analysis.

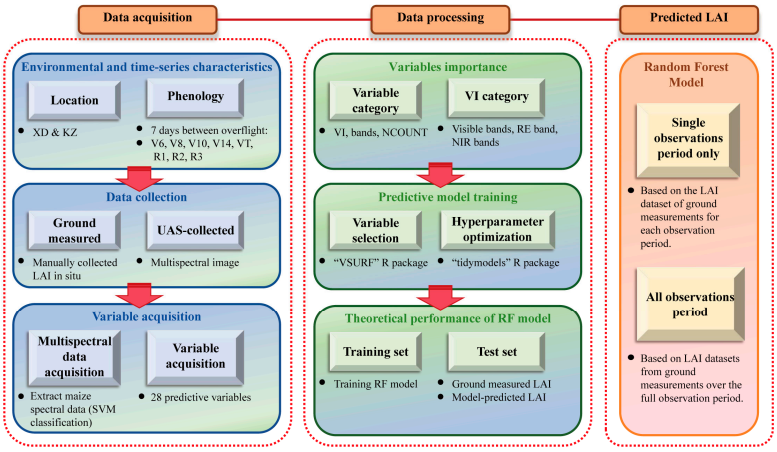

Figure 2. A framework regarding leaf area index (LAI) estimation in this study.

Table 1. Vegetation indices used in this study.

Full Name	Abbreviation	Formulas [a]	Ref.	Category
Excess green vegetation index	ExG	$(2G - R - B)/(G + R + B)$	[22]	Visible bands
Normalized green–red difference index	NGRDI	$(G - R)/(G + R)$	[19]	
Normalized pigment chlorophyll ratio index	NPCI	$(R - B)/(R + B)$	[19]	
Red–green ratio index	RGRI	R/G	[19]	
True color vegetation index	TCVI	$1.4 \times (2R - 2B)/2R - G - 2B + 0.4)$	[19]	
Green chlorophyll index	CIgreen	$NIR/G - 1$	[15]	NIR bands
Difference vegetation index	DVI	$NIR - R$	[19]	
Enhanced vegetation index	EVI	$2.5 \times (NIR - R)/(NIR + 6R - 7.5G + 1)$	[16]	
Green–blue normalized difference vegetation index	GBNDVI	$(NIR - (G + B))/(NIR + (G + B))$	[19]	
Green normalized difference vegetation index	GNDVI	$(NIR - G)/(NIR + G)$	[19]	
Normalized difference vegetation index	NDVI	$(NIR - R)/(NIR + R)$	[24]	
Plant senescence reflectance index	PSRI	$(R - G)/NIR$	[25]	
Ratio vegetation index	RVI	NIR/R	[15]	
Soil-adjusted vegetation index	SAVI	$(NIR - R)/(NIR+ R +0.25) + 0.25$	[15]	
Chlorophyll index red edge	CIre	$NIR/RE - 1$	[24]	RE bands
Modified enhanced vegetation index	MEVI	$2.5 \times (NIR - RE)/(NIR + 6RE - 7.5G + 1)$	[19]	
Normalized difference red edge index	NDRE	$(NIR - RE)/(NIR + RE)$	[11]	
NIR-RE normalized difference vegetation index	NIRRENDVI	$((NIR + RE)/2 - R)/((NIR + RE)/2 + R)$	[26]	
Red edge difference vegetation index	REDVI	$NIR - RE$	[27]	
Red edge normalized difference vegetation index	RENDVI	$(RE - R)/(RE + R)$	[19]	
Red edge ratio vegetation index	RERVI	NIR/RE	[28]	
Soil-adjusted red edge index	SARE	$(NIR - RE)/(NIR + RE + 0.25) + 0.25$	[19]	

[a] B, G, R, RE, and NIR represent the blue, green, red, red edge, and near-infrared bands, respectively.

The RF model has the capacity to efficiently process complex nonlinear relationships and avoid redundancy of variable information, thereby meeting the requirements for developing predictive models under multiple variables [20]. For model training, the RF model incorporates 28 predictive variables (NCOUNT, five spectral bands, and 22 VIs), with in situ LAI serving as the response variable. Two RF models were constructed for the ground-measured LAI datasets using a training dataset based on (1) a single observation period only and (2) all observation periods. Each dataset was randomly divided into two subsets: 75% served as the training set and 25% as the test set. In

developing the models, the optimal number of variables in each model was initially determined. Variable selection method was implemented in the "VSURF" R package [21]. Then, hyperparameter optimization was conducted using the "tidymodels" package (https://tidymodels.tidymodels.org/authors.html#citation, accessed on 15 October 2024.) in R software (version 4.3.3) to find the optimal hyperparameter combination. A set of values for the number of trees (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000) and the number of mtry (mtry is determined based on the outcomes of the variable selection in steps of 1) were tested to identify the optimal hyperparameter combination of RF. The hyperparameter tuning was conducted on each training model. The selected hyperparameter combinations of the optimal model were used for LAI prediction. During the process of tuning hyperparameters, the model-building dataset was randomly divided into training and tuning set and 5-fold cross-validation was performed. The significance of each variable was rated based on the percentage increase in mean squared error (%IncMSE) [19]. During model training, the reliability of the models was assessed by the mean squared error (MSE), root mean squared error (RMSE), and determination coefficient (R^2). The RF models were implemented using the R (version 4.3.3) package "randomForest (version 4.7.1.1)".

2.4. Soil Sampling and Measurement

In each field (XD and KZ), seven sampling points were chosen along the S-shape. Soil samples at four depths (0–20, 20–40, 40–60, and 60–80 cm) were collected in the quadrat using soil drill after removing the surface plants and litter. Five samples collected at the same depth from each sampling point were thoroughly mixed, resulting in a total of 56 composite samples. The samples were transported to the laboratory and subjected to drying to remove plant residues and other impurities. Subsequently, they were screened through different sizes (0.25 mm and 1 mm) of sieves before being used for further analysis.

A pH meter and a conductivity meter (SevenExcellence-S470, Zurich, Switzerland) were used to determine the soil pH value (pH) and electrical conductivity (EC), respectively, after water extraction (with a 1:5 ratio of soil to deionized water). The soil organic carbon (SOC) content was determined through $K_2Cr_2O_7$-H_2SO_4 oxidation. A flame photometer (Shanghaiyuefeng FP6400, Shanghai, China) was used to determine soil available potassium (K) and total potassium (TK). Furthermore, the alkaline hydrolysis diffusion approach was employed to measure the available soil nitrogen (N). Soil available phosphorus (P) was determined using spectrophotometry (Shimadzu UV-1780, Kyoto, Japan). Soil total nitrogen (TN) content was determined using the Kjeldahl digestion–distillation approach [29]. The soil carbon to nitrogen ratio (C/N) was computed as the ratio of SOC to TN.

2.5. Statistical Analyses

For statistical analysis and graphical representation, R (version 4.3.3) was used. The Least Significant Difference (LSD) test, performed at a 95% significance level, was employed to compare the average values of various degraded areas, soil depths, and physical and chemical properties of the soil and to assess the normality and uniformity of variance. The analysis primarily relied on the R package "randomForest " to examine the impact of fertilization factors on the LAI of maize.

3. Results

3.1. Presentation of Data Collected In Situ

Figures S1 and 3 illustrate the manually collected LAI in situ and multispectral image data, respectively. The manually collected and multispectral image data were obtained eight times during the maize growing period. The analysis of the ground measurements dataset indicated that the LAI exhibited a slightly skewed normal distribution until the V10 period. However, the ground measurement data were based on a portion of the maize material and may not be fully representative of the entire maize population.

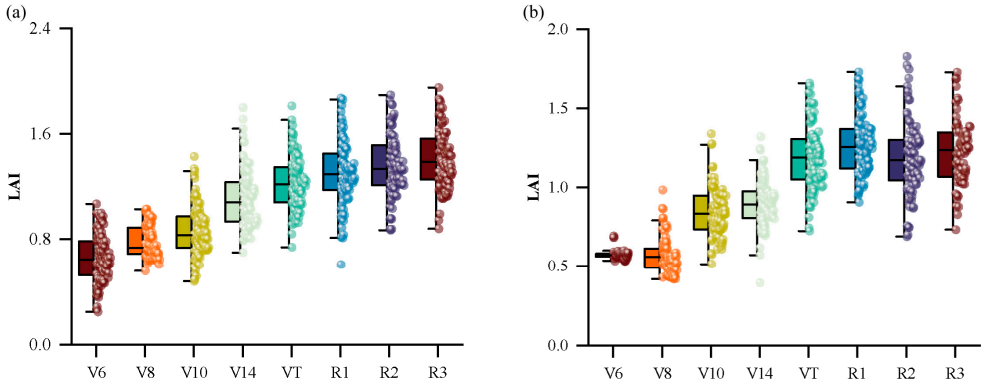

Figure 3. LAI data collected manually in situ during 8 periods. (**a**) KZ field; (**b**) XD field.

3.2. Significance of Variables in Various Random Forest Models

The sensitivity of each variable was evaluated based on the statistics of the variable importance ranking and the variable selection results in the LAI prediction model across different field locations. The significance of the 28 variables in estimating the LAI using an RF model was evaluated, and the findings are shown in Figure S2. In each model, the top seven (top 25%) of twenty-eight variables were defined as significant variables (Figure 4a,b). For all models in both fields, the VI incorporating the NIR band (excluding the red edge band) was the most important, followed by the VI with the RE band only. In minority models, a single band appears only in the top 25% of most important variables, indicating limited contributions of a single band to LAI predictions. Three VIs, namely RVI, NDVI, and NPCI, were among the seven highest-ranked variables when predicting the LAI in the KZ field. Conversely, in the XD field, REDVI, NCOUNT, and RVI are the most important LAI prediction models. In the prediction models established across all observation periods, NCOUNT is the most crucial variable in both fields.

Variable selection is often a necessary part of prediction model development. To provide a reference on the variable importance in the RF models across various datasets (single observation period only and all observation periods) for LAI prediction, the VIs obtained from the variable selection are labeled (Figure S3 and 4c). For all models in both fields, NCOUNT was the most frequently occurring variable, followed by DVI. Compared to the single band and VI calculated in visible bands, the VI incorporating the NIR band (excluding the red edge band) and the VI with the RE band only occur with greater frequency. These findings indicate that VIs computed using NIR and RE can reflect more information about maize growth compared to other VIs, as confirmed in different prediction models. Of these, DVI and NCOUNT are more sensitive to changes in the LAI and are therefore more effective in predicting the LAI.

3.3. Predictive Accuracy of the Models

To investigate the generality of the RF model, the models developed on a single observation period and all observation periods were applied to the XD and KZ fields. The model was developed using different datasets, and the results are presented in Figure 5 and Table S1. For prediction models trained with the dataset from a single observation period, models developed before the VT generally yield higher accuracy (MSE: 0.001 to 0.017) compared to those developed after the VT of maize (MSE: 0.003 to 0.041). Among the models established using a single observation period in the KZ field, the prediction accuracy before the VT of maize is significantly high. The goodness-of-fit metrics for the test dataset range from 0.62 to 0.94 for R^2 and from 0.19 to 0.03 for RMSE. In the single observation period model of the XD field, the prediction accuracy of the LAI is notably lower only during R1 and R3. The R^2 values for the test dataset are 0.50 and 0.34, respectively, with corresponding RMSE values of 0.18 and 0.20. In some models established in a single observation period, there was a considerable gap between the prediction accuracy of the

training dataset and that of the test dataset. Compared with the models established using a single observation period, the gap in prediction accuracy between the training and test datasets was smaller in the model developed using all observation periods (all stage). The R^2 values for the test dataset were 0.79 and 0.86, with corresponding RMSE values of 0.18 and 0.12, respectively. Although some models established using a single observation period were reliable, not every single observation period model was suitable for LAI prediction. The model constructed using all observation periods offers a greater advantage in predicting the LAI.

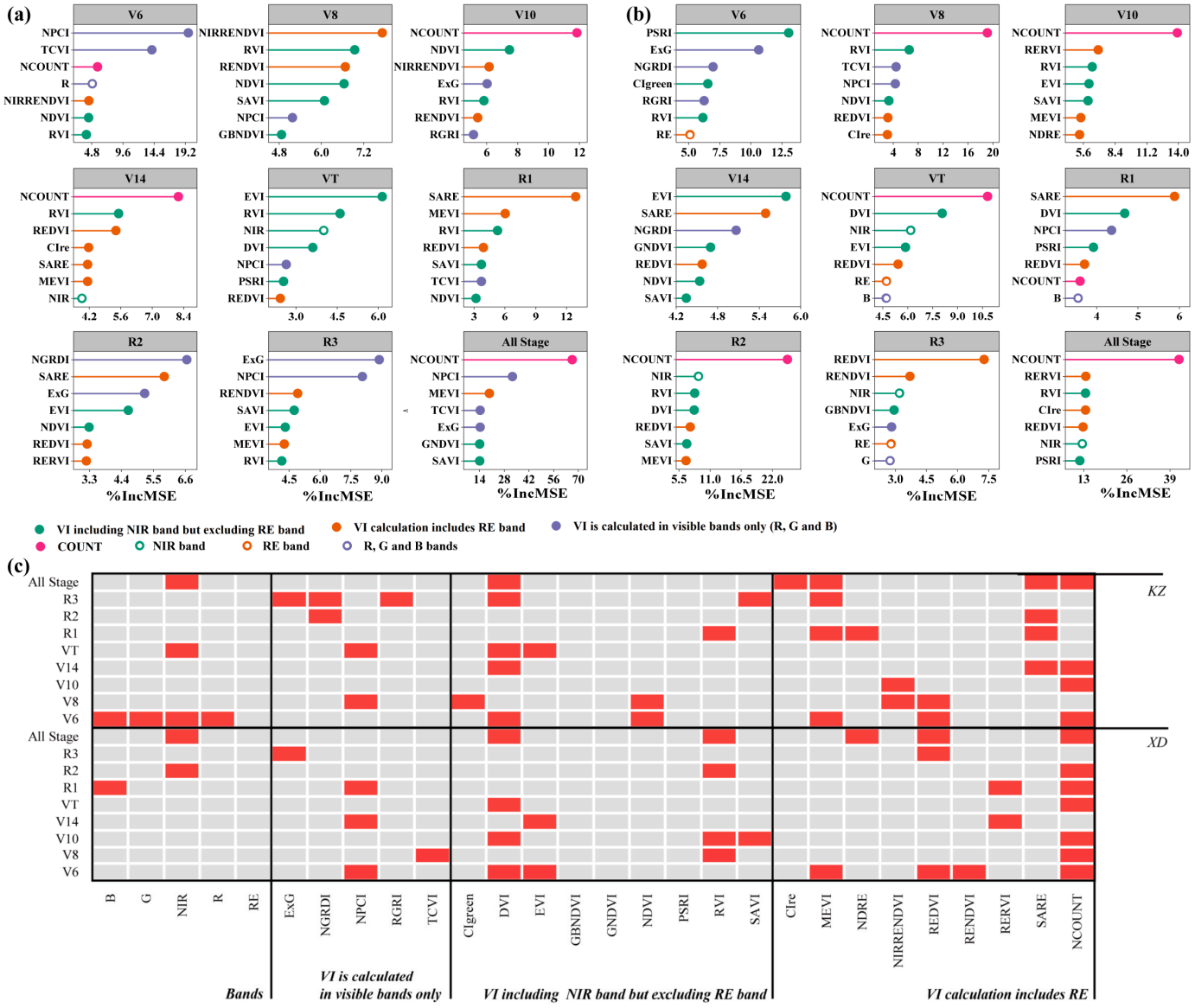

Figure 4. In (**a**) KZ and (**b**) XD field, top 25% most important variables in the random forest models and (**c**) variables obtained through variable selection on the KZ and XD.

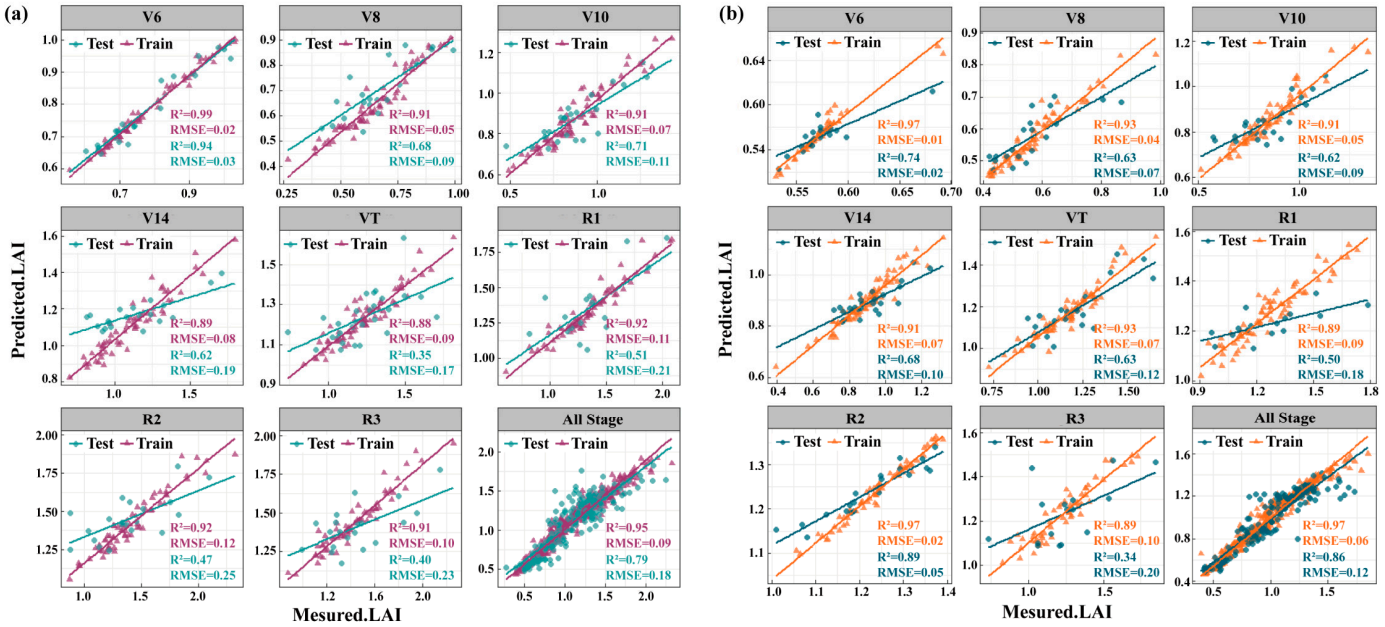

Figure 5. The UAV on (**a**) KZ field and (**b**) XD field. Comparison between manually measured LAI with LAI predicted by RF models.

3.4. Dynamics of the Maize Leaf Area Index

The RF models developed using all observation periods were employed to predict the LAI phenotype of maize. Figure 6 shows the spatial patterns of the LAI predicted by the RF model for the KZ and XD fields at eight growth stages. Prior to the maize VT stage, most of the KZ field and XD field had LAI values between 0.1 and 1.2. The LAI values higher than 1.4 were mainly distributed in the KZ field (after R1). Normal distributions were observed for V10, V14, VT, R1, R2, and R3 in the XD and KZ fields, whereas V6 and V8 showed slightly skewed normal distributions (Figure 7a,b). Notably, in the XD field, the LAI of maize exhibited a significant increase with the growth and development of maize in adjacent observation periods. In the KZ field, the LAI of maize exhibited no significant change during V6 and V8 and gradually increased after V8. A substantial difference in the LAI was observed between KZ and XD in each observation period, indicating that the diversity in the maize LAI between the two fields is attributable to the environmental conditions at various locations (Figure 7c).

To reduce the dimensionality of the time-series data, a principal component analysis (PCA) was conducted using the LAI data for all eight observation periods in the KZ and XD fields (Figure 8a,b). The first two principal components accounted for 55.0% and 21.1%, respectively, of the growing variation in the LAI data in the KZ field. In the XD field, PC1 (57.9%) and PC2 (17.0%) explained 74.9% of the growing variation in the LAI data. PC1 primarily demonstrated the magnitude of the LAI value, whereas PC2 represented the growth rate of maize leaves. Notably, this analysis yielded similar results in the KZ and XD fields. The PCA effectively captured most of the variations in the time-series LAI. A cluster analysis based on time-series data was conducted to further investigate the growth variation types of the maize LAI in the two fields (Figure 8c,d). In the KZ field, the 399 maize inbred lines were categorized into three large clusters (A, B, and C), whereas in the XD field, they were classified into two categories (A and B). These results indicate that environmental conditions influenced the variation categories of the LAI time-series data but not the variation patterns.

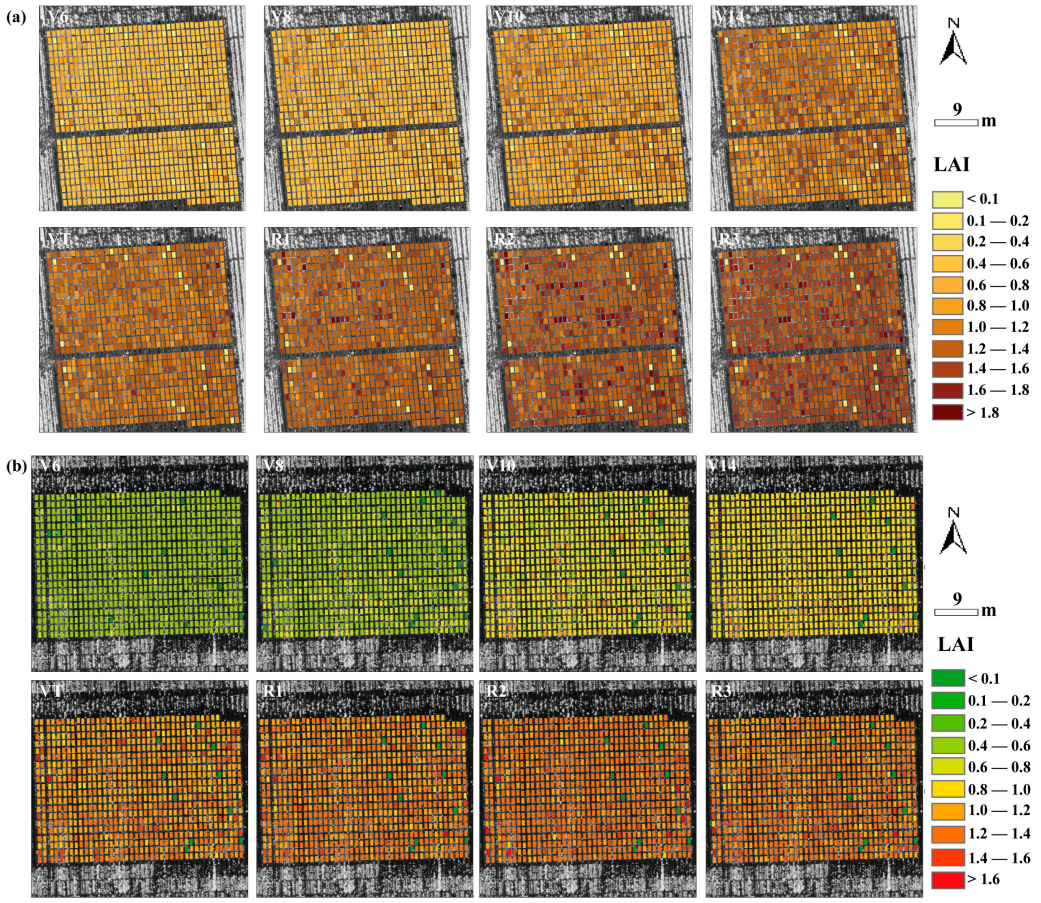

Figure 6. Maps of LAI using the random forest prediction model trained on all stages. (**a**) KZ; (**b**) XD.

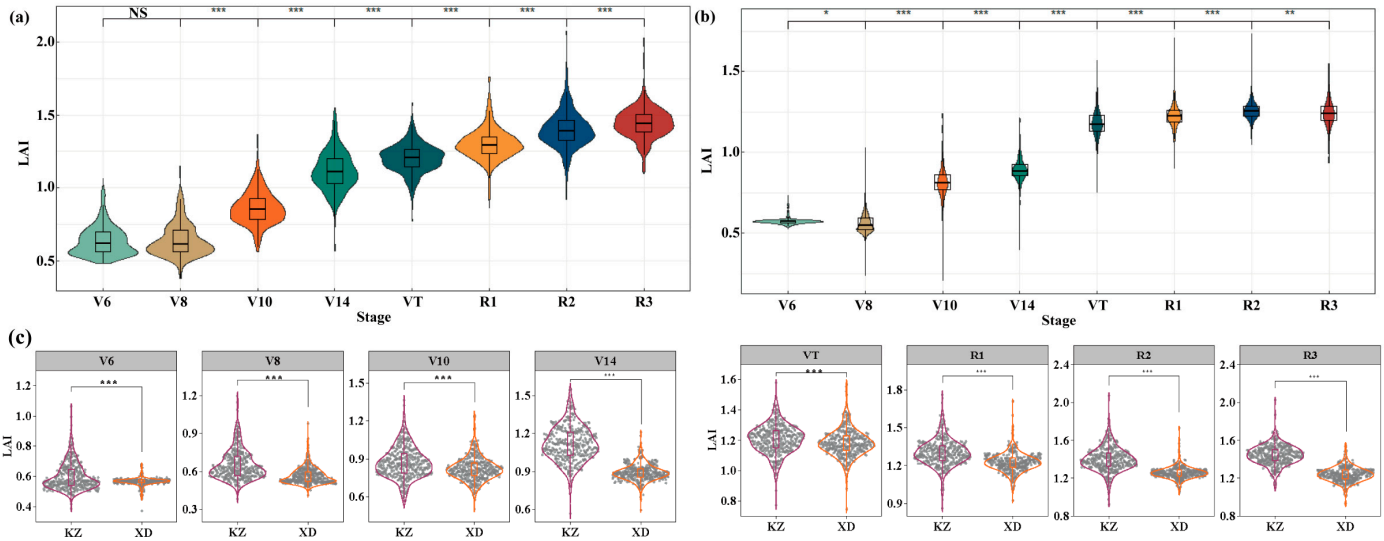

Figure 7. Violin plots of maize LAI using the RF model on (**a**) KZ field and (**b**) XD field. The horizontal axis represents different observation periods. (**c**) Comparison of maize LAI between KZ field and XD field in different observation periods. The central short line represents the median value. *, ** and *** are significant correlation at $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively. NS: not significantly different.

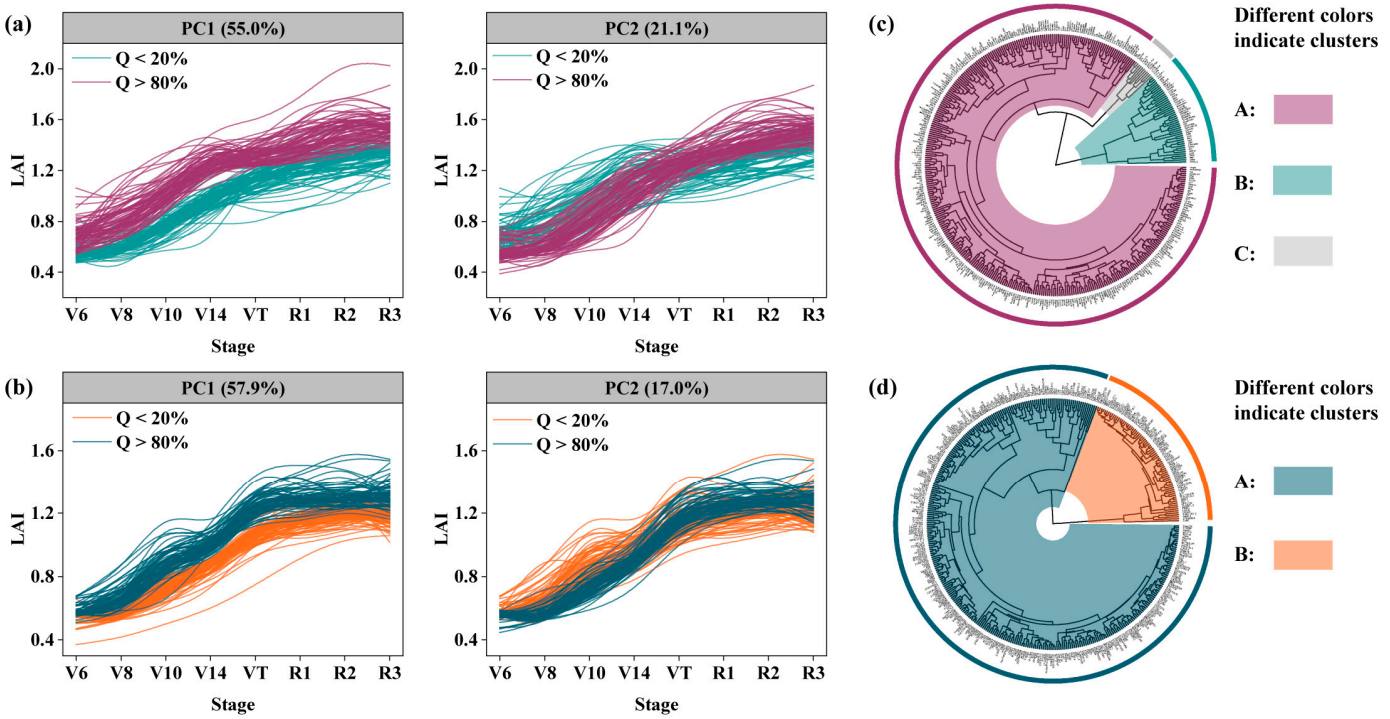

Figure 8. Principal component analysis (PCA) and cluster analysis of maize LAI change on KZ field (**a**,**c**) and XD field (**b**,**d**). Q, quantile. Different colors indicate clusters.

The analysis of variance for the eight observation periods revealed highly significant variations in genotypes, environments, and genotype-by-environment interactions (Table 2). Maize exhibited no significant effects from genotype-by-environment interaction in the later stage of growth (after entering VT). After entering the R2 stage, maize is primarily influenced by genotypes and environments, emphasizing the significance of this period for environmental impacts on the LAI. Various soil conditions resulting from fertilization play a vital role in changing the maize phenotype under the same climatic conditions.

Table 2. Analysis of variance for LAI.

Stage	F-Value		
	Environment	Genotype	Environment × Genotype
V6	73.14 ***	1.46 ***	1.40 ***
V8	242.87 ***	1.61 ***	1.42 ***
V10	55.29 ***	1.71 ***	1.22 **
V14	164.96 ***	1.60 ***	1.18 *
VT	22.58 ***	2.09 ***	1.17 *
R1	192.18 ***	1.78 ***	1.18 *
R2	459.75 ***	1.51 ***	0.92
R3	1357.43 ***	1.67 ***	1.02

*, ** and *** are significant correlation at $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively.

3.5. Factors Influencing the Leaf Area Index

To understand the difference in soil environments between KZ and XD resulting from various fertilization approaches, soil samples from the maize R2 stage were examined in layers. The findings are illustrated in Figure 9a. Significant differences were observed between the two fields (KZ and XD) in soil pH, SOC, TN, K, and C/N in the 0~20 cm soil layer. In the 20~40 cm soil layer, significant differences were observed in soil TK, N, P, and

K between KZ and XD. Furthermore, EC, TN, and C/N exhibited significant differences in deep-seated soil (40~60 cm and 60~80 cm).

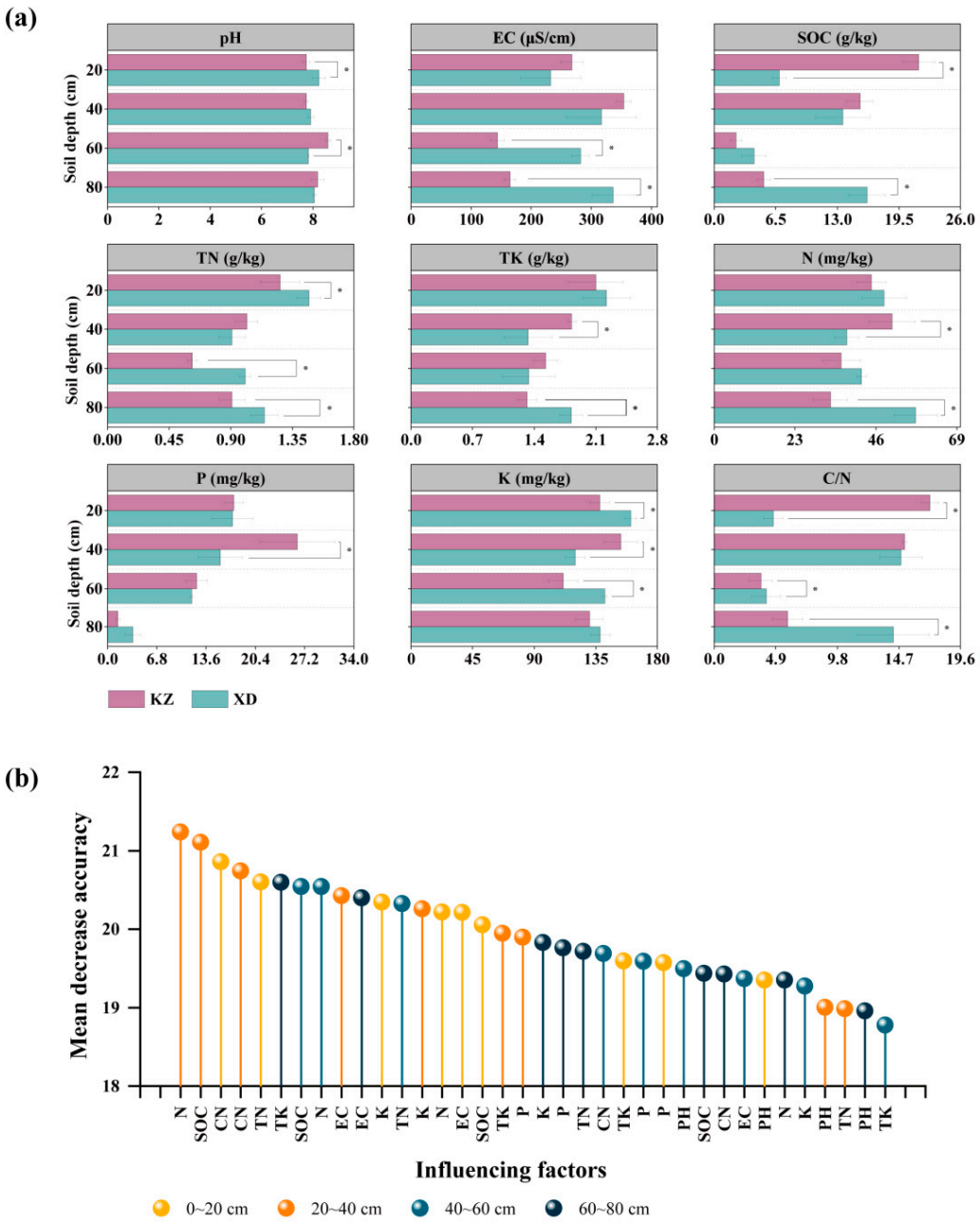

Figure 9. Effects of different fertilization treatments on LAI of maize. (**a**) Nutrient conditions of different soil layers in KZ field and XD field. (**b**) Effects of soil nutrient conditions in different soil layers on maize LAI. Higher values of mean decrease in accuracy indicate variables that are more important to the LAI. * is significant correlation at $p < 0.05$.

The multiple regression findings of the RF model illustrated the influence of different factors on the LAI of maize, as assessed by the Mean Decrease Accuracy metric. A higher Mean Decrease Accuracy value indicates a significant impact of the factor on the model's prediction accuracy (Figure 9b). In general, the upper soil (0~20 cm and 20~40 cm) exhibits a significant impact on the LAI of maize, whereas the lower (40~60 cm and 60~80 cm) soil has a relatively minor influence. The first five primary factors influencing the LAI of maize are associated with carbon and nitrogen in the soil. These findings indicate that the

contents of carbon and nitrogen in soil are the primary factors affecting the LAI of maize. However, it is crucial to recognize that these findings are based on the analysis conducted using an RF model. Additional statistical analysis and domain knowledge validation are required to fully understand the impact of these factors.

4. Discussion

4.1. Importance of Variables in Prediction Models

To predict LAI in maize at the plot level, RF machine learning models were employed to assess the impact of various variables on prediction accuracy. The findings indicate that VIs were typically more important than the individual spectral bands in predicting the LAI, which is consistent with previous research [19,29]. In this study, VIs computed using the NIR and RE bands proved useful in predicting the LAI. Similarly, in previous studies, VIs associated with the NIR and RE bands consistently exhibited strong correlations with the crop LAI, as evidenced in various crops such as vines [12], spinach plants [6], and maize [9]. In addition, NCOUNT, a non-VI variable, exhibited significant importance in some LAI prediction models (particularly in the RF model based on all observations). Among these models, NCOUNT exhibited significant importance for LAI estimation. This could be attributed to NCOUNT representing the purest crop pixels of the upper leaves, which directly reflects the canopy growth of maize plots [16]. Since NCOUNT is pixel data, it can avoid some VI saturation when the canopy structure reaches its peak [6,20]. The prediction models established multiple times during the pilot study underscore the potential of NCOUNT in investigating canopy dynamics. Simultaneously, it also indicates that NCOUNT as a variable has some limitations. These limitations may arise from the drought of farmland soil and background noise from the surrounding environment (e.g., weeds) [16,20,30,31]. Some prediction models were established under conditions of low soil moisture content in farmland due to the separation of farmland management between the two fields (KZ and XD). Low soil moisture conditions can cause crop leaves to curl, a scenario that directly affects the number of crop pixels extracted (COUNT). Nevertheless, manually measured data are unaffected by curved leaves, resulting in a mismatch between pixel data and manually measured data, consequently leading to the low contribution of NCOUNT in some models. Furthermore, the noise generated by weeds in the plot can lead to an incorrect estimation of NCOUNT [20], which will further affect the contribution of NCOUNT in the model.

4.2. Prediction Accuracy and Influencing Factors of the Random Forest Model

The findings also indicated disparities in the importance ranking of variables in the LAI prediction model across various observation periods or fields. The primary reason may be that the VI, as a commonly employed input variable, is sensitive to the environment during the estimation process [15]. Variations in environmental conditions, such as solar radiation, sun angle, and cloud coverage, across various measurement dates could lead to inconsistent measurements obtained by sensors deployed using UAVs [32]. Furthermore, the occlusion factors must be considered, as spectral features only account for apparent factors, while the LAI represents an aggregated level [19]. These observations indicate the importance of employing multiple VIs in constructing prediction models. However, the RF model developed using multiple variables may be overfitted and less robust than expected [33]. In this case, the model may exhibit a good fit in the training dataset but fail to achieve accurate predictions in the test dataset. This limitation could stem from the abundance and distribution of the training datasets [29]. In this study, using measured data from all observation periods to develop the model would broaden the abundance and distribution of data, which may enhance the model's robustness [33].

Overall, the findings demonstrate the utility of utilizing multiple types of variables and an RF model to predict the LAI. Therefore, the incorporation of multiple variables is recommended for predicting phenotypic traits across various environments. Although the prediction findings of the model developed during the observation period are promising,

there is still room for improvement. Since VIs are sensitive to environmental conditions, future studies could consider conducting multiple data collections under different environmental conditions to reduce the impact of environmental changes on sensor measurement consistency. Furthermore, one advantage of RF is its ability to accommodate extensive datasets [34]. By combining multiple variables, the accuracy of the RF model can be enhanced [35]. However, the challenge in multivariate estimations of crop phenotypes lies in the redundancy of information among variables, which often impacts accuracy. Therefore, it is crucial to use an appropriate number of variables [15]. The results indicate that multispectral sensors incorporating both RE and NIR bands provide appropriate additional information, thereby facilitating the identification of suitable variables. These advancements will further improve the efficacy of developing predictive models across multiple environments.

4.3. Environmental Effects on Time-Series LAI Data

Plant growth is a dynamic and complex process that responds to environmental changes. It is regulated by various sets of genes at different times and in different environments [36,37]. Analyzing the dynamics of plant growth can help identify geno loci that cannot be detected with single growth period data [4]. A previous investigation revealed that 30.6% of the loci linked to dynamic traits detected in genome-wide association studies were not identified in previous nondynamic studies [38]. Collecting time-series phenotypes of crops is advantageous for breeding programs. In this study, analyzing time-series data provides additional insights into the dynamics and trends crucial for breeding. PCA provides valuable information on the dynamics of the LAI over time for maize populations [4,6]. Our natural population of maize accessions exhibited varied time-series LAI patterns, as reflected in the PCA results, where the first two principal components represented the magnitude of the LAI value and growth rate of maize leaves, respectively. These findings could facilitate the selection and breeding of maize varieties associated with leaf characteristics, particularly those associated with growth rate. In addition, a cluster analysis of time-series trait data can effectively categorize inbred lines exhibiting similar trait dynamics and distinguish inbred line differences in the time dimension [4]. In this study, natural population accessions of maize were categorized into different clusters exhibiting distinct LAI growth patterns under the two environments. The categorization of LAI variations through cluster analysis prompted further investigation into whether growth patterns can be explained by environmental information. This interest stems from the differences in soil nutrient conditions influenced by the environment. Notably, soil nutrients associated with carbon and nitrogen, such as SOC, N, and C/N, were dominant factors. These carbon- and nitrogen-related soil nutrients not only directly provide essential nutrients for crop growth but also indirectly regulate the release of other nutrients in the soil [39]. Thus, soil nutrients associated with carbon and nitrogen play a crucial role in influencing the shifts in LAI over time.

High-throughput phenotyping techniques improved statistical capabilities by allowing large-scale field phenotyping data collection to be rapidly realized. These statistical capabilities provide the means to collect time-series data in multiple environments. Time-series data can provide more phenotypic variation information, rendering it a valuable addition to quantitative genetics investigations and plant breeding programs. With future developments in the UAS, the scope of applications for time-series data in crop genetics and breeding research is expected to broaden significantly.

5. Conclusions

This study employed a multispectral UAS based on an RF model to estimate the LAI of maize during the entire growth stage under two fertilization approaches. In addition, the primary variation patterns and environmental factors of the maize LAI varying with time were examined. The findings of this study indicated that (1) among the variables in the prediction model, VIs computed using NIR and RE bands exhibit superior performance

in LAI prediction; (2) integration multispectral sensors and RF model in large-scale maize breeding trials facilitate the prediction of the LAI at various growth stages; and (3) the soil environment significantly impacts the variation in the LAI in maize after the R2 stage, and such variation is substantially influenced by carbon and nitrogen in the upper soil.

This study holds significance in developing suitable prediction models for LAI estimation across specific crop and phenology periods, thereby enhancing the applications of UAVs in crop breeding.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/agronomy14112688/s1. Table S1: MSE statistics for different prediction models. Figure S1: Multispectral image data. Similar multispectral image data were acquired eight times during the maize growth period. Figure S2: Importance variables in the random forest models. Figure S3: Results of variable selection for random forest models.

Author Contributions: X.W. wrote the original draft and the review and editing. P.W. was responsible for funding acquisition, supervision, and project administration. J.R. engaged in writing—review and editing this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Research and Development Program of China (2021B02002-2, 2021A02001-2); Tianshan innovation team (2022D14017); Xinjiang Uygur Autonomous Region Natural Science Foundation key project (2022D01D34); Tianshan Yingcai (2022TSYCJU0003); Xinjiang Uygur Autonomous Region Major Science and Technology Special Projects (2022A02003-4); National Natural Foundation of China (Grant numbers: 32060484, U2003304, 32001561); and XinJiang Agriculture Research System (XJARS-02).

Data Availability Statement: The original contributions presented in this study are included in the article; further inquiries can be directed to the corresponding authors. Raw code is available upon request.

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

References

1. Wu, X.; Feng, H.; Wu, D.; Yan, S.; Zhang, P.; Wang, W.; Zhang, J.; Ye, J.; Dai, G.; Fan, Y.; et al. Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. *Genome Biol.* **2021**, *22*, 1–26. [CrossRef] [PubMed]
2. Zheng, Y.; Yuan, F.; Huang, Y.; Zhao, Y.; Jia, X.; Zhu, L.; Guo, J. Genome-wide association studies of grain quality traits in maize. *Sci. Rep.* **2021**, *11*, 9797. [CrossRef]
3. Singh, I.; Sheoran, S.; Kumar, B.; Kumar, K.; Rakshit, S. Speed breeding in maize (*Zea mays*) vis-à-vis in other crops: Status and prospects. *Indian J. Agric. Sci.* **2021**, *91*, 1267–1273. [CrossRef]
4. Wang, J.; Li, X.; Guo, T.; Dzievit, M.J.; Yu, X.; Liu, P.; Price, K.P.; Yu, J. Genetic dissection of seasonal vegetation index dynamics in maize through aerial based high-throughput phenotyping. *Plant Genome* **2021**, *14*, 20155. [CrossRef] [PubMed]
5. Sankaran, S.; Zhou, J.; Khot, L.R.; Trapp, J.J.; Mndolwa, E.; Miklas, P.N. High-throughput field phenotyping in dry bean using small unmanned aerial vehicle based multispectral imagery. *Comput. Electron. Agric.* **2018**, *151*, 84–92. [CrossRef]
6. Li, D.; Bai, D.; Tian, Y.; Li, Y.H.; Zhao, C.; Wang, Q.; Guo, S.; Gu, Y.; Luan, X.; Wang, R.; et al. Time series canopy phenotyping enables the identification of genetic variants controlling dynamic phenotypes in soybean. *J. Integr. Plant Biol.* **2023**, *65*, 117–132. [CrossRef]
7. Yamaguchi, H.; Yasutake, D.; Hirota, T.; Nomura, K. Nondestructive measurement method of leaf area index using near-infrared radiation and photosynthetically active radiation transmitted through a leafy vegetable canopy. *HortScience* **2023**, *58*, 16–22. [CrossRef]
8. Bao, X.; Wen, X.; Sun, X. Effects of environmental conditions and leaf area index changes on seasonal variations in carbon fluxes over a wheat–maize cropland rotation. *Int. J. Biometeorol.* **2022**, *66*, 213–224. [CrossRef]
9. Dimitrov, P.; Kamenova, I.; Roumenina, E.; Filchev, L.; Ilieva, I.; Jelev, G.; Alexander, G.; Martin, B.; Veneta, K.; Victor, K. Estimation of biophysical and biochemical variables of winter wheat through Sentinel-2 vegetation indices. *Bulg. J. Agric. Sci.* **2019**, *25*, 819.
10. Shao, G.; Han, W.; Zhang, H.; Liu, S.; Wang, Y.; Zhang, L.; Cui, X. Mapping maize crop coefficient Kc using random forest algorithm based on leaf area index and UAV-based multispectral vegetation indices. *Agric. Water Manag.* **2021**, *252*, 106906. [CrossRef]
11. Hashimoto, N.; Saito, Y.; Yamamoto, S.; Ishibashi, T.; Ito, R.; Maki, M.; Homma, K. Feasibility of yield estimation based on leaf area dynamics measurements in rice paddy fields of farmers. *Field Crops Res.* **2022**, *286*, 108609. [CrossRef]
12. Kang, Y.; Gao, F.; Anderson, M.; Kustas, W.; Nieto, H.; Knipper, K.; Yang, Y.; White, W.; Alfieri, J.; Torres-Rua, A. Evaluation of satellite Leaf Area Index in California vineyards for improving water use estimation. *Irrig. Sci.* **2022**, *40*, 531–551. [CrossRef]

13. Potgieter, A.B.; George-Jaeggli, B.; Chapman, S.C.; Laws, K.; Suárez, C.L.A.; Wixted, J.; Watson, J.; Eldridge, M.; Jordan, D. Multi-spectral imaging from an unmanned aerial vehicle enables the assessment of seasonal leaf area dynamics of sorghum breeding lines. *Front. Plant Sci.* **2017**, *8*, 1532. [CrossRef] [PubMed]
14. Rodene, E.; Xu, G.; Delen, P.S.; Zhao, X.; Smith, C.; Ge, Y.; Schnable, J.; Yang, J. A UAV-based high-throughput phenotyping approach to assess time-series nitrogen responses and identify trait-associated genetic components in maize. *Plant Phenome J.* **2022**, *5*, 20030. [CrossRef]
15. Shao, M.; Nie, C.; Zhang, A.; Shi, L.; Zha, Y.; Xu, H.; Yang, H.; Yu, X.; Bai, Y.; Liu, S. Quantifying effect of maize tassels on LAI estimation based on multispectral imagery and machine learning methods. *Comput. Electron. Agric.* **2023**, *211*, 108029. [CrossRef]
16. Zhu, W.; Sun, Z.; Yang, T.; Li, J.; Peng, J.; Zhu, K.; Li, S.; Gong, H.; Yun, L.; Li, B. Estimating leaf chlorophyll content of crops via optimal unmanned aerial vehicle hyperspectral data at multi-scales. *Comput. Electron. Agric.* **2020**, *178*, 105786. [CrossRef]
17. Li, M.; Shamshiri, R.R.; Schirrmann, M.; Weltzien, C.; Shafian, S.; Laursen, M.S. UAV oblique imagery with an adaptive micro-terrain model for estimation of leaf area index and height of maize canopy from 3D point clouds. *Remote Sens.* **2022**, *14*, 585. [CrossRef]
18. Banerjee, B.P.; Joshi, S.; Thoday-Kennedy, E.; Pasam, R.K.; Tibbits, J.; Hayden, M.; Spangenberg, G.; Kant, S. High-throughput phenotyping using digital and hyperspectral imaging-derived biomarkers for genotypic nitrogen response. *J. Exp. Bot.* **2020**, *71*, 4604–4615. [CrossRef]
19. Jiang, J.; Johansen, K.; Stanschewski, C.S.; Wellman, G.; Mousa, M.A.; Fiene, G.M.; Asiry, K.A.; Tester, M.; McCabe, M.F. Phenotyping a diversity panel of quinoa using UAV-retrieved leaf area index, SPAD-based chlorophyll and a random forest approach. *Precis. Agric.* **2022**, *23*, 961–983. [CrossRef]
20. Zhou, H.; Zhou, G.; Zhou, L.; Lv, X.; Ji, Y.; Zhou, M. The interrelationship between water use efficiency and radiation use efficiency under progressive soil drying in maize. *Front. Plant Sci.* **2021**, *12*, 794409. [CrossRef]
21. Speiser, J.L.; Miller, M.E.; Tooze, J.; Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. [CrossRef] [PubMed]
22. Chen, Q.; Zheng, B.; Chenu, K.; Hu, P.; Chapman, S.C. Unsupervised plot-scale LAI phenotyping via UAV-based imaging, modelling, and machine learning. *Plant Phenomics* **2022**, *4*, 9768253. [CrossRef] [PubMed]
23. Poudyal, C.; Costa, L.F.; Sandhu, H.; Ampatzidis, Y.; Odero, D.C.; Arbelo, O.C.; Cherry, R.H. Sugarcane yield prediction and genotype selection using unmanned aerial vehicle-based hyperspectral imaging and machine learning. *Agron. J.* **2022**, *114*, 2320–2333. [CrossRef]
24. Njane, S.N.; Tsuda, S.; Marrewijk, B.M.; Polder, G.; Katayama, K.; Tsuji, H. Effect of varying UAV height on the precise estimation of potato crop growth. *Front. Plant Sci.* **2023**, *14*, 1233349. [CrossRef]
25. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]
26. Xie, Q.; Dash, J.; Huang, W.; Peng, D.; Qin, Q.; Mortimer, H.; Casa, R.; Pignatti, S.; Laneve, G.; Pascucci, S. Vegetation indices combining the red and red-edge spectral information for leaf area index retrieval. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1482–1493. [CrossRef]
27. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
28. Jasper, J.; Reusch, S.; Link, A. Active sensing of the N status of wheat using optimized wavelength combination: Impact of seed rate, variety and growth stage. *Precis. Agric.* **2009**, *9*, 23–30.
29. Chen, M.; Yang, Z.; Abulaizi, M.; Hu, Y.; Tian, Y.; Hu, Y.; Yu, G.; Zhu, X.; Yu, P.; Jia, H. Soil bacterial communities in alpine wetlands in arid Central Asia remain stable during the seasonal freeze–thaw period. *Ecol. Indic.* **2023**, *156*, 111164. [CrossRef]
30. Liu, S.; Jin, X.; Nie, C.; Wang, S.; Yu, X.; Cheng, M.; Shao, M.; Wang, Z.; Tuohuti, N.; Bai, Y. Estimating leaf area index using unmanned aerial vehicle data: Shallow vs. deep machine learning algorithms. *Plant Physiol.* **2021**, *187*, 1551–1576. [CrossRef]
31. Murguia-Cozar, A.; Macedo-Cruz, A.; Fernandez-Reynoso, D.S.; Transito, J.A.S. Recognition of maize phenology in sentinel images with machine learning. *Sensors* **2021**, *22*, 94. [CrossRef] [PubMed]
32. Bhandari, M.; Ibrahim, A.M.; Xue, Q.; Jung, J.; Chang, A.; Rudd, J.C.; Maeda, M.; Rajan, N.; Neely, H.; Landivar, J. Assessing winter wheat foliage disease severity using aerial imagery acquired from small Unmanned Aerial Vehicle (UAV). *Comput. Electron. Agric.* **2020**, *176*, 105665. [CrossRef]
33. Ganeva, D.; Roumenina, E.; Dimitrov, P.; Gikov, A.; Jelev, G.; Dragov, R.; Bozhanova, V.; Taneva, K. Phenotypic traits estimation and preliminary yield assessment in different phenophases of wheat breeding experiment based on UAV multispectral images. *Remote Sens.* **2022**, *14*, 1019. [CrossRef]
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
35. Xu, H.; Zhang, X.; Ye, Z.; Jiang, L.; Qiu, X.; Tian, Y.; Zhu, Y.; Cao, W. Machine learning approaches can reduce environmental data requirements for regional yield potential simulation. *Eur. J. Agron.* **2021**, *129*, 126335. [CrossRef]
36. Miao, C.; Xu, Y.; Liu, S.; Schnable, P.S.; Schnable, J.C. Increased power and accuracy of causal locus identification in time series genome-wide association in sorghum. *Plant Physiol.* **2020**, *183*, 1898–1909. [CrossRef]
37. Wang, C.; Pan, W.; Song, X.; Yu, H.; Zhu, J.; Liu, P.; Li, X. Predicting plant growth and development using time-series images. *Agronomy* **2022**, *12*, 2213. [CrossRef]

38. Fei, S.; Hassan, M.A.; Xiao, Y.; Rasheed, A.; Xia, X.; Ma, Y.; Fu, L.; Chen, Z.; He, Z. Application of multi-layer neural network and hyperspectral reflectance in genome-wide association study for grain yield in bread wheat. *Field Crops Res.* **2022**, *289*, 108730. [CrossRef]
39. Sainju, U.M.; Liptzin, D.; Ghimire, R.; Dangi, S. Relationship between soil carbon and nitrogen, soil properties, and dryland crop yields. *Agron. J.* **2022**, *114*, 395–414. [CrossRef]