agronomy

MDPI

Article

Time-Resolved UAV Multispectral Characterization of Cultivar-Dependent Yield Variation in Rice Under Planting-Density Treatments

Jing Yang [1,†], Jian Liu [1,†], Xuetong Zhu [1], Yikai Ding [1], Mengya Li [1], Xin Lu [1], Xiuru Wu [2] and Fei Zhao [1,*]

[1] College of Agronomy and Resources and Environment, Tianjin Agricultural University, Tianjin 300392, China
[2] China United Property Insurance Co., Ltd., Tianjin Branch, 6/F, Tower A, Tianjin Lujiazui Financial Plaza, No. 170 Beima Road, Hongqiao District, Tianjin 300121, China
* Correspondence: 13920357546@126.com
† These authors contributed equally to this work.

Abstract

UAV multispectral imagery can support rice yield estimation, but spectral–yield relationships may change with cultivar and planting density. We evaluated three japonica rice cultivars under five planting-density treatments in a split-plot experiment and collected UAV multispectral imagery on 13 dates. We examined five vegetation indices: NDVI, RVI, NDRE, OSAVI, and CLRE. Cultivar significantly affected grain yield and yield components. Planting density had no significant main effect, while the cultivar × density interaction was significant for grain yield. Spectral–yield correlations were strongest during 19–25 August, which broadly matched reproductive development based on field observations. NDRE and CLRE showed the most consistent associations. A four-index PLSR model using 19 August data achieved an outer-LOOCV R^2 of 0.773 and an RMSE of 530.28 kg ha^{-1}. Tuned RF gave comparable performance (R^2 = 0.783, RMSE = 518.60 kg ha^{-1}). Adjustment for cultivar, planting density, and block substantially reduced the pooled mid-August correlations. However, the four-index spectral set still improved prediction beyond cultivar, planting density, and their interaction. Grouped validation changed little when density levels or cultivar × density combinations were withheld, but performance declined sharply when an entire cultivar was excluded. These results show that mid-August UAV observations were informative in this experiment, but wider use will require validation across more cultivars, years, and sites.

Keywords: rice; UAV multispectral imagery; planting density; cultivar variation; red-edge vegetation indices; observation timing; yield estimation

Received: 17 August 2026
Revised: 10 September 2026
Accepted: 18 September 2026
Published: 20 September 2026

1. Introduction

Rice yield formation depends on the coordinated development of crop population structure and source–sink processes, both of which are sensitive to planting density. Changes in planting density can modify tiller production, panicle number, canopy light distribution, photosynthetic performance, and grain filling, thereby affecting the formation of grain yield [1–3]. However, these responses are not uniform across rice genotypes. Cultivars differ in plant architecture, tillering capacity, photosynthetic characteristics, panicle traits, and adaptation to population density, and may therefore exhibit different canopy and yield responses under the same planting-density treatment [4,5]. Cultivar and planting density can consequently generate structured agronomic variation in canopy development and yield formation. This heterogeneity is also relevant to remote-sensing-based yield estimation because canopy spectral signals integrate structural and physiological properties that are themselves influenced by genotype and management [6,7].

Unmanned aerial vehicle (UAV) remote sensing provides high spatial resolution and flexible temporal coverage and has become an established tool for rice growth monitoring and yield estimation [8]. Previous studies have shown that multispectral vegetation indices, canopy features, and multi-temporal observations can capture substantial variation in grain yield, while statistical, machine-learning, and deep-learning approaches have further expanded the available prediction strategies [9–13]. Multi-year and multi-variety studies have also demonstrated that useful yield predictions can be obtained under considerably more heterogeneous experimental conditions than those represented in many early UAV studies [7,11–14]. Thus, the feasibility of UAV-based rice yield prediction is well established. A more important question is whether an apparently strong spectral–yield relationship reflects information that remains stable across different agronomic conditions. When cultivars or management treatments contribute simultaneously to canopy characteristics and final yield, prediction accuracy obtained from pooled experimental plots may partly reflect this underlying agronomic structure rather than a uniformly transferable spectral response [6,7,15].

The timing of image acquisition introduces another important source of variation. Rice canopy structure, leaf area, chlorophyll status, panicle exposure, and the relative contribution of vegetation and background change continuously during crop development, causing spectral–yield relationships to vary across observation periods [9,10,16]. Importantly, the most informative sensing period has not been identical among studies. Strong single-stage relationships have been reported around booting or initial heading, cumulative time-series analyses have identified particular combinations of developmental stages rather than the full seasonal sequence as optimal, and recent multi-variety studies have reported increasing predictive performance toward reproductive development and flowering–grain filling [9,10,16,17]. These differences argue against assuming a universally optimal calendar date for rice yield prediction. Instead, observation timing should be interpreted in relation to crop development and the experimental population being monitored. This issue becomes particularly relevant when cultivars differ in growth duration [18]. Phenology-aware studies have shown that developmental misalignment among cultivars, environments, or management conditions can reduce the consistency of remote-sensing relationships [7,19], whereas phenological adjustment or alignment can improve model performance [7,19]. UAV imagery has also been used to detect rice phenological stages directly, further emphasizing the importance of developmental context in temporal monitoring [20]. Accordingly, identifying an informative observation window may be more meaningful than simply increasing the number of UAV acquisitions.

Model transferability is closely related to, but distinct from, within-experiment prediction accuracy. Independent rice studies have shown that empirical UAV yield relationships can change across years, cultivars, climatic environments, and sensors [6,7,15]. More stringent validation strategies have therefore been used to evaluate these sources of domain shift [21,22]. Zheng et al. independently assessed model transfer across years, cultivars, and sensors, while Zhou et al. compared stratified and cultivar-group sampling in multi-variety rice breeding trials [11,15]. More recently, independent temporal validation across three growing seasons has demonstrated that models can retain useful predictive ability across years under multi-variety and multi-nitrogen conditions [14], whereas a large multi-genotype study found that yield prediction remained strongly genotype dependent despite relatively robust remote sensing of nitrogen status [7]. These studies show that conventional plot-level validation and agronomic-domain transfer address different aspects of model performance. Nevertheless, temporal dependence, cultivar effects, management heterogeneity, and transferability have rarely been examined together within a structured cultivar × management experiment. In particular, it remains unclear to what extent strong pooled spectral–yield relationships retain incremental predictive information after cultivar and management structure are accounted for, and whether predictive stability differs when the withheld domain represents a management level rather than an entire cultivar.

Based on these gaps, we used a field experiment with three japonica cultivars, five planting-density treatments, and 13 UAV multispectral observation dates. We asked three questions. First, when were spectral–yield associations most informative during the growing season? Second, did multispectral variables add predictive value after cultivar and planting-density structure were considered? Third, how did model performance change when density levels, cultivar × density combinations, or entire cultivars were withheld? We expected spectral–yield associations to become stronger during reproductive development than during early vegetative growth (H1). We also expected part of the pooled association to reflect cultivar and management structure, while some spectral information would remain after adjustment (H2). Finally, we expected transfer to an unseen cultivar to be more difficult than transfer across the tested planting-density range (H3).

2. Materials and Methods

2.1. Plant Materials

Three japonica rice cultivars, Jingrun, Tiannongdao 307, and Jinchuan 1, were used in this study. Jingrun is a conventional japonica cultivar, Tiannongdao 307 is a registered rice cultivar in China, and Jinchuan 1 is characterized by high eating quality. All three cultivars were developed by the Quality Rice Breeding Team of Tianjin Agricultural University. The cultivars were included to provide multiple genetic backgrounds within the same field experiment rather than to represent the full diversity of japonica rice germplasm.

2.2. Experimental Site, Split-Plot Design, and Field Management

The field experiment was conducted in 2025 in Lintingkou Town, Baodi District, Tianjin, China (39.57° N, 117.46° E). The region has a warm-temperate, semi-humid continental monsoon climate, with an annual mean temperature of approximately 11.6 °C and annual precipitation of approximately 612.5 mm. The experimental field has fluvo-aquic soil with a medium-loam texture and relatively uniform irrigation and drainage conditions.

The experiment followed a split-plot arrangement with three blocks. Planting density was the whole-plot factor, and cultivar was the subplot factor within each density whole plot. Each block contained all five density treatments, and each density whole plot contained the three cultivar subplots, giving a total of 3 blocks × 5 densities × 3 cultivars = 45 experimental subplots. Each subplot measured 6.7 m × 2.4 m, corresponding to an area of 16.08 m^2. The 45 subplots covered 723.6 m^2 in total, excluding the isolation strips between blocks. Two 0.6-m-wide isolation strips separated the three blocks. Density treatments and cultivar subplots were randomized at their corresponding experimental levels. All factorial analyses followed this split-plot structure. All 45 subplots had complete data for the variables used in each analysis.

Row spacing was fixed at 30 cm, whereas hill spacing was 14, 17, 20, 23, and 26 cm for D1–D5, corresponding to approximately 23.81, 19.61, 16.67, 14.49, and 12.82 × 10^4 hills ha^{-1}, respectively. Three seedlings were transplanted per hill. Seedlings were raised in a dry-bed nursery covered with plastic film from 15 April and were transplanted on 20 May at approximately the 3.8-leaf stage.

