



# Remote Estimation of Rice Yield With Unmanned Aerial Vehicle (UAV) Data and Spectral Mixture Analysis

Bo Duan<sup>1</sup>, Shenghui Fang<sup>1,2\*</sup>, Renshan Zhu<sup>2,3</sup>, Xianting Wu<sup>2,3</sup>, Shanqin Wang<sup>4</sup>, Yan Gong<sup>1,2</sup> and Yi Peng<sup>1,2</sup>

<sup>1</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, <sup>2</sup> Lab for Remote Sensing of Crop Phenotyping, Wuhan University, Wuhan, China, <sup>3</sup> College of Life Sciences, Wuhan University, Wuhan, China, <sup>4</sup> College of Resource and Environment, Huazhong Agricultural University, Wuhan, China

## OPEN ACCESS

### Edited by:

Yanbo Huang,  
United States Department  
of Agriculture, United States

### Reviewed by:

Jingcheng Zhang,  
Hangzhou Dianzi University, China  
Keisuke Nagai,  
Nagoya University, Japan

### \*Correspondence:

Shenghui Fang  
shfang@whu.edu.cn

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
Frontiers in Plant Science

**Received:** 28 September 2018

**Accepted:** 07 February 2019

**Published:** 27 February 2019

### Citation:

Duan B, Fang S, Zhu R, Wu X,  
Wang S, Gong Y and Peng Y (2019)  
Remote Estimation of Rice Yield With  
Unmanned Aerial Vehicle (UAV) Data  
and Spectral Mixture Analysis.  
*Front. Plant Sci.* 10:204.  
doi: 10.3389/fpls.2019.00204

The accurate assessment of rice yield is crucially important for China's food security and sustainable development. Remote sensing (RS), as an emerging technology, is expected to be useful for rice yield estimation especially at regional scales. With the development of unmanned aerial vehicles (UAVs), a novel approach for RS has been provided, and it is possible to acquire high spatio-temporal resolution imagery on a regional scale. Previous reports have shown that the predictive ability of vegetation index (VI) decreased under the influence of panicle emergence during the later stages of rice growth. In this study, a new approach which integrated UAV-based VI and abundance information obtained from spectral mixture analysis (SMA) was established to improve the estimation accuracy of rice yield at heading stage. The six-band image of all studied rice plots was collected by a camera system mounted on an UAV at booting stage and heading stage respectively. And the corresponding ground measured data was also acquired at the same time. The relationship of several widely-used VIs and Rice Yield was tested at these two stages and a relatively weaker correlation between VI and yield was found at heading stage. In order to improve the estimation accuracy of rice yield at heading stage, the plot-level abundance of panicle, leaf and soil, indicating the fraction of different components within the plot, was derived from SMA on the six-band image and *in situ* endmember spectra collected for different components. The results showed that VI incorporated with abundance information exhibited a better predictive ability for yield than VI alone. And the product of VI and the difference of leaf abundance and panicle abundance was the most accurate index to reliably estimate yield for rice under different nitrogen treatments at heading stage with the coefficient of determination reaching 0.6 and estimation error below 10%.

**Keywords:** rice, yield, remote sensing (RS), unmanned aerial vehicle (UAV), vegetation index (VI), spectral mixture analysis (SMA)

## INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most important grain crops in the world, especially in China. There are more than half of the China's population regarding rice as the staple food (Xiao et al., 2002). The accurate estimation of rice yield is of significance to ensure food security and promote sustainable development.

Remote sensing (RS) is a technique which can obtain the information about an object without making physical contact with the object (Wikipedia<sup>1</sup>) and it can efficiently obtain canopy spectra data in a non-destructive way, which carries valuable information indicating the canopy interaction with solar radiation such as vegetation absorption and scattering (Thenkabail et al., 2011). The vegetation canopy spectra are closely related to vegetation growth. In the visible range, vegetation has strongly absorption of light due to the pigments and presents as low reflectance (Woolley, 1971). But in the near-infrared (NIR) range, vegetation reflectance is relatively high and affected by thick plant tissues and canopy structure (Gausman et al., 1969). A series of studies have been developed to relate the vegetation spectra to vegetation growth parameters such as chlorophyll content (Gitelson et al., 2006; Wu et al., 2008; Feret et al., 2011), leaf area index (LAI) (Broge and Leblanc, 2001; Viña et al., 2011) and biomass (Thenkabail et al., 2000; Hansen and Schjoerring, 2003), and thus a lot of vegetation indices (VIs) calculated from reflectance of different spectra ranges (Hatfield et al., 2008) have been proposed to accurately estimate these parameters. For example, Peng et al. (2013) used Enhanced Vegetation Index (EVI) and Wide Dynamic Range Vegetation Index (WDRVI) obtained from MODIS to accurately estimate the gross primary productivity in crops with coefficients of variation below 20% in maize and 25% in soybean (Peng et al., 2013). Li et al. (2014) applied several red-edge based spectral indices to estimate plant nitrogen uptake with the coefficient of determination above 0.76. Parametric statistical approaches based on VIs are by far the simplest and most studied variable estimation approaches and have been widely used in monitoring crop growth (Verrelst et al., 2015). And the changes of crop growth status, which can be effectively monitored by spectral measures, directly determine its ultimate yield. Therefore, VIs have also exhibited good potential in remote estimation of crop yield especially at the large scales (Rahman et al., 2012; Sun et al., 2017). Becker-Reshef et al. (2010) estimated the wheat yield in Kansas and Ukraine with a 7% error by using time series Normalized Difference Vegetation Index (NDVI) data from MODIS. Holzman et al. (2014) came up with a new method to estimate regional crop yield using the Temperature Vegetation Dryness Index (TVDI) with the RMSE values ranged from 12 to 13% for soybean and 14 to 22% for wheat. Sakamoto et al. (2014) mapped U.S. corn yields successfully using Wide Dynamic Range Vegetation Index (WDRVI) derived from timeseries MODIS data with the estimation error below 30% at the state level. Generally, VI-based methods are the mainstream approach for crop yield prediction, and a lot of regression algorithms of using VIs to estimate crop yield were established including simple

linear functions and complex non-linear functions (Tucker et al., 1980). Many experiments showed that the use of appropriate VI was the key to crop yield estimation instead of the complex function structures in VI-based methods (Gholizadeh et al., 2015). Therefore, it is particularly important for crop yield evaluation to get the exact VIs.