Basal, regreening, tillering, and panicle fertilizers were applied on 19 May, 30 May, 12 June, and 15 July at 600.0, 90.0, 165.0, and 67.5 kg ha^{-1}, respectively. The compound fertilizer contained N–P_2O_5–K_2O at a ratio of 24–14–10, with a total nutrient content of at least 48%. The fertilization schedule is summarized in Table 1. Except for planting density, all treatments received the same fertilizer schedule and rate, irrigation management, and pest, disease, and weed control practices.

Table 1. Fertilization schedule and application rates in the rice field.

Fertilization Stage	Application Date	Relation to Transplanting	Compound Fertilizer Application Rate (kg ha^{-1})
Basal fertilizer	19 May 2025	1 day before transplanting	600.0
Regreening fertilizer	30 May 2025	10 days after transplanting	90.0
Tillering fertilizer	12 June 2025	23 days after transplanting	165.0
Panicle fertilizer	15 July 2025	56 days after transplanting	67.5

2.3. Temporal and Phenological Context of UAV Observations

UAV multispectral imagery was acquired on 13 dates throughout the rice growing season using a DJI Mavic 3M platform (DJI, Shenzhen, China). Field observations were used to place the UAV acquisition dates in a phenological context. These observations indicated that the mid-August acquisitions broadly coincided with the transition into reproductive development, although developmental timing differed somewhat among cultivars. We analyzed the UAV data by calendar date and did not apply formal phenological normalization. We therefore treated 19–25 August as a calendar-defined sensitive observation window rather than as a synchronized physiological stage across cultivars.

2.4. UAV Multispectral Data Acquisition, Image Processing, and ROI Extraction

Multispectral imagery was acquired using a DJI Mavic 3 Multispectral UAV (DJI, Shenzhen, China) equipped with a real-time kinematic positioning system and four multispectral bands centered at approximately 560 nm (green), 650 nm (red), 730 nm (red edge), and 860 nm (near infrared). Flights were conducted under clear or slightly cloudy conditions with relatively stable illumination and low wind speed and, whenever possible, between 10:00 and 14:00. Flight altitude was 20 m, forward and side overlaps were 80% and 70%, respectively, and flight speed was approximately 1.7 m s^{-1}.

Images were acquired on 13 dates: 16 June, 24 June, 27 June, 4 July, 9 July, 16 July, 23 July, 6 August, 14 August, 19 August, 25 August, 10 September, and 14 October. Multispectral images were processed in DJI Terra version 5.0.1 (DJI, Shenzhen, China) to generate georeferenced single-band orthomosaics, which were subsequently imported into ENVI 5.6 (Harris Geospatial Solutions, Inc., Broomfield, CO, USA) for band stacking, vegetation-index calculation, and plot-level extraction.

Five vegetation indices were calculated:

$$\mathrm{NDVI} = \frac{NIR - R}{NIR + R} \tag{1}$$

$$\mathrm{RVI} = \frac{NIR}{R} \tag{2}$$

$$\mathrm{NDRE} = \frac{NIR - RE}{NIR + RE} \tag{3}$$

$$\mathrm{OSAVI} = \frac{(1 + 0.16)(NIR - R)}{NIR + R + 0.16} \tag{4}$$

$$\mathrm{CLRE} = \frac{NIR}{RE} - 1 \tag{5}$$

where *NIR*, *R*, and *RE* denote near-infrared, red, and red-edge band values, respectively. These expressions define the five vegetation indices used in the temporal analyses [23–26].

Forty-five plot-level regions of interest (ROIs) were manually delineated within the visible boundaries of the experimental subplots. ROI boundaries were placed inside the target subplot to avoid areas outside the experimental plot and to minimize inclusion of border rows, ridges, and pixels from adjacent plots. No additional pixel-level soil/background or shadow mask was applied after ROI delineation. Vegetation indices were calculated using all pixels within each retained ROI, and the mean value was used as the subplot-level spectral measurement. Plot identification and spatial correspondence were maintained across acquisition dates to preserve one-to-one matching among UAV observations, agronomic measurements, and yield data. Residual contributions from within-plot background, canopy gaps, and illumination variability were therefore retained in the plot-level spectral measurements.

2.5. Agronomic Measurements and Grain Yield

At maturity, three hills were randomly selected from the interior of each subplot, while border rows were avoided. The sampled hills were used for yield-component measurements, including effective panicle number, grains per panicle, seed-setting rate, and thousand-grain weight. Effective panicles were counted from the three sampled hills and converted to panicles m^{-2} based on the corresponding planting density. The values obtained from the three sampled hills were averaged to represent each subplot.

Grain yield was measured independently of the yield-component sampling. At maturity, the entire 16.08 m^2 area of each experimental subplot was harvested separately for grain-yield determination. Grain was threshed and weighed, and yield was standardized to a moisture content of 15%. Plot yield was then converted to kg ha^{-1} based on the harvested subplot area.

2.6. Statistical Analysis of Agronomic and Spectral Relationships

Cultivar and planting density were treated as fixed experimental factors, and block represented the replication structure. Factorial analyses followed the split-plot design. The density effect was tested against the Block × Density whole-plot error term, whereas cultivar and Cultivar × Density were tested against the subplot residual. Partial η^2 was calculated for each experimental effect. When an interaction was significant, simple effects were examined to characterize treatment responses within cultivars. Statistical significance was evaluated at $p < 0.05$.

Temporal relationships between vegetation indices and grain yield were evaluated using Pearson correlations across all 45 subplots. Five indices were assessed on each of 13 UAV acquisition dates, giving 65 tests. The complete family of tests was subjected to Benjamini–Hochberg false-discovery-rate and Holm corrections, and both adjusted and unadjusted results were retained.

To examine whether pooled correlations masked differences among cultivars, we also calculated Pearson correlations separately for each cultivar across all 65 date × vegetation-index combinations. Each cultivar analysis included 15 observations. We treated these results as structural diagnostics rather than as cultivar-specific predictive models.

For each of the 65 combinations, we also calculated two partial correlations. The first analysis removed cultivar effects from both the vegetation index and grain yield. The second removed cultivar, planting density, and block. We then calculated Pearson correlations between the two sets of residuals. We calculated two-sided p values using the residual degrees of freedom and the rank of the adjustment design matrix. We corrected the 65 tests within each analysis family separately using the Benjamini–Hochberg and Holm procedures. These analyses tested how much of the pooled temporal association was related to cultivar and experimental structure. We did not interpret these associations causally.

2.7. Yield Modeling, Model Comparison, and Validation

For single-index benchmarking, NDVI, RVI, NDRE, OSAVI, and CLRE acquired on 19 August were each fitted separately using univariate linear regression. Each model was evaluated by outer leave-one-out cross-validation (LOOCV), and predictive performance was calculated from the resulting out-of-fold predictions using the same metrics as those applied to the primary multivariable models. The primary single-date yield model was a partial least squares regression (PLSR) using NDVI, RVI, NDRE, and CLRE acquired on 19 August [27]. This four-index specification was fixed for the primary analysis. The choice of 19 August was based on the preceding within-experiment temporal screening and the preference for the earlier of the two strongly associated dates. Accordingly, the reported outer-LOOCV performance is conditional on the fixed 19 August specification; acquisition-date selection itself was not nested within the outer validation loop.

PLSR was evaluated using nested leave-one-out cross-validation (LOOCV). In each of 45 outer folds, one subplot was withheld and the remaining 44 observations were used for model development. Within each outer training set, one to four latent components were compared by shuffled five-fold cross-validation with random seed 2026, and the component count with the lowest inner-CV RMSE was selected. Predictor scaling was estimated from the applicable training data only. The selected model was refitted to the full outer training set and predicted the held-out subplot, yielding one out-of-fold prediction per plot.

Predictive performance was calculated from outer-loop out-of-fold predictions using R^2, root mean square error (RMSE), mean absolute error (MAE), RMSE normalized by mean observed yield, and bias. Bias was defined as predicted minus observed; positive values indicate systematic overprediction. As a naïve benchmark, each held-out subplot was predicted using the mean yield of the corresponding outer training set, providing a fold-matched reference for evaluating the spectral models. Because all 45 observations came from one site and growing season, outer LOOCV was interpreted as within-experiment validation rather than external validation.

$$\text{Bias} = \frac{1}{n}\sum\nolimits_{i=1}^{n}(\hat{y}_i - y_i)$$

For comparison with PLSR, random forest (RF) hyperparameters were tuned within the same nested cross-validation framework and on the same outer held-out observations [28]. The inner grid comprised 300, 500, or 1000 trees; minimum leaf sizes of 1, 2, 3, 5, or 8; and mtry values of 1, 2, 3, or 4. Candidates were evaluated by shuffled five-fold cross-validation with seed 2026. The minimum inner-CV RMSE determined selection; when candidates were within 1.0 kg ha^{-1} of that minimum, fewer trees, a larger minimum leaf size, and then a smaller mtry were preferred. All fitting and tuning were restricted to the corresponding training data.

To characterize collinearity among the four spectral predictors, pairwise Pearson correlations and the condition number of the standardized predictor matrix were calculated. PLSR variable importance in projection (VIP), loadings, and standardized regression coefficients were obtained from a model fitted to the complete dataset for interpretation only. VIP values were interpreted as contributions within a correlated predictor set rather than as unique or causal effects. Predictive performance was calculated only from out-of-fold predictions.

To determine whether spectral information provided predictive value beyond the experimental treatment structure, Ridge regression models were evaluated by outer LOOCV. Cultivar and density were dummy-coded categorical predictors, with Jinchuan 1 and

D1 as reference categories. T0 contained Cultivar + Density, and T1 added Cultivar × Density through eight interaction terms formed from two non-reference cultivar and four non-reference density indicators. Block was not included in the predictive agronomic baselines because it represented experimental replication rather than an agronomic predictor intended for application beyond the experimental layout. The evaluated specifications were T0, T1, T0 + 4VI, T1 + 4VI, T0 + RE, and T1 + RE, where 4VI comprised NDVI, RVI, NDRE, and CLRE and RE comprised NDRE and CLRE. Within each outer training set, the Ridge regularization parameter α was selected independently from {0.01, 0.1, 1, 10, 100} by minimizing RMSE in shuffled five-fold inner cross-validation with random seed 2026. Continuous vegetation indices were median-imputed and standardized from training-fold estimates; dummy and interaction variables were not standardized, and the intercept was retained. These comparisons were interpreted as within-experiment incremental predictive information, not evidence of external transferability.

Because ordinary leave-one-plot-out validation does not evaluate performance under a withheld agronomic domain, three grouped schemes were used as structured transfer diagnostics within the present experiment [21,22]. Leave-one-density-out validation withheld all nine subplots at one density, trained on 36 plots, and repeated this procedure for five density groups. Leave-one-cultivar × density-combination-out validation withheld the three block replicates of one combination, trained on the remaining 42 plots, and repeated this procedure for all 15 combinations. Leave-one-cultivar-out validation withheld all 15 subplots of one cultivar, trained on 30 plots, and repeated this procedure for three cultivar groups. Each scheme used the four 19 August indices. Preprocessing was re-estimated within each training subset, and one to four PLS components were selected by shuffled five-fold inner cross-validation with seed 2026. These diagnostics remained internal to one site and season and did not substitute for independent validation.

Finally, uncertainty in cross-validated performance was quantified by cluster bootstrap resampling of fixed out-of-fold predictions. Fifteen Block × Density clusters were defined, each containing the three cultivar subplots. A total of 2000 replicates were generated with seed 2026, and percentile 95% confidence intervals were calculated for R^2, RMSE, MAE, and bias. The same cluster resamples were applied to paired model predictions when comparing PLSR with RF and when assessing the incremental contribution of spectral predictors relative to the interaction-inclusive Ridge baseline (T1). Paired bootstrap distributions were obtained for ΔR^2, ΔRMSE, and ΔMAE. These intervals characterize uncertainty in within-experiment cross-validation metrics and are not prediction intervals for independent future fields.

2.8. Green-Band Sensitivity Analysis

A green-band sensitivity analysis was conducted for 19 August because the green band was not represented in the primary four-index PLSR model. GNDVI was calculated for each pixel before ROI averaging:

$$\mathrm{GNDVI} = \frac{NIR - G}{NIR + G}$$

where G and NIR denote green and near-infrared band-level values, respectively. Plot-level mean GNDVI was then extracted for the same 45 ROIs, preserving one-to-one correspondence with the experimental plots.

Two nested PLSR specifications were compared: the primary model containing NDVI, RVI, NDRE, and CLRE, and the same model with GNDVI added. Both specifications used outer LOOCV, with one to four latent components evaluated by shuffled five-fold inner cross-validation using seed 2026. The component count minimizing the inner-CV

RMSE was selected, and predictor scaling was estimated from the corresponding training data only. This analysis assessed whether GNDVI provided additional within-experiment predictive information beyond the primary four-index specification.

A separate 19 August dataset containing green, red, red-edge, and near-infrared values was also examined as a sensitivity dataset. Because no traceable reflectance-panel calibration record was available for this export, these variables are referred to as band-level values. The dataset comprised 45 plot-matched observations with no missing values.

2.9. Software and Reproducibility

Single-index regression and the initial PLSR workflow were implemented in MATLAB R2019b (MathWorks, Natick, MA, USA). The nested PLSR, tuned RF, Ridge, multiple-testing, grouped-validation, and bootstrap procedures were implemented in Python 3.12.13. Predictive models used scikit-learn 1.9.0, with supporting statistical and data-processing libraries. Final figure preparation used R version 4.6.0 (R Foundation for Statistical Computing, Vienna, Austria).

Random seeds were fixed at 2026 for stochastic inner cross-validation and resampling procedures where applicable. Unless otherwise specified, predictive performance was calculated from out-of-fold predictions.

2.10. Use of Generative AI Tools

ChatGPT (GPT-5.6 Sol; OpenAI, San Francisco, CA, USA) was used during manuscript preparation solely for English translation and language polishing. It was not used for study design, data collection, data generation, statistical analysis, model calculation, figure generation, or scientific interpretation. All scientific content, results, and conclusions were independently produced, reviewed, and verified by the authors.

3. Results

3.1. Agronomic and Spectral Responses to Cultivar and Planting Density

Split-plot ANOVA showed that cultivar had significant effects on grain yield and all measured yield-component traits, including effective panicle number, grains per panicle, seed-setting rate, and thousand-grain weight (all $p < 0.001$; Table 2). Planting density showed no significant main effect on grain yield (F = 0.48, $p = 0.751$) or on any measured yield-component trait. The density effect on effective panicle number was marginal but did not reach statistical significance (F = 2.90, $p = 0.093$). Grains per panicle, seed-setting rate, and thousand-grain weight also showed no significant density main effects, with p-values of 0.186, 0.932, and 0.941, respectively. A significant cultivar × planting density interaction was detected for grain yield (F = 5.29, $p = 0.001$), whereas the interactions for the four yield-component traits were not significant. For grain yield, the corresponding partial η^2 values were 0.957 for cultivar, 0.193 for planting density, and 0.679 for the cultivar × density interaction, indicating a particularly large cultivar effect and a substantial cultivar × density interaction relative to the density main effect.

The grain-yield patterns in Figure 1A further illustrate the large cultivar effect and the cultivar-dependent response to planting density. Tiannongdao 307 generally maintained higher yield levels than Jingrun and Jinchuan 1, whereas Jinchuan 1 showed lower overall yield under most density treatments. Although the cultivar × planting density interaction for grain yield was significant, the within-cultivar density simple-effect tests did not reach $p < 0.05$. The observed density-related changes were therefore interpreted as cultivar-dependent response patterns within the tested density range, not as evidence for a statistically supported optimal density for each cultivar.

Table 2. Effects of cultivar, planting density, and their interaction on agronomic traits and 19 August vegetation indices in rice.

Indicator	Cultivar F (Partial η^2)	p	Planting Density F (Partial η^2)	p	Cultivar × Planting Density F (Partial η^2)	p
Grain yield	221.74 (0.957)	<0.001	0.48 (0.193)	0.751	5.29 (0.679)	0.0012
Effective panicle number (m^{-2})	31.78 (0.761)	<0.001	2.90 (0.592)	0.093	1.05 (0.296)	0.432
Grains per panicle	151.69 (0.938)	<0.001	2.01 (0.501)	0.186	0.28 (0.102)	0.964
Seed-setting rate	15.47 (0.607)	<0.001	0.20 (0.090)	0.932	0.61 (0.195)	0.761
Thousand-grain weight	21.01 (0.678)	<0.001	0.18 (0.084)	0.941	2.09 (0.455)	0.087
NDVI	2343.00 (0.996)	<0.001	0.68 (0.254)	0.625	4.64 (0.650)	0.0025
RVI	577.78 (0.983)	<0.001	0.66 (0.249)	0.636	1.37 (0.354)	0.267
NDRE	963.15 (0.990)	<0.001	0.62 (0.236)	0.661	1.79 (0.417)	0.139
OSAVI	240.01 (0.960)	<0.001	0.47 (0.191)	0.756	2.14 (0.461)	0.080
CLRE	560.75 (0.982)	<0.001	0.69 (0.257)	0.619	1.26 (0.335)	0.317

Note: Values in parentheses are partial η^2. Planting density was tested against the block × density mean square. Cultivar and cultivar × planting density were tested against the subplot residual error. NDVI, RVI, NDRE, OSAVI, and CLRE refer to UAV observations acquired on 19 August. Degrees of freedom were (2, 20) for cultivar, (4, 8) for planting density, and (8, 20) for the cultivar × planting-density interaction.

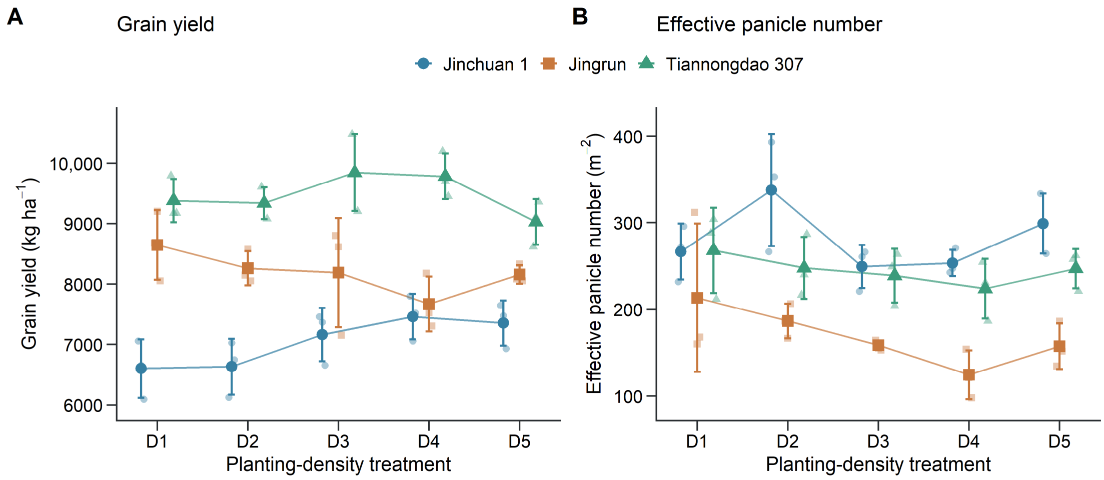

Figure 1. Grain yield and effective panicle number (m^{-2}) of three rice cultivars under five planting-density treatments. (**A**) Grain yield. (**B**) Effective panicle number (m^{-2}). Individual points represent subplot observations, and connected symbols with error bars represent treatment means ± SD across three replicates (n = 3). D1–D5 correspond to 23.81, 19.61, 16.67, 14.49, and 12.82 × 10^4 hills ha^{-1}, respectively. Cultivars are distinguished by color and symbol.

Effective panicle number (m^{-2}) also differed visibly among cultivars and density treatments (Figure 1B), but the split-plot analysis did not support a significant density main effect. Jinchuan 1 generally showed higher effective panicle numbers than Jingrun, whereas Tiannongdao 307 showed intermediate values. Because neither the density main effect nor the cultivar × density interaction was significant for effective panicle number, no density-based multiple-comparison letters were assigned.

Canopy spectral responses on 19 August also showed clear cultivar dependence. Cultivar had significant effects on all five vegetation indices measured on 19 August, including NDVI, RVI, NDRE, OSAVI, and CLRE (Table 2). In contrast, planting density showed no significant main effect on any vegetation index. The p-values for the density main effect were 0.625 for NDVI, 0.636 for RVI, 0.661 for NDRE, 0.756 for OSAVI, and 0.619 for CLRE. The cultivar × density interaction was significant only for NDVI

($F = 4.64$, $p = 0.003$), but not for RVI, NDRE, OSAVI, or CLRE. Simple-effect analysis showed that planting density significantly affected NDVI only in Jinchuan 1 ($F(4, 8) = 7.61$, Holm-adjusted $p = 0.023$), whereas the density effects were not significant in Tiannongdao 307 or Jingrun (Holm-adjusted $p > 0.05$). Within Jinchuan 1, NDVI was significantly higher under D4 and D5 than under D1 (Tukey-adjusted $p = 0.047$ and 0.005, respectively), whereas the remaining density comparisons were not significant. These results indicate that canopy spectral variation on 19 August was structured primarily by cultivar differences, whereas planting-density main effects on the vegetation indices were comparatively weak.

3.2. Temporal Dynamics and the Calendar-Defined Sensitive Observation Window

The relationships between grain yield and UAV-derived vegetation indices varied markedly across the 13 acquisition dates (Figure 2). All 15 date × vegetation-index combinations during the three June acquisitions showed negative pooled correlations. However, the direction of these associations was inconsistent among the three cultivars for all 15 combinations. After adjustment for cultivar, the absolute correlation coefficient decreased in 13 of 15 combinations and the sign reversed in 10; none of the cultivar-adjusted June associations remained significant after Benjamini–Hochberg or Holm correction. After further adjustment for cultivar, planting density, and block, the absolute correlation coefficient was lower than the pooled value in 11 of 15 combinations. Two full-structure-adjusted associations had unadjusted $p < 0.05$, but none remained significant after either multiple-testing correction. For example, the pooled NDVI–yield correlation on 16 June was $r = -0.464$, whereas the cultivar-specific correlations were -0.014, 0.234, and -0.549 for Tiannongdao 307, Jingrun, and Jinchuan 1, respectively; the cultivar-adjusted and full-structure-adjusted correlations were -0.125 and -0.231. This pattern shows that cultivar differences contributed strongly to the pooled early-season negative correlations. Full structural adjustment produced mixed residual correlations. The early negative associations do not support a common inverse relationship between early canopy development and final yield (Table S1; Figures S1 and S2).

The strongest pooled associations were concentrated between 19 and 25 August. On 19 August, NDRE and CLRE were strongly correlated with grain yield ($r = 0.866$ and 0.863, respectively), and similarly strong associations were observed on 25 August ($r = 0.875$ and 0.871, respectively). NDVI and RVI also showed relatively strong positive associations during this interval, whereas OSAVI generally exhibited weaker relationships.

Across the complete family of 65 date × vegetation-index tests, 46 associations were significant at the unadjusted $p < 0.05$ level, 41 remained significant after Benjamini–Hochberg false-discovery-rate correction, and 27 remained significant after Holm correction. Importantly, the NDRE and CLRE associations on both 19 and 25 August remained significant under both correction procedures (Table S1).

Structural adjustment also reduced the magnitude of the mid-August associations. Across the ten date × vegetation-index combinations on 19 and 25 August, cultivar-adjusted correlations remained positive but ranged from $r = 0.266$ to 0.489, whereas correlations adjusted for cultivar, planting density, and block ranged from $r = 0.040$ to 0.392. None of the full-structure-adjusted associations remained significant after Benjamini–Hochberg or Holm correction. The strong pooled correlations during 19–25 August were therefore partly related to cultivar and treatment structure. We next used multivariable models to test whether the four spectral indices still added predictive value beyond the agronomic treatment terms (Table S1; Figure S1).

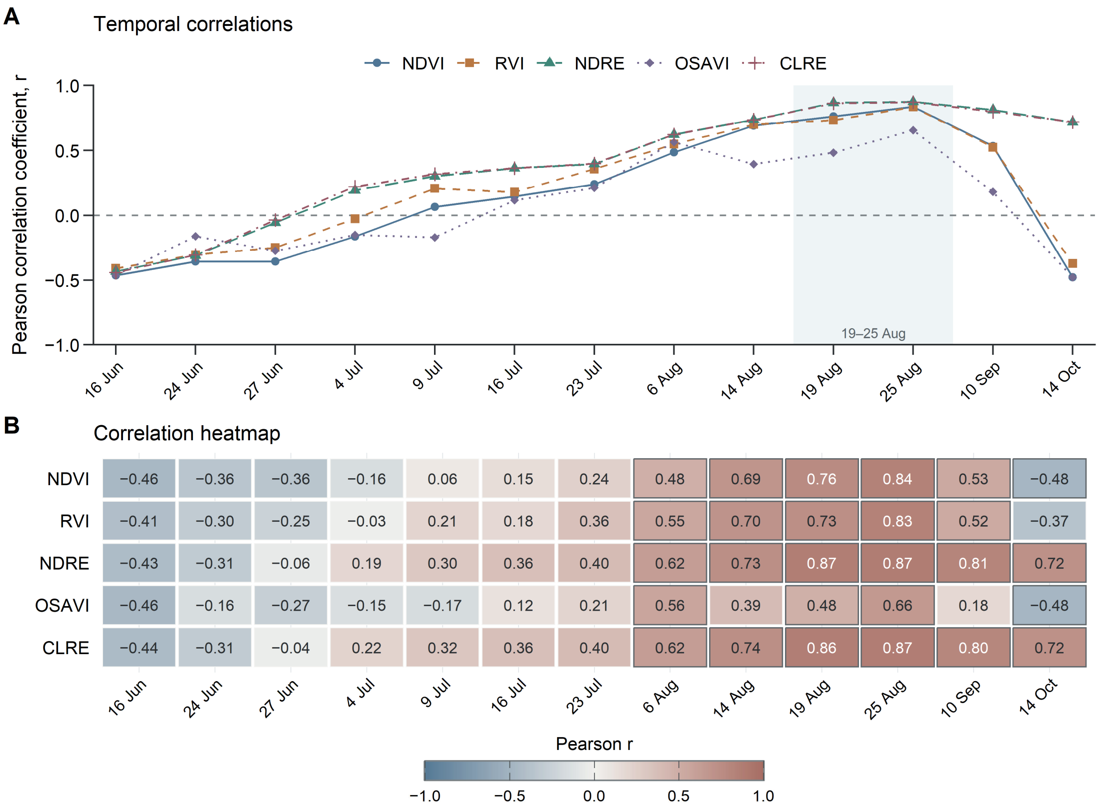

Figure 2. Temporal variation in the relationships between UAV-derived vegetation indices and grain yield across the rice growing season. (**A**) Pearson correlation coefficients (r) between grain yield and NDVI, RVI, NDRE, OSAVI, and CLRE across 13 UAV acquisition dates. The shaded region denotes the calendar-defined sensitive observation window from 19 to 25 August. (**B**) Heatmap of the corresponding Pearson correlation coefficients; values within cells indicate r. Blue and red tones represent negative and positive correlations, respectively. Thin dark outlines identify associations that remained significant after Holm adjustment across the complete family of 65 date × vegetation-index tests.

We therefore identified 19–25 August as the calendar-defined sensitive observation window in this experiment. Field phenological observations indicated that this period broadly coincided with reproductive development, although developmental timing differed among cultivars. Of the two dates, 19 August was retained as the primary single-date modeling point because it provided strong spectral–yield associations while allowing earlier yield assessment than 25 August.

3.3. Within-Experiment Yield Estimation, Model Comparison, and Sensitivity Analyses

Among the five single-index models evaluated for 19 August, NDRE provided the strongest cross-validated performance (R^2 = 0.728, RMSE = 580.54 kg ha^{-1}), followed closely by CLRE (R^2 = 0.720, RMSE = 588.58 kg ha^{-1}). NDVI, RVI, and particularly OSAVI showed lower predictive performance (Figure 3A).

Combining spectral predictors through PLSR improved within-experiment prediction. The four-index model based on NDVI, RVI, NDRE, and CLRE achieved an outer-LOOCV R^2 of 0.773, with RMSE = 530.28 kg ha^{-1}, MAE = 406.11 kg ha^{-1}, and bias = 2.81 kg ha^{-1}. The RMSE corresponded to 6.44% of the mean observed yield and approximately 0.47 of the plot-level yield standard deviation (Figure 3B).

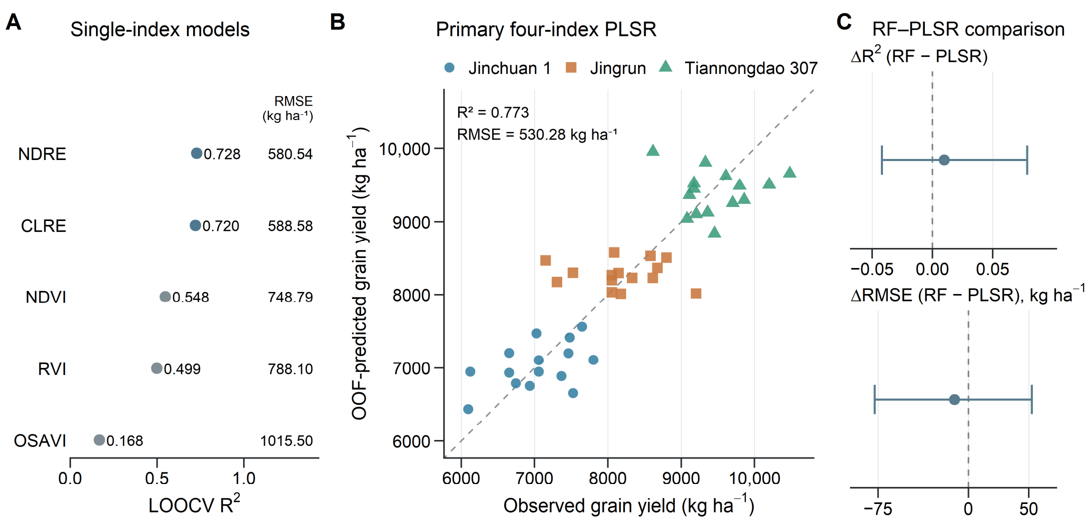

Figure 3. Within-experiment grain-yield prediction performance of UAV-based models. (**A**) Leave-one-out cross-validation (LOOCV) performance of five single-index univariate linear-regression models using vegetation indices acquired on 19 August. Points indicate LOOCV R^2, and the aligned values report RMSE (kg ha^{-1}). Point colors are used only for visual differentiation among the five single-index models. (**B**) Observed versus outer-LOOCV-predicted grain yield for the primary four-index PLSR model using NDVI, RVI, NDRE, and CLRE. All 45 out-of-fold predictions are shown; cultivars are distinguished by color and symbol, and the dashed line denotes the 1:1 relationship. (**C**) Paired differences between tuned random forest (RF) and PLSR (RF − PLSR) for R^2 and RMSE. Points indicate paired differences and horizontal error bars denote 95% confidence intervals obtained by Block × Density cluster bootstrap (B = 2000); vertical dashed lines indicate zero difference.

After nested hyperparameter tuning, RF achieved an outer-LOOCV R^2 of 0.783, with RMSE = 518.60 kg ha^{-1}, MAE = 422.15 kg ha^{-1}, and bias = 25.86 kg ha^{-1}. The cluster-bootstrap 95% confidence intervals were [0.628, 0.867] for PLSR R^2 and [409.38, 633.90] kg ha^{-1} for PLSR RMSE, compared with [0.670, 0.858] for RF R^2 and kg ha^{-1} for RF RMSE. Although RF had a slightly lower point RMSE, the paired bootstrap confidence intervals for the RF–PLSR differences included zero for both ΔR^2 and ΔRMSE. The two models were therefore regarded as showing broadly comparable within-experiment predictive performance (Figure 3C). Both substantially reduced prediction error relative to the training-mean benchmark (RMSE = 1138.44 kg ha^{-1}).

Strong collinearity was present among the four PLSR predictors. NDVI and RVI were highly correlated (r = 0.9929), as were NDRE and CLRE (r = 0.9981), and the standardized predictor matrix had a condition number of 95.29. VIP values were correspondingly similar (CLRE = 1.036, NDRE = 1.034, RVI = 0.974, and NDVI = 0.953), indicating that predictor contributions were largely shared within a highly correlated spectral signal rather than uniquely attributable to a single vegetation index (Table S2).

The green-band sensitivity analysis did not materially change model performance. Adding pixel-wise GNDVI to the four-index PLSR model increased R^2 by only 0.0018 and reduced RMSE by 2.06 kg ha^{-1}. A model based directly on the four available band-level values likewise did not improve performance (R^2 = 0.768, RMSE = 535.79 kg ha^{-1}). These sensitivity results did not materially alter the performance of the primary four-index specification, which was therefore retained for subsequent analyses (Table S3).

3.4. Agronomic Structure and Incremental Spectral Information

The treatment-only Ridge baseline containing cultivar and planting density (T0) achieved R^2 = 0.729 and RMSE = 579.21 kg ha^{-1}. Adding the cultivar × planting density interaction (T1) resulted in R^2 = 0.730 and RMSE = 577.96 kg ha^{-1}.

On 19 August, pooled spectral–yield associations varied among cultivars and were markedly attenuated after adjustment for cultivar, planting density, and block (Table S4).

When the four vegetation indices acquired on 19 August were added to the interaction-inclusive T1 baseline, R^2 increased to 0.801 and RMSE decreased to 496.33 kg ha^{-1}. Relative to T1, this corresponded to ΔR^2 = +0.071 (95% cluster-bootstrap CI: +0.0003 to +0.1279) and an RMSE reduction of 81.62 kg ha^{-1} (95% CI: +0.42 to +148.46 kg ha^{-1}). MAE was also reduced by 91.67 kg ha^{-1} (95% CI: 7.73 to 168.37 kg ha^{-1}). Thus, the joint four-index spectral set retained incremental predictive information after cultivar, planting density, and their interaction were represented in the model.

Adding only the two red-edge indices (NDRE and CLRE) to T1 yielded R^2 = 0.785 and RMSE = 515.65 kg ha^{-1}. The corresponding point estimates indicated ΔR^2 = +0.055 and an RMSE reduction of 62.31 kg ha^{-1}; however, the paired cluster-bootstrap intervals crossed zero for both ΔR^2 (95% CI: −0.0328 to +0.1233) and RMSE reduction (95% CI: −33.49 to +145.78 kg ha^{-1}). The red-edge-only augmentation therefore showed greater uncertainty than the joint four-index augmentation (Figure 4A; Table S5). These comparisons describe incremental predictive information within the present experiment and should not be interpreted as evidence of equivalent gains under external transfer (Table S5).

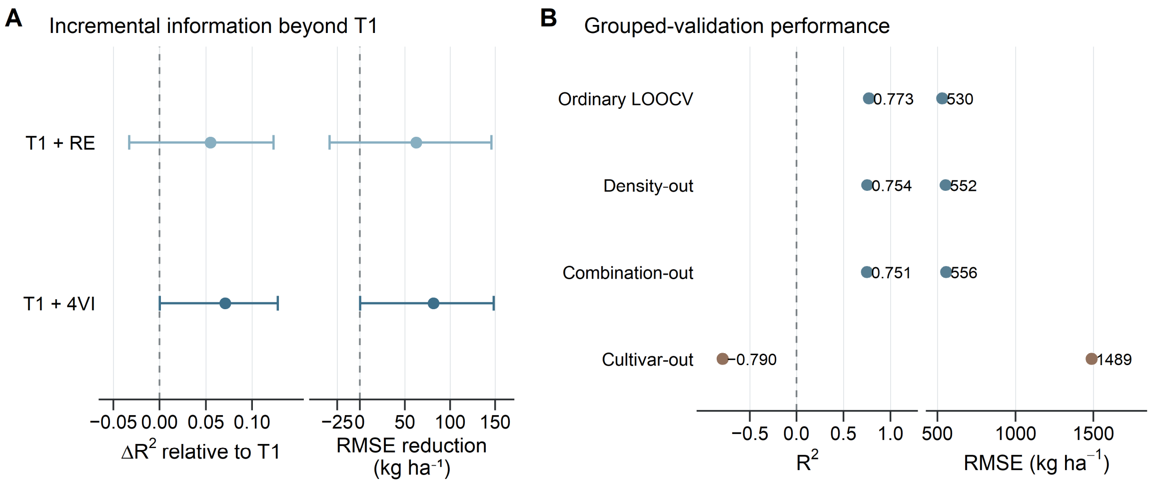

Figure 4. Incremental spectral information and grouped-validation performance. (**A**) Changes in predictive performance after adding spectral predictors to the interaction-inclusive agronomic baseline T1 (cultivar + planting density + cultivar × planting density). RE comprises NDRE and CLRE, whereas 4VI comprises NDVI, RVI, NDRE, and CLRE. Points show ΔR^2 and RMSE reduction relative to T1, and horizontal error bars denote paired Block × Density cluster-bootstrap 95% confidence intervals (B = 2000). Light- and dark-blue points denote T1 + RE and T1 + 4VI, respectively. Vertical dashed lines indicate zero change; positive RMSE reduction indicates lower prediction error than T1. (**B**) R^2 and RMSE of the primary four-index PLSR model under ordinary LOOCV and three grouped-validation diagnostics: leave-one-density-out, leave-one-cultivar × density-combination-out, and leave-one-cultivar-out validation. Blue points denote ordinary LOOCV, density-out, and combination-out validation, whereas the brown point denotes cultivar-out validation. The dashed vertical line in the R^2 panel denotes R^2 = 0. All grouped-validation analyses were conducted within the same site and growing season and represent internal structured diagnostics rather than independent external validation.

3.5. Grouped Validation and Model Applicability

Grouped validation produced markedly different results depending on the type of withheld agronomic domain (Figure 4B). Leave-one-density-out validation retained performance close to ordinary LOOCV (R^2 = 0.754, RMSE = 551.60 kg ha^{-1}). Similarly, leave-one-cultivar × density-combination-out validation yielded R^2 = 0.751 and RMSE = 555.68 kg ha^{-1}.

In contrast, predictive performance deteriorated strongly when an entire cultivar was withheld. Leave-one-cultivar-out validation yielded $R^2 = -0.790$ and RMSE = 1489.10 kg ha^{-1}, indicating poor performance under the leave-one-cultivar-out diagnostic within this dataset (Table S6).

The three validation schemes showed a clear contrast. Performance changed little when density levels or cultivar × density combinations were withheld, but it declined sharply when an entire cultivar was excluded. Because all grouped analyses were conducted within the same site and growing season, they define an internal applicability boundary rather than demonstrating external generalization.

4. Discussion

4.1. Temporal Dynamics of Spectral–Yield Relationships and the Sensitive Observation Window

The relationships between UAV-derived vegetation indices and grain yield changed substantially across the growing season. Among the 13 acquisition dates, the strongest associations were observed during 19–25 August, when NDRE and CLRE showed consistently high correlations with yield. On 19 August, the correlation coefficients for NDRE and CLRE were 0.866 and 0.863, respectively, and increased slightly to 0.875 and 0.871 on 25 August. Similar temporal dependence has been reported in previous rice UAV studies, although the most informative stage has not been identical among experiments [9,10,16,17]. Booting, initial heading, and heading have each been identified as favorable periods for yield-related spectral assessment under different cultivar, management, and sensing conditions [9,10,16]. These differences suggest that a single optimal calendar date is unlikely to apply across experiments. A more useful goal is to identify a period when canopy spectral variation is closely related to yield under the conditions being studied.

The strong performance of NDRE and CLRE during this period is consistent with the known sensitivity of red-edge information to changes in crop canopy condition. Previous studies in rice have shown that red-edge indices can respond to variation in canopy nitrogen and chlorophyll-related traits and may retain useful sensitivity as canopy development increases [25,26,29–31]. However, these relationships should not be interpreted as evidence that the yield associations observed here were driven by chlorophyll or nitrogen status alone. UAV multispectral canopy signals integrate variation in leaf area, canopy architecture, panicle emergence, and within-plot background exposure, all of which change during crop development. Multi-temporal rice studies have similarly shown that the relative contributions of leaves, panicles, and background components to yield-related spectral signals shift from booting through heading and early grain filling [16]. The early-season negative correlations should therefore be interpreted cautiously. Across all 15 June date × vegetation-index combinations, the directions of the within-cultivar associations were inconsistent, and adjustment for cultivar reduced the absolute correlation in 13 combinations and reversed the sign in 10. Additional adjustment for cultivar, planting density, and block produced a more mixed residual pattern: most associations remained weaker in absolute magnitude than their pooled counterparts, although several red-edge associations became more negative. None of the cultivar-adjusted or full-structure-adjusted June associations remained significant after Benjamini–Hochberg or Holm correction. These results show that cultivar differences contributed strongly to the early pooled negative correlations. However, full structural adjustment did not remove all of these relationships. The early negative associations do not support a common inverse relationship between early canopy development and final yield.

Field phenological observations indicated that the mid-August acquisitions broadly coincided with the transition into reproductive development, although developmental timing differed somewhat among cultivars. Accordingly, the 19–25 August interval is best interpreted as a calendar-defined sensitive observation window for this experiment rather than as a universally optimal physiological stage. This distinction is important because the timing of a strong pooled spectral–yield relationship may depend not only on crop development itself, but also on the agronomic structure represented by different cultivars and planting-density treatments.

4.2. Agronomic Structure Shapes Apparent Spectral–Yield Associations

Strong pooled spectral–yield relationships were not equally expressed across agronomic backgrounds. On 19 August, NDRE and CLRE showed high pooled correlations with grain yield, but cultivar-specific analyses revealed clear heterogeneity, and both relationships were markedly attenuated after adjustment for cultivar, planting density, and block. Across the ten date $\times$ vegetation-index combinations on 19 and 25 August, cultivar-adjusted correlations ranged from 0.266 to 0.489, whereas correlations adjusted for cultivar, planting density, and block ranged from only 0.040 to 0.392. For example, the pooled correlations of NDRE and CLRE with yield on 19 August decreased from 0.866 and 0.863 to full-structure-adjusted correlations of 0.154 and 0.112, respectively. These results show that cultivar and management differences explained part of the strong pooled correlation. The relationship was not equally strong across all plots. High pooled correlations therefore do not show that the spectral–yield relationship is consistent across cultivars and management conditions.

This structural dependence is agronomically plausible. Rice cultivars differ in developmental duration, canopy architecture, and population structure, all of which can affect the integrated canopy signal observed by UAV sensors. Previous multi-cultivar UAV studies have likewise shown that cultivar background can alter yield-estimation performance, and recent work involving a large set of rice genotypes has further demonstrated genotype dependence in remote-sensing-based yield forecasting [6,7]. Planting density represents an additional source of management structure. Although density did not exert a significant overall main effect on grain yield in the present experiment, previous field studies have shown that density can modify light distribution and nitrogen distribution within rice canopies, providing a pathway through which management may alter remotely sensed canopy expression [3]. The weaker correlations after adjustment show that part of the pooled spectral signal was associated with cultivar and management structure. The spectral response was not constant across treatments.

However, the spectral predictors did more than reproduce cultivar or treatment labels. Using the interaction-inclusive Ridge model containing cultivar, planting density, and their interaction as the agronomic baseline, addition of the four vegetation indices acquired on 19 August increased cross-validated R^2 from 0.730 to 0.801 and reduced RMSE from 577.96 to 496.33 kg ha^{-1}. Relative to the agronomic baseline, the corresponding paired Block $\times$ Density cluster-bootstrap 95% confidence intervals excluded zero for both ΔR^2 (+0.0003 to +0.1279) and RMSE reduction (+0.42 to +148.46 kg ha^{-1}), although the lower bounds were close to zero. The four-index spectral set therefore added predictive value beyond the categorical treatment terms in this experiment. This should be interpreted as conditional predictive information rather than evidence for an independent physiological effect of any specific canopy trait.

The remaining predictive information should also not be attributed to the red-edge indices alone. Adding only NDRE and CLRE to the interaction-inclusive baseline produced favorable point estimates, but the paired bootstrap 95% confidence intervals for both ΔR^2 and RMSE reduction crossed zero, indicating greater uncertainty for the red-edge-only augmentation. In addition, NDRE and CLRE were highly correlated, as were NDVI and RVI, while the four PLSR predictors had similar VIP values. The model therefore appears to have used shared information across a correlated multispectral feature set rather than an independent contribution that could be assigned to a single index. Agronomic adjustment did not remove the value of the UAV data. Instead, it changed how the spectral information should be interpreted. The pooled correlations contained substantial cultivar and management structure, but the combined spectral predictors still added predictive value beyond the treatment terms. Whether this information remains useful when the agronomic domain itself changes requires a different validation framework and is examined in the following section.

4.3. Grouped Validation Reveals Agronomic-Domain-Specific Transferability

The ordinary LOOCV results indicated good predictive performance within the present experiment, but this validation scheme mainly evaluates prediction for another plot drawn from the same experimental population. Because each held-out plot shares cultivar backgrounds and most treatment structure with the training data, ordinary LOOCV does not directly test performance under a shifted agronomic domain. Previous methodological studies have emphasized that validation design should reflect the intended prediction target when data contain hierarchical or grouped structure, as randomly mixed training and validation samples may not adequately represent out-of-domain performance [21,22]. We therefore used grouped validation as an internal diagnostic to examine how model performance changed when specific types of agronomic information were systematically excluded from model training.

Performance degradation differed markedly among the three grouped schemes. Leave-one-density-out and leave-one-cultivar × density-combination-out validation produced only modest reductions relative to ordinary LOOCV, whereas leave-one-cultivar-out validation resulted in a pronounced loss of accuracy. This contrast reflects how much relevant agronomic information remained represented in the training set. When one density level was withheld, all three cultivars were still represented at other densities; similarly, when one cultivar × density combination was omitted, both the cultivar and density remained represented elsewhere in the training data. By contrast, leave-one-cultivar-out validation removed an entire genetic background. The results therefore do not simply indicate that the model was transferable across density but not across cultivar. Rather, performance loss depended on the extent to which the withheld domain remained represented by related agronomic backgrounds during model calibration.

The stronger cultivar shift is consistent with the structural patterns observed elsewhere in this study. Cultivar had a dominant effect on grain yield, the cultivar × density interaction was also significant, and cultivar-specific and structure-adjusted analyses showed that pooled spectral–yield relationships differed among cultivar backgrounds. Previous multi-cultivar rice UAV studies have likewise reported that differences in developmental duration, canopy architecture, and yield potential can alter the relationship between remotely sensed canopy features and grain yield [6,15]. More recent work involving a broader range of rice genotypes has also demonstrated genotype dependence in remote-sensing-based yield forecasting [7]. The poor leave-one-cultivar-out result suggests that the spectral–yield relationship differed among genetic backgrounds. It does not indicate a failure of UAV spectral sensing itself. Because the grouped diagnostics were conducted only for the predefined PLSR model, these results do not establish whether alternative nonlinear algorithms could fully resolve the cultivar shift.

This result highlights a practical limit within this dataset. Within this dataset, performance was less degraded when withheld groups retained cultivar representation in the training data, whereas prediction deteriorated substantially for a cultivar absent from model calibration. Application to previously unseen cultivars may therefore require recalibration, more diverse training material, or modeling strategies that explicitly account for genotype heterogeneity. Similar limitations in cross-cultivar, cross-year, and cross-sensor transfer have been reported in previous rice UAV studies [15], and recent breeding-oriented studies have increasingly adopted group-based validation to assess model robustness under varietal heterogeneity [11]. Nevertheless, all grouped analyses in the present study were conducted within a single site and growing season and should therefore be regarded as structured within-experiment transfer diagnostics rather than external validation. Model transfer depended on how similar the training and target agronomic populations were.

4.4. Implications for UAV-Assisted Rice Yield Assessment and Future Research

UAV multispectral sensing can support more than stand-alone yield prediction. It can also provide rapid, repeatable, and non-destructive canopy information within a defined cultivar and management context. In this study, yield-related spectral information was concentrated within a specific observation window. The structural analyses also showed that part of the spectral variation was related to cultivar and treatment structure, while the spectral indices still added predictive value beyond these treatment terms. UAV observations may therefore be most useful as a complement to conventional agronomic measurements in cultivar comparison, planting-density experiments, and breeding trials, rather than as signals interpreted independently of crop background. The 19–25 August interval may provide a useful reference under similar experimental conditions, but it should not be treated as a fixed calendar recommendation or as a universally synchronized physiological stage.

This study has several limitations. All 45 subplots were obtained from a single site and growing season, and the three japonica cultivars were selected to introduce contrasting genetic backgrounds rather than to represent the wider diversity of japonica germplasm. Cluster bootstrap and grouped validation helped quantify uncertainty and structured shifts within this experiment. However, these analyses cannot replace independent validation across years, sites, or broader cultivar sets. The temporal analysis was based on calendar-date UAV observations supported by field phenological observations, rather than formal stage-based normalization. Consequently, direct comparison among cultivars at precisely synchronized developmental stages was beyond the scope of the present study. In addition, the reported cross-validation metrics characterize the fixed 19 August specification identified from the preceding within-experiment temporal screening, rather than the performance of a fully data-driven pipeline in which acquisition date is selected within each outer validation fold. The present analysis also relied mainly on plot-level vegetation indices. Although red-edge indices were strongly associated with yield, these data cannot separate the respective contributions of chlorophyll status, nitrogen condition, canopy architecture, panicle development, and other physiological processes. We therefore interpret these spectral results as predictive associations, not as evidence of specific physiological mechanisms.

Future work should focus on two related questions: whether improved developmental alignment can increase consistency across rice materials, and whether broader training domains can reduce cultivar-related performance loss. Multi-year and multi-location experiments should include a wider range of cultivars and management conditions and incorporate standardized developmental alignment using growth stage, thermal time, or other reproducible metrics [19,20]. Recent large-genotype rice studies have adopted accumulated growing degree days to normalize remote-sensing time series across environments while still identifying substantial genotype dependence in yield prediction, indicating that phenological alignment and genetic diversity need to be considered jointly [7]. More stringent model evaluation should also reserve entire years, locations, or independent cultivar populations for testing rather than relying only on randomly mixed or within-experiment grouped validation [12,14]. Recent multi-year rice UAV studies have already used an independent growing season for validation, providing a stronger test of performance under practical use [14].

5. Conclusions

This study examined UAV multispectral–yield relationships in a structured rice experiment involving three cultivars and five planting-density treatments, with particular attention to observation timing, agronomic structure, and model transferability. Spectral–yield associations varied markedly across the season and were strongest within the calendar-defined 19–25 August observation window. The four-index PLSR model based on NDVI, RVI, NDRE, and CLRE on 19 August achieved an outer-LOOCV R^2 of 0.773 and an RMSE of 530.28 kg ha^{-1}, while the tuned RF model showed broadly comparable within-experiment performance. These results support the use of mid-season UAV multispectral observations for characterizing yield-related canopy variation under the conditions examined here, but do not imply that the same calendar window will be optimal across cultivars, years, or environments.

The strong pooled spectral–yield relationships were partly structured by cultivar and management differences. After cultivar, planting density, and their interaction were represented in the agronomic baseline, the joint four-index spectral set still provided incremental predictive information: cross-validated R^2 increased from 0.730 to 0.801, while RMSE decreased from 577.96 to 496.33 kg ha^{-1}. Paired cluster-bootstrap confidence intervals for both ΔR^2 and RMSE reduction excluded zero, although their lower bounds were close to zero. However, grouped validation showed that this predictive information was not equally transferable across agronomic domains. Performance was only modestly degraded when density levels or cultivar $\times$ density combinations were withheld, whereas leave-one-cultivar-out validation performed poorly ($R^2 = -0.790$). Thus, the spectral–yield relationship differed among cultivar backgrounds.

UAV multispectral sensing provided useful yield-related information in this experiment, but its value depended on agronomic structure and validation context. Future studies should include multiple years and locations, more cultivars, standardized developmental alignment, and independent validation before UAV-based yield models are used more widely.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy16181855/s1, Table S1: Complete pooled, cultivar-specific, and structure-adjusted temporal associations between UAV-derived vegetation indices and grain yield across 13 acquisition dates; Figure S1: Temporal structure of pooled, cultivar-specific, and structure-adjusted spectral–yield associations across the growing season; Figure S2: Illustrative early-season cultivar structure for the strongest negative pooled June association; Table S2: PLSR predictor collinearity and diagnostics; Table S3: Green-band and GNDVI sensitivity analysis; Table S4: Cultivar-specific and structure-adjusted associations between UAV-derived vegetation indices and grain yield on 19 August; Table S5: Agronomic baseline and incremental spectral models; Table S6: Fold-level performance of the four-index PLSR model under grouped-validation diagnostics.

Author Contributions: Conceptualization, J.Y., J.L. and F.Z.; methodology, J.Y. and J.L.; software, J.Y.; validation, J.Y. and J.L.; formal analysis, J.Y.; investigation, J.Y., X.Z., Y.D., M.L., X.L. and X.W.; resources, X.L., X.W. and F.Z.; data curation, J.Y.; writing—original draft preparation, J.Y.; writing—review and editing, J.L. and F.Z.; visualization, J.Y.; supervision, J.L. and F.Z.; project administration, X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript. J.Y. and J.L. contributed equally to this work.

Funding: This research was funded by the National Key Research and Development Program of China, under the subproject "Creation and Application of New Lodging-Resistant Northern Japonica Rice Varieties", grant number 2024YFD1201005-03. Additional support for experimental consumables was provided by the Tianjin Science and Technology Bureau, contract number 24ZYCGSN00300.

Institutional Review Board Statement: Not applicable. This study did not involve humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request. The raw UAV imagery is not publicly available due to file-size limitations and ongoing related research.

Acknowledgments: The authors thank the Quality Rice Breeding Team of Tianjin Agricultural University for providing experimental materials and field support. During the preparation of this manuscript, the authors used ChatGPT (GPT-5.6 Sol; OpenAI, San Francisco, CA, USA) solely for English translation and language polishing. All scientific analyses, results, interpretations, and conclusions were independently conducted, reviewed, and verified by the authors, who take full responsibility for the final content of the manuscript.

Conflicts of Interest: Xiuru Wu is employed by China United Property Insurance Co., Ltd., Tianjin Branch. The company had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The remaining authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Yun, Y. Changes in the growth and yield of an extremely early-maturing rice variety according to transplanting density. *Agriculture* **2023**, *13*, 717. [CrossRef]
2. Zhou, W.; Yan, L.; Fu, Z.; Guo, H.; Zhang, W.; Liu, W.; Ye, Y.; Long, P. Increasing planting density and reducing N application improves yield and grain filling at two sowing dates in double-cropping rice systems. *Plants* **2023**, *12*, 2298. [CrossRef] [PubMed]
3. Liu, Z.C.; Shang, L.Y.; Dai, S.J.; Ye, J.Y.; Sheng, T.; Deng, J.; Liu, K.; Fahad, S.; Tian, X.H.; Zhang, Y.B.; et al. Optimizing nitrogen application and planting density improves yield and resource use efficiency via regulating canopy light and nitrogen distribution in rice. *J. Integr. Agric.* **2026**, *25*, 81–91. [CrossRef]
4. Mi, K.L.; Lu, Y.Y.; Xu, F.F.; Yang, Y.J.; Zhang, H.P.; Zhang, H.C. Planting density affects yield of indica-japonica hybrid rice by regulating population and photosynthetic characteristics. *Crop J.* **2026**, *14*, 592–605. [CrossRef]
5. Zhang, H.; Mi, K.; Ma, M.; Chen, T.; Zhang, H.; Yang, Y. Effects of planting density on population characteristics and yield formation in conventional japonica rice with different yield levels. *Agron. J.* **2025**, *117*, e70233. [CrossRef]
6. Duan, B.; Fang, S.; Gong, Y.; Peng, Y.; Wu, X.; Zhu, R. Remote estimation of grain yield based on UAV data in different rice cultivars under contrasting climatic zone. *Field Crops Res.* **2021**, *267*, 108148. [CrossRef]
7. Della-Bellver, F.; Franch, B.; Tarin-Mestre, J.; Guerrero-Benavent, C.J.; Domingo, C.; Català Forner, M.; Marti-Jerez, K.; Marqués, L. Assessing rice genotype dependency in remote sensing: Challenges in nitrogen discrimination and yield forecasting. *Precis. Agric.* **2026**, *27*, 133. [CrossRef]
8. Aasen, H.; Honkavaara, E.; Lucieer, A.; Zarco-Tejada, P.J. Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows. *Remote Sens.* **2018**, *10*, 1091. [CrossRef]
9. Zhou, X.; Zheng, H.B.; Xu, X.Q.; He, J.Y.; Ge, X.K.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.X.; Tian, Y.C. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 246–255. [CrossRef]
10. Wan, L.; Cen, H.; Zhu, J.; Zhang, J.; Zhu, Y.; Sun, D.; Du, X.; Zhai, L.; Weng, H.; Li, Y.; et al. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer–a case study of small farmlands in the South of China. *Agric. For. Meteorol.* **2020**, *291*, 108096. [CrossRef]
11. Zhou, H.K.; Huang, F.D.; Lou, W.D.; Gu, Q.; Ye, Z.R.; Hu, H.; Zhang, X.B. Yield prediction through UAV-based multispectral imaging and deep learning in rice breeding trials. *Agric. Syst.* **2025**, *223*, 104214. [CrossRef]

12. Huo, J.; Wang, Q.; Ji, C.; Huang, S.; Sun, B.; Xiao, H.; Yang, C.; Yan, F.; Zhang, Y.; Zhou, W. Rice yield prediction using UAV-based multispectral imagery and AutoGluon across regions and field scales. *Front. Plant Sci.* **2026**, *17*, 1866530. [CrossRef] [PubMed]
13. Tian, X.; Cen, F.; Wei, J.; Zhao, Q.; Gao, Z. Rice multi-variety yield prediction using multi-temporal fusion of thermal infrared and multispectral UAV data based on random forest feature selection. *Ecol. Inform.* **2026**, *97*, 103894. [CrossRef]
14. Pokharel, R.; Gentimis, T.; Kongchum, M.; Tubana, B.; Adhikari, R.; Setiyono, T. Rice yield estimation based on machine learning applied to UAV remote sensing data. *Remote Sens.* **2026**, *18*, 1575. [CrossRef]
15. Zheng, H.; Ji, W.; Wang, W.; Lu, J.; Li, D.; Guo, C.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y.; et al. Transferability of models for predicting rice grain yield from unmanned aerial vehicle (UAV) multispectral imagery across years, cultivars and sensors. *Drones* **2022**, *6*, 423. [CrossRef]
16. Su, X.; Wang, J.; Ding, L.; Lu, J.; Zhang, J.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.; Tian, Y. Grain yield prediction using multi-temporal UAV-based multispectral vegetation indices and endmember abundance in rice. *Field Crops Res.* **2023**, *299*, 108992. [CrossRef]
17. Li, J.; Wang, W.; Sheng, Y.; Anwar, S.; Su, X.; Nian, Y.; Yue, H.; Ma, Q.; Liu, J.; Li, X. Rice yield estimation based on cumulative time series vegetation indices of UAV MS and RGB images. *Agronomy* **2024**, *14*, 2956. [CrossRef]
18. Luo, S.; Jiang, X.; Jiao, W.; Yang, K.; Li, Y.; Fang, S. Remotely sensed prediction of rice yield at different growth durations using UAV multispectral imagery. *Agriculture* **2022**, *12*, 1447. [CrossRef]
19. Yang, Q.; Shi, L.S.; Han, J.Y.; Chen, Z.W.; Yu, J. A VI-based phenology adaptation approach for rice crop monitoring using UAV multispectral images. *Field Crops Res.* **2022**, *277*, 108419. [CrossRef]
20. Yang, Q.; Shi, L.; Han, J.; Yu, J.; Huang, K. A near real-time deep learning approach for detecting rice phenology based on UAV images. *Agric. For. Meteorol.* **2020**, *287*, 107938. [CrossRef]
21. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
22. Habibi, L.N.; Matsui, T.; Tanaka, T.S.T. Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing. *J. Agric. Food Res.* **2024**, *16*, 101096. [CrossRef]
23. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
24. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [CrossRef]
25. Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* **2005**, *32*, L08403. [CrossRef]
26. Liang, T.; Duan, B.; Luo, X.; Ma, Y.; Yuan, Z.; Zhu, R.; Peng, Y.; Gong, Y.; Fang, S.; Wu, X. Identification of high nitrogen use efficiency phenotype in rice (*Oryza sativa* L.) through entire growth duration by unmanned aerial vehicle multispectral imagery. *Front. Plant Sci.* **2021**, *12*, 740414. [CrossRef] [PubMed]
27. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
29. Zheng, H.; Ma, J.; Zhou, M.; Li, D.; Yao, X.; Cao, W.; Zhu, Y.; Cheng, T. Enhancing the nitrogen signals of rice canopies across critical growth stages through the integration of textural and spectral information from unmanned aerial vehicle (UAV) multispectral imagery. *Remote Sens.* **2020**, *12*, 957. [CrossRef]
30. Wang, Z.; Tan, X.; Ma, Y.; Liu, T.; He, L.; Yang, F.; Shu, C.; Li, L.; Fu, H.; Li, B.; et al. Combining canopy spectral reflectance and RGB images to estimate leaf chlorophyll content and grain yield in rice. *Comput. Electron. Agric.* **2024**, *221*, 108975. [CrossRef]
31. Li, S.; Ding, X.; Kuang, Q.; Ata-Ul-Karim, S.T.; Cheng, T.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. Potential of UAV-based active sensing for monitoring rice leaf nitrogen status. *Front. Plant Sci.* **2018**, *9*, 1834. [CrossRef] [PubMed]