However, there may be a considerable discrepancy between pixel sizes of RS images (e.g., 1 km for MODIS satellite image) and much smaller sizes of croplands (e.g., usually 33 m × 20 m in South China) due to the limitation of the spatial resolution as well as the landscape fragmentation (Gong et al., 2018). In this case, one pixel of a RS image may contain several land cover types and the signal of this pixel (mixed pixel) is the outcome of various land cover components that have significantly different spectra. VI, derived from the spectra of such mixed pixels, may encompass the unexpected information of the components not related to yield, which could lower the precision of yield estimation. The mixed pixel is always a problem that have to be discussed in the application of RS technique. Even for the high spatial resolution images, the problem of mixed pixel still exists because of the smaller cropland components such as flower, fruit and grain. This problem is more obvious when addressing rice yield prediction with RS data in VI-based method. Rice is a type of grain crop, the panicle will gradually emerge in paddy rice field when rice steps into reproductive phase and may last until maturation. In the early stage of rice panicle emergence, the panicle was bright green and scarcely distributed in rice canopy, while rice leaves were still the dominant component in paddy rice field. With the continuous growth of rice, the amount of panicle increased and its color gradually changed to yellow. Since rice panicle had the dramatically different spectra from rice leaves, the remotely detected canopy spectra of rice after panicle emergence were greatly mixed by panicle and leaf spectra, and the accuracy of estimated crop growth parameters in VI-based method would decrease. Zhou et al. (2017) found that the appearance of the rice panicle enhances the difficulty of LAI estimation and yield prediction in the later growth stages of rice. And Sakamoto et al. (2011) found the same result in rice that the color index performed well-before the booting stage but poorly in the late growth stages. Harrell et al. (2011) reported that the reduced predictive ability at heading stage was most likely associated with the uneven emergence of panicles into the sensor field of view. These studies showed that the canopy spectra remotely obtained from rice in late growth stages were susceptible to the interference of uneven emergence of panicles, and the factor of spectral mixture that influence the rice yield estimation must be considered.

Spectral mixture analysis (SMA) was extensively used to quantify the spectral contributions from different components in a mixed pixel. It assumed that the spectrum of a mixed pixel was a linear or non-linear combination of its constituent spectral components (called endmember) weighted by their subpixel fractional cover (called abundance) (Somers et al., 2011). Once the pure spectra of endmembers were obtained, the fraction of each endmember within a mixed pixel can be estimated based on its mixed spectra (Bioucas-Dias et al., 2012). SMA has been widely applied in RS for evaluating vegetation properties. Gitelson et al.

<sup>1</sup>[https://en.wikipedia.org/wiki/Remote\\_sensing](https://en.wikipedia.org/wiki/Remote_sensing)

(2002) developed an approach to estimate vegetation fraction in sampling zones based on measured spectra of two endmembers (bare soil and dense vegetation). Lobell and Asner presented a SMA approach that employs time series of MODIS data to estimate subpixel fractions of land cover types with  $R^2$  reaching 0.8 (Lobell and Asner, 2004). Therefore, SMA can be a good tool to analyze the influence of spectral mixture to rice yield estimation and it also has good potential to improve the accuracy of rice yield estimation by combining with VIs.

In recent years, unmanned aerial vehicles (UAVs) have become an increasingly used platform for RS application in Precision agriculture due to its high spatial and temporal resolutions (Zarcotejada et al., 2013; Candiago et al., 2015; Fang et al., 2016; Aasen and Bolten, 2018). And the UAV-collected data is playing a more and more important role in monitoring crop growth. Based on multi-band images obtained by an UAV system, this study explores to improve VI-based approach for rice yield estimation by combining with SMA.

## MATERIALS AND METHODS

### Study Area

The study site was located at the Rice Experiment and Research Base of Huazhong Agricultural University near Wuxue City, Hubei Province, China (30.1117°N, 115.5892°E). In this investigation, the data from 24 rice plots was studied – **Figure 1A**. They were of the size about 20 m<sup>2</sup> including around 330 plants and all planted with the same hybrid of rice. The field management for these plots were similar except that different levels of nitrogen fertilizer were applied. Eight levels of nitrogen fertilizer (0, 3, 5.5, 8.5, 11, 14, 16.5, and 19.5 kg/ha) were utilized, and each level was repeated on three randomly distributed plots – **Figure 1B**. The growing period for rice in our study was from June to September, and field experiments were conducted during booting (13 August, 2015) and heading (29 August, 2015) stages of rice growth. At these two stages, rice has gradually completed the transformation from vegetative period to reproductive period. There was no obvious panicle in rice field at booting stage; on the contrary, the panicle began to emerge at heading stage. During the course of each experiment, one UAV flight was arranged to obtain the image of all rice plots. After the UAV flight (from 11:00 am to 2:00 pm), the corresponding ground measurements were carried out *in situ* immediately.

### Field Data Collection

The LAI and leaf chlorophyll content were measured at booting (13 August, 2015) and heading (29 August, 2015) stage of rice growth. Ground rice canopy LAI was measured using a SunScan canopy analysis system (Delta-T Devices, Ltd., Burwell, Cambridge, United Kingdom) under windless conditions and stable light levels. For each plot, three LAI readings were acquired and their average represented the canopy LAI of the plot. A SPAD-502 Chlorophyll Meter (Soil and Plant Analyzer Development Chlorophyll Meter, Spectrum Technologies, Inc., Plainfield, IL, United States; abbreviated as SPAD) was utilized to measure the leaf chlorophyll content of rice. Five positions

were selected in each plot, and the rice leaf chlorophyll content was measured in each position. For each position in plot, three SPAD values of top layer leaf, middle layer leaf, and bottom layer leaf were recorded and the average SPAD indicated the rice leaf chlorophyll content of the position. And the plot level leaf chlorophyll content was the average of five SPAD values in five positions.

At the heading stage of rice, some samples of typical ground features were collected and their spectra were measured as the endmember spectra used for SMA. The spectra of six kinds of endmember were measured using an ASD Field Spec 4 spectrometer (Analytical Spectral Devices Inc., Boulder, CO, United States), including top layer leaf (TL), bottom layer leaf (BL), top layer panicle (TP), bottom layer panicle (BP), dry soil (DS), and wet soil (WS). Samples of different endmember were collected from the studied area and their spectra were measured *in situ* immediately, and the measurements were carried out in a stable light condition after UAV flight. For soil spectra collection, we walked into the rice field and selected several representative sample-collecting spots in the ridges between plots. The measurement of soil spectra was conducted using ASD probe pointing to the ground vertically at the appropriate height to make sure the field of view was full of wet or dry soil with no other land cover features, and the averaged spectra were used as soil spectra. In the same way, the panicle spectra were obtained with ASD probe pointing downward approximately 10 cm above the panicle samples. Since the rice panicle was small and granular, the samples of panicle were put on a black background and evenly spread. As for the leaf spectra, a self-illuminated leaf clip of ASD was used and then the spectra of top layer leaf and bottom layer leaf were taken respectively. In the whole study area, five positions were selected randomly and the spectra of top layer leaf, bottom layer leaf, top layer panicle, and bottom layer panicle were gained respectively for each plot. Similarly, the averaged spectra were used as their endmember spectra.

At maturity, the all rice plants in each plot were harvested manually for determination of grain yield. The seeds were cleaned and exposed to the sun until their weight did not change. And then all the dry seeds were weighted according to plots and the plot-level yield was obtained.

### Surface Reflectance and Vegetation Index Derived From UAV Data

The image of study plots was acquired using a Mini-MCA system mounted on a UAV (S1000, SZ DJI Technology, Co., Ltd., Shenzhen, China) on 13 August and 29 August, 2015 – **Figure 2**. The Mini-MCA system consists of an array of six individual miniature digital cameras (Mini-MCA 6, Tetracam, Inc., Chatsworth, CA, United States) – **Figure 2C**. Each camera imager was equipped with a customer-specified band pass filter centered at a wavelength of 490, 550, 670, 720, 800, or 900 nm, respectively, and the band width was 10 nm. These bands were selected since they were commonly employed to analyze vegetation growth-related parameters (Behrens et al., 2006; Ray et al., 2010; Kira et al., 2015).

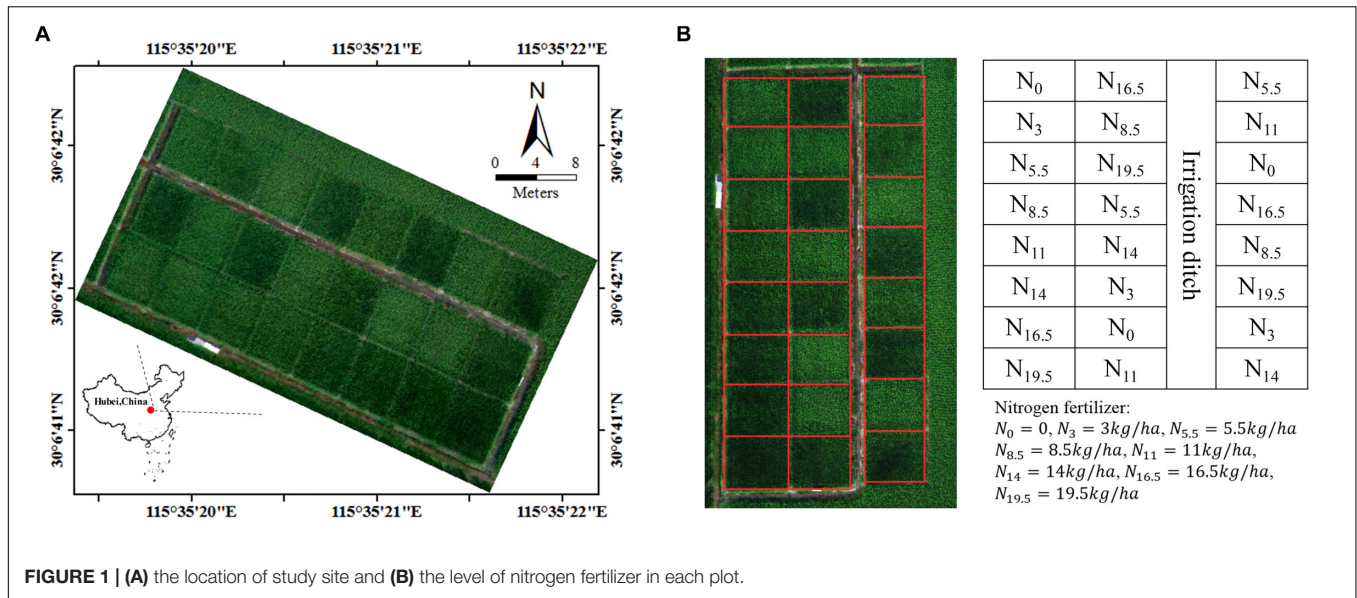


FIGURE 1 | (A) the location of study site and (B) the level of nitrogen fertilizer in each plot.

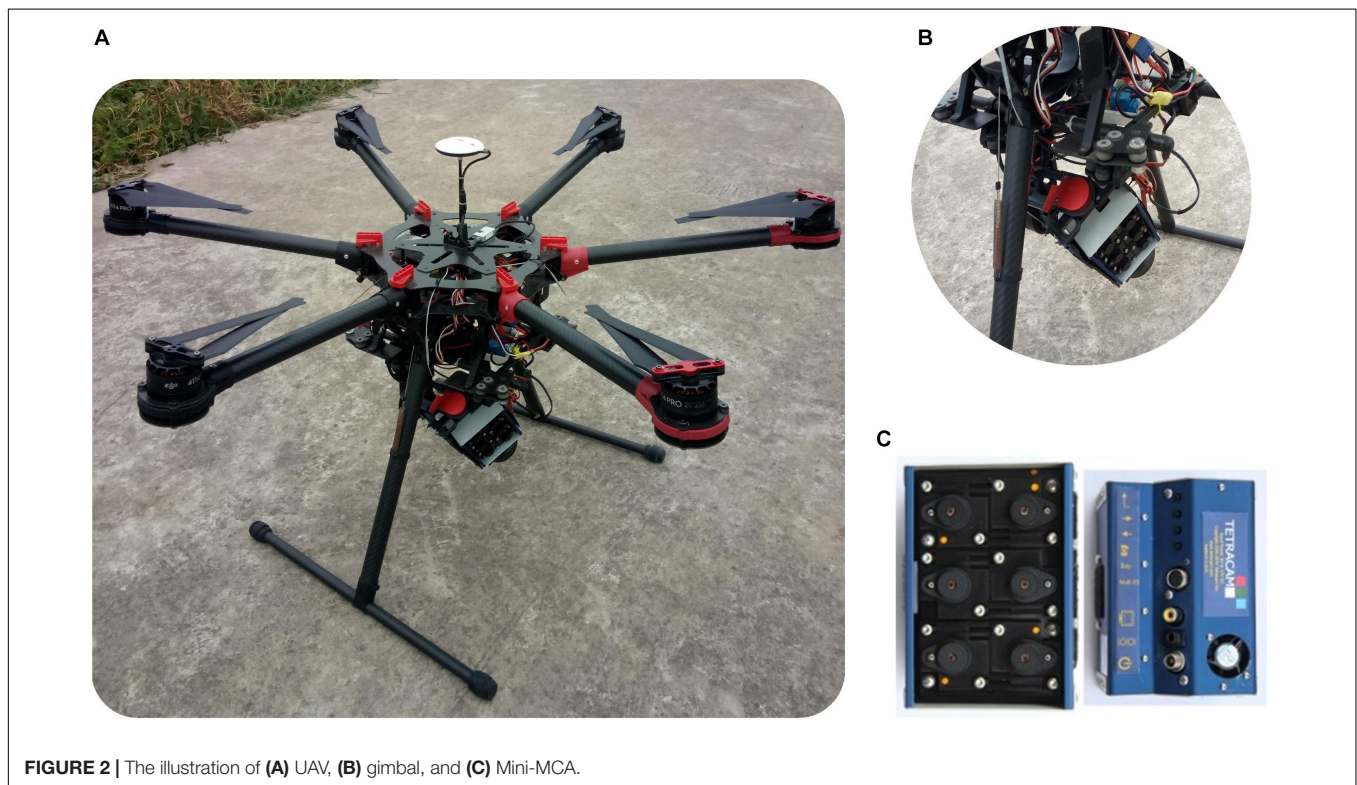


FIGURE 2 | The illustration of (A) UAV, (B) gimbal, and (C) Mini-MCA.

The Mini-MCA system was attached to the UAV on a gimbal which can help to compensate for the UAV movement (pitch and roll) during the flight and guarantee close to nadir image collection – **Figure 2B** (Turner et al., 2014). Since the imaging system had a significant camera mis-registration effect, six cameras were co-registered in the laboratory prior to the flight so that corresponding pixels of each lens were spatially overlapping in the same focal plane (Jhan et al., 2016). UAV flight was conducted under clear skies with little cloud cover between

10:00 and 14:00 local time when the changes in the solar zenith angle were minimal. The image was collected at approximately 60 m above the ground with the spatial resolution of 3 cm around.

In this study, an empirical linear correction method was applied to transform image Digital Numbers (DNs) into surface Reflectance ( $\rho$ ) (Dwyer et al., 1995; Laliberte et al., 2011). Four calibration ground targets were placed in the cameras' field of view as a standard for image radiometric corrections and the image taken from Mini-MCA system included all of them. The

calibration targets used in this study have the relatively constant reflectance of 0.06, 0.24, 0.48, and 1 respectively throughout the visible to NIR wavelengths and are specially utilized for aerial image radiometric calibration. As a linear relationship was assumed between DNs and  $\rho$ , the canopy surface reflectance can be calculated as

$$\rho_\lambda = DN_\lambda \times Gain_\lambda + Offset_\lambda (\lambda = 490, 550, 670, 720, 800 \text{ and } 900 \text{ nm}) \quad (1)$$

Where  $\rho_\lambda$  and  $DN_\lambda$  are the surface reflectance and digital number, respectively, of a given pixel at wavelength  $\lambda$ ;  $Gain_\lambda$ ; and  $Offset_\lambda$  are gains and bias of the camera at different wavelengths respectively. For each wavelength,  $Gain_\lambda$  and  $Offset_\lambda$  can be calculated using the least-square method from  $\rho$  and  $DN$  values (referring to  $DN_{0.06}$ ,  $DN_{0.24}$ ,  $DN_{0.48}$ , and  $DN_1$ ) of four calibration targets.

$$\begin{pmatrix} 0.06 \\ 0.24 \\ 0.48 \\ 1 \end{pmatrix} = \begin{pmatrix} DN_{0.06} \\ DN_{0.24} \\ DN_{0.48} \\ DN_1 \end{pmatrix} \times Gain_\lambda + Offset_\lambda \quad (2)$$

$$\begin{pmatrix} Offset_\lambda \\ Gain_\lambda \end{pmatrix} = \left[ \begin{pmatrix} 1 & DN_{0.06} \\ 1 & DN_{0.24} \\ 1 & DN_{0.48} \\ 1 & DN_1 \end{pmatrix}^T \begin{pmatrix} 1 & DN_{0.06} \\ 1 & DN_{0.24} \\ 1 & DN_{0.48} \\ 1 & DN_1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & DN_{0.06} \\ 1 & DN_{0.24} \\ 1 & DN_{0.48} \\ 1 & DN_1 \end{pmatrix}^T \begin{pmatrix} 0.06 \\ 0.24 \\ 0.48 \\ 1 \end{pmatrix} \quad (3)$$

For each of 24 plots, we defined a maximum rectangle in the image that fit the plot – **Figure 1B**. The rectangle included approximated 18000 pixels, and the plot-level VI was retrieved by averaging all of the per-pixel values within the rectangle. The VI tested in this study are shown in **Table 1**.

## Fully Constrained Least Squares Linear Spectral Mixture Analysis

Although UAV made it possible to get high resolution data, one pixel on a UAV image still encompassed several land cover components especially for crops with distinct grains like rice. When rice was at heading stage, there were six dominant components visible in the field of view – **Figure 3**, including top layer leaf (TL), bottom layer leaf (BL), top layer panicle (TP), bottom layer panicle (BP), dry soil (DS), and wet soil (WS). And their continuous spectra were collected by an ASD spectrometer used as the endmember spectra for SMA – **Figure 4**. Compared with the ground measured reflectance spectra, the spectra acquired by the MCA camera onboard UAV were discrete wavebands (490, 550, 670, 720, 800, and 900 nm, 10 nm band width). For each endmember, with reference to the wavelength range of each MCA waveband, the average of ground measured endmember reflectance in the corresponding wavelength range was calculated as the endmember reflectance used for the SMA of UAV images. Therefore, six kinds of endmember reflectance were obtained as:  $\rho(TL)$ ,  $\rho(BL)$ ,  $\rho(TP)$ ,  $\rho(BP)$ ,  $\rho(DS)$ , and  $\rho(WS)$ .

In this study, fully constrained least squares linear SMA method was employed to estimate the abundance fractions of components present in an image pixel. For linear spectral mixing model, a mixed pixel was defined as a linear combination of components with their relative concentrations (Heinz, 2001). In this case, the reflectance  $\rho$  of a mixed pixel at wavelength  $\lambda$  can be approximated as

$$\rho_\lambda = \sum_{i=1}^N Abd_i \rho_\lambda(i) + e \quad (4)$$

where  $e$  is noise or can be interpreted as a measurement error,  $N$  is the number of selected endmembers,  $Abd_i$  denotes the abundance fraction of endmember  $i$ , and  $\rho_\lambda(i)$  is the reference reflectance of endmember  $i$  at wavelength  $\lambda$ .

And for the fully constrained linear mixing model, two partially constraints were considered which were referred to as non-negatively constraint and sum-to-one constraint

$$0 \leq Abd_i \leq 1; \text{ and } \sum_{i=1}^N Abd_i = 1 \quad (5)$$

**TABLE 1** | Vegetation Indices tested in this study.

Vegetation Indices	Formula	Reference
Simple Ratio (SR)	$\rho_{800} / \rho_{670}$	Jordan, 1969
Red-edge Chlorophyll Index ( $CI_{rededge}$ )	$\rho_{800} / \rho_{720} - 1$	Gitelson et al., 2005
Green-edge Chlorophyll Index ( $CI_{green}$ )	$\rho_{800} / \rho_{550} - 1$	Gitelson et al., 2005
Normalized Difference Vegetation Index (NDVI)	$(\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670})$	Rouse et al., 1974
Green Normalized Difference Vegetation Index (GNDVI)	$(\rho_{800} - \rho_{550}) / (\rho_{800} + \rho_{550})$	Gitelson et al., 1996
Normalized Difference Red edge (NDRE)	$(\rho_{800} - \rho_{720}) / (\rho_{800} + \rho_{720})$	Glenn et al., 2010
Visible Atmospherically Resistant Index (VARI)	$(\rho_{550} - \rho_{670}) / (\rho_{550} + \rho_{670})$	Gitelson et al., 2002
MERIS Terrestrial Chlorophyll Index (MTCI)	$(\rho_{800} - \rho_{720}) / (\rho_{720} - \rho_{670})$	Dash and Curran, 2004
Enhanced Vegetation Index (EVI)	$2.5(\rho_{800} - \rho_{670}) / (\rho_{800} + 6\rho_{670} - 7.5\rho_{490} + 1)$	Liu and Huete, 1995
Two-band Enhanced Vegetation Index (EVI2)	$2.5(\rho_{800} - \rho_{670}) / (\rho_{800} + 2.4\rho_{670} + 1)$	Jiang et al., 2008

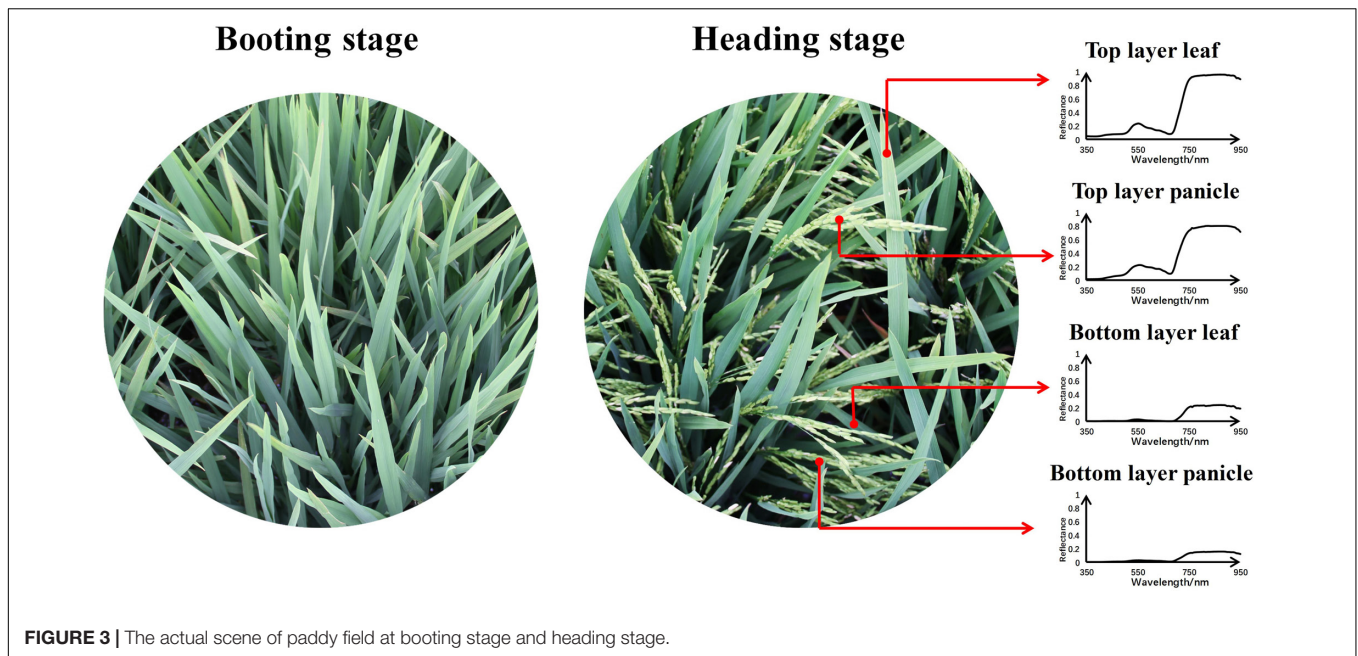


FIGURE 3 | The actual scene of paddy field at booting stage and heading stage.

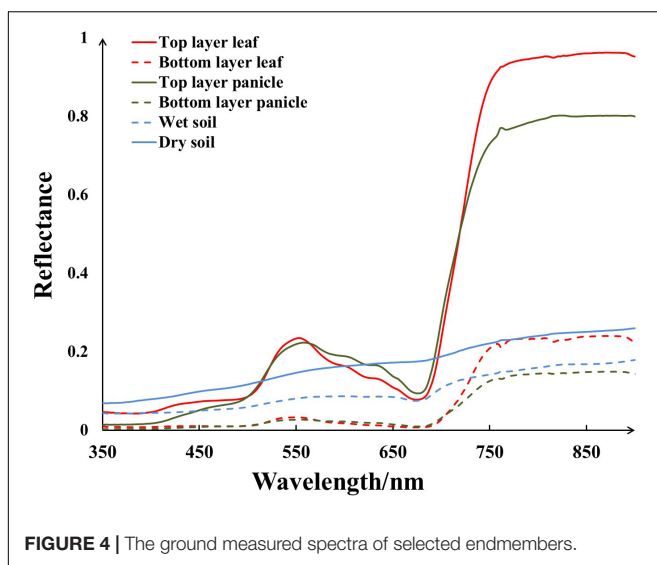


FIGURE 4 | The ground measured spectra of selected endmembers.

As shown above, the value of abundance was non-negative with its maximum constrained to 1 and for each mixed pixel the sum of the endmember abundance equals to 1. Generally, the abundance that meets these conditions can indicate the proportion of the corresponding endmember within the mixed pixel. An abundance of 0 means no presence of the particular endmember in this pixel, while an abundance of 1 indicates that this pixel is a pure pixel of the particular endmember.

To solve the abundance of these selected six components, a least squares method was applied (Pu et al., 2014). In this study, we used MATLAB (MATLAB 2016a, MathWorks, Inc., Natick, MA, United States) to derive the abundance gray images of six selected endmembers whose gray level values meant the

estimated abundance of the corresponding endmember in the image pixel. The plot-level abundance was calculated in the same manner as plot-level reflectance, and for each plot the same maximum rectangle was utilized again. In consideration of endmember selection, the panicle abundance ( $Abd_p$ ) and leaf abundance ( $Abd_l$ ) were the sum of the corresponding elements in top layer and bottom layer together.

### Data Analysis Among UAV Data, Ground Measured Data, and Rice Yield

In this study, the IBM SPSS Statistics (Statistical Product and Service Solutions 22.0, IBM, Armonk, NY, United States) was used to statistically describe and analyze data. Firstly, a normally distribution test was applied to the rice yield data and the product data of LAI and SPAD value ( $LAI \times SPAD$ ) at booting and heading stage. The result of Shapiro–Wilk test served as the indicator to test whether the data is normally distributed or not. And then, the correlation analysis and regression analysis were successively applied. The Pearson correlation coefficient ( $r$ ) was exhibited as the result of correlation analysis. And for regression analysis, adjusted R square ( $R^2$ ), root mean square error (RMSE) and  $p$ -value were analyzed and compared. The detailed calculation method of above evaluation indices can be found in the help document of SPSS.

The relationship between ground measured data ( $LAI \times SPAD$ ) and rice yield was firstly established at booting and heading stage respectively. And the plot-level VI derived from UAV data was correlated with rice yield directly and compared with the ground measured data. The difference of these two rice growth stages was discussed and analyzed.

At the heading stage of rice, in consideration of the impact of spectral mixture within 1 pixel, the leaf and panicle make different spectral contributions to the reflectance of the mixed

pixel. For each studied plot, different proportion of leaf and panicle may affect the accuracy of yield estimation based on plot-level VI at heading stage. In order to eliminate the influence, plot-level VI was incorporated with abundance information to relate with rice yield and four relationships were developed using 23 samples (one of the rice yield data was obviously wrong): (1) yield vs. VI, (2) yield vs. VI  $\times$  Abd<sub>L</sub>, (3) yield vs. VI  $\times$  Abd<sub>P</sub> and (4) yield vs. VI  $\times$  Abd<sub>L-P</sub>. The different estimation ability for rice yield of these four kinds of index was compared and evaluated.

## Algorithm Establishment Using Leave One Out Cross-Validation

The final rice yield estimation model was established using a leave one out cross-validation method. Leave one out cross-validation is a statistics method widely applied in model establishment and validation (Fielding and Bell, 1997). It divided the samples into two groups, one for training and the other one is used for validation. For leave one out cross-validation, the validation set just included one sample and the training and validating process was repeated K times (K equals to the number of samples,  $K = 23$  in this study). For each time  $i$ ,  $K-1$  samples were used iteratively as training data to calibrate the coefficients ( $Coef_i$ ) of the algorithm with the accuracy was measured in terms of coefficients of determination ( $R_i^2$ ), and the remaining single sample was used for validation to obtain the estimation error ( $E_i$ ). The training and validating process was repeated K times until every single sample was used exactly one time for validation data. After K iterations, the coefficients and accuracy of the final algorithm can be produced as

$$Coef = \frac{\sum_{i=1}^K Coef_i}{K} \quad R^2 = \frac{\sum_{i=1}^K R_i^2}{K} \quad RMSE = \sqrt{\frac{\sum_{i=1}^K E_i^2}{K}}$$

## RESULTS

### Relationship Between Ground Measured Data and Rice Yield

In general, the product of LAI and leaf-level SPAD (LAI  $\times$  SPAD) was used to estimate canopy chlorophyll status of rice and has been proven to be a promising index to predict rice yield (Liu et al., 2017). In this study, LAI  $\times$  SPAD was calculated according to the ground measured LAI and SPAD value both at booting and heading stage. Among all plots

which have yield data, one of the plot-level LAI records was missing and 22 LAI  $\times$  SPAD values was calculated at these two stages. Before correlation analysis and regression analysis, a normal distribution test was applied to the rice yield data and LAI  $\times$  SPAD data. The result of Shapiro–Wilk test turned out that the yield and LAI  $\times$  SPAD data followed normal distribution – **Table 2**. And a simple linear regression analysis of LAI  $\times$  SPAD and yield was presented in **Figure 5A**. At booting stage, the result provided a satisfactory linear fitting equation between yield and LAI  $\times$  SPAD with a relatively high  $R^2$  value ( $R^2 = 0.627^{**}$ ). However, at heading stage, an obviously lower  $R^2$  value ( $R^2 = 0.379^{**}$ ) was found compared with booting stage.

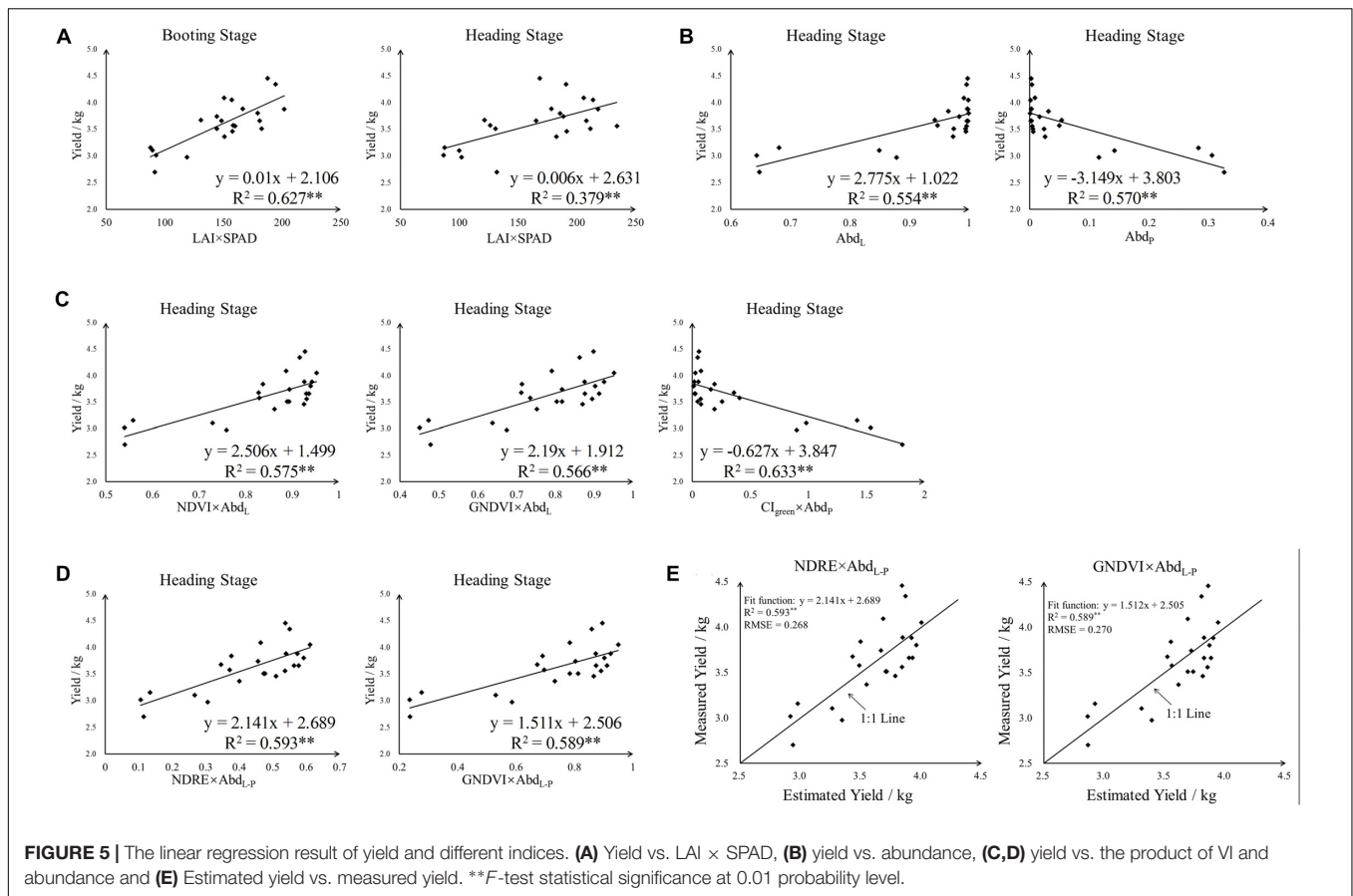
### Correlations of Vegetation Index With Yield and LAI $\times$ SPAD

Vegetation index calculated from reflectance of different wavebands could be used to estimate rice growth parameters and thus has a good potential to predict the rice yield. To compare the relationships of VI to LAI  $\times$  SPAD and yield, we performed a correlation analysis at booting and heading stages. The result of correlation analysis indicated that the VIs were correlated positively to LAI  $\times$  SPAD and yield except VARI, with 0.01 significance levels at two individual stages – **Table 3**. On the whole, the Pearson correlation coefficients ( $r$ ) of VI and LAI  $\times$  SPAD were obviously higher than that of VI and yield at these two stages. Compared with heading stage, most VIs exhibited better correlation with yield and LAI  $\times$  SPAD at booting stage. Generally, the VIs which had better correlation with LAI  $\times$  SPAD also produced higher  $r$  values with yield at both two stages – **Figure 6**. At booting stage, SR and NDRE showed most strong correlations with yield ( $r$  was 0.756 and 0.730 respectively), and the  $r$  values of most VIs were above 0.7 except  $CI_{green}$ , EVI2 and EVI. Where it came to heading stage, the situation was different, the  $r$  values of all VIs were below 0.7. Among the test indices, NDRE and NDVI had the highest  $r$  values with yield ( $r$  was 0.697 and 0.695 respectively). In addition, the relationships of VARI vs. yield appeared extremely weak correlation at both booting and heading stage. For the most relevant VI with yield at each stage (SR at booting stage and NDRE at heading stage), their correlation with LAI  $\times$  SPAD had little difference ( $r$  was 0.844 and 0.818 respectively), but the result was obviously different when correlated with yield ( $r$  was 0.756 and 0.697 respectively).

**TABLE 2** | The statistical description and Shapiro–Wilk test results of LAI  $\times$  SPAD, abundance and yield.

		Observation plots	Minimum value	Maximum value	Mean value	$p$ -Value	Coefficient of variation
LAI $\times$ SPAD	Booting stage	22	87.66	201.74	148.79	0.079	23.25%
	Heading stage	22	86.74	233.92	164.96	0.076	27.88%
Leaf abundance	Heading stage	23	0.64	1.00	0.93	0.000	–
Panicle abundance	Heading stage	23	0.00	0.33	0.06	0.000	–
Yield		23	2.70	4.46	3.61	0.948	11.95%

$p$ -Value is the result of Shapiro–Wilk test.



**TABLE 3 |** The Pearson correlation coefficients of VI with yield and LAI × SPAD at booting and heading stage.

	Growth stage	SR	NDRE	GNDVI	NDVI	CI <sub>rededge</sub>	MTCI	CI <sub>green</sub>	EVI2	EVI	VARI
Yield	Booting stage	0.756**	0.730**	0.728**	0.721**	0.710**	0.704**	0.696**	0.616**	0.586**	-0.381
	Heading stage	0.685**	0.697**	0.644**	0.695**	0.656**	0.661**	0.565**	0.624**	0.575**	-0.386
LAI × SPAD	Booting stage	0.844**	0.896**	0.875**	0.830**	0.896**	0.894**	0.869**	0.757**	0.726**	-0.546**
	Heading stage	0.774**	0.818**	0.794**	0.841**	0.763**	0.756**	0.589**	0.614**	0.557**	-0.293

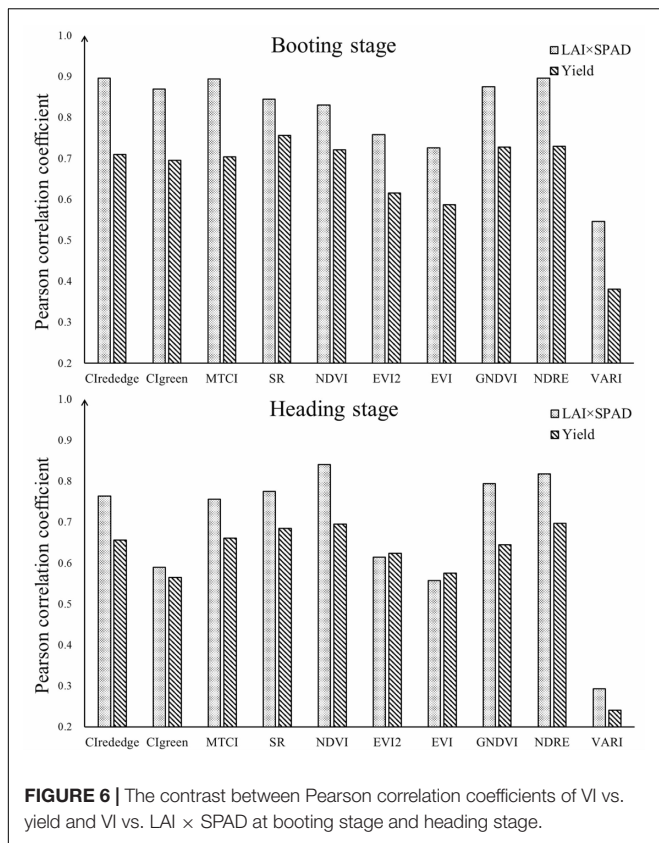
\*\*Correlation is significant at the 0.01 level (two-tailed).

### Spectral Mixture Analysis in Rice Field

In consideration of the impact of mixed pixel, SMA was applied to the UAV image obtained at rice heading stage. According to the ground measured spectra, there was obvious spectral difference in various endmembers – **Figure 4**. In the top layer, both leaf and panicle spectra appeared the peak and valley configuration as that of other typical vegetation spectra characteristics. Nevertheless, leaf reflectance was a little higher than panicle reflectance in blue bands (7 vs. 5%) and it is much higher than panicle reflectance in NIR bands. On the contrary, leaf reflectance was a bit lower than panicle reflectance in red bands. Compared to the top layer, both leaf reflectance and panicle reflectance were low in the bottom layer, but leaf reflectance was still higher than panicle reflectance in NIR bands. As for soil endmember, the reflectance decreased at all wavelengths with soil moisture increasing.

In this study, abundance image of each component was obtained using fully constrained least squares linear SMA and the spectra of selected endmember measured by ASD. As shown in **Figure 7**, significant differences existed in abundance images of different endmembers. On the whole, the brightness of the dry and wet soil abundance images was relatively low. And the bright pixels were mainly clustered in the ridges surrounding the plots in these two abundance images, while the pixels located at rice growing area were really dark. For the other four abundance images, leaf abundance images were obviously brighter than panicle abundance images both in top and bottom layers which was fully corresponded to the actual occurrence in paddy field. Noted that obvious brightness heterogeneity was existed among different plots in the images, and such heterogeneity patterns were quite different in panicle abundance images which indicated the uneven emergence of rice panicles.



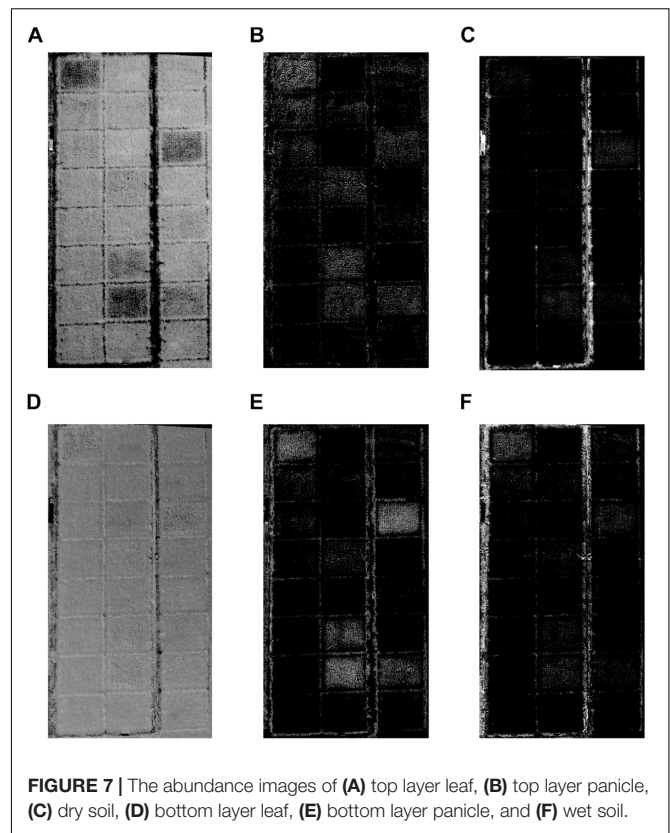


**FIGURE 6 |** The contrast between Pearson correlation coefficients of VI vs. yield and VI vs. LAI × SPAD at booting stage and heading stage.

In consideration of endmember selection, the sum of panicle abundance in top layer and bottom layer was calculated as panicle abundance ( $Abd_p$ ), and the leaf abundance ( $Abd_L$ ) was obtained in the same way. A normal distribution test was also used to  $Abd_L$  and  $Abd_p$  data, the result of statistical description and Shapiro–Wilk test was presented in **Table 2**. Evidently,  $Abd_L$  concentrated near 1, while  $Abd_p$  concentrated near 0. The aggregation effect of abundance data was more apparent in the fit plot of yield with  $Abd_L$  and  $Abd_p$  – **Figure 5B**.

### Rice Yield Estimation Using Vegetation Index and Abundance Data

At heading stage, the uneven emergence of panicles may influence the accuracy of rice yield estimation. Since panicle abundance and leaf abundance were the indicators of how much panicle had been emerging, in our proposed approach the VIs were incorporated with the information of plot-level panicle abundance ( $Abd_p$ ) and leaf abundance ( $Abd_L$ ) to estimate the yield of rice. In this part, yield was firstly correlated with VI,  $VI \times Abd_L$ , and  $VI \times Abd_p$  respectively, and the Pearson correlation coefficients ( $r$ ) were compared – **Table 4**. Generally, after multiplied by  $Abd_L$  or  $Abd_p$ , most VIs produced relatively higher  $r$  values with yield than VI alone. The result of correlation analysis indicated that the  $VI \times Abd_L$  was correlated positively with yield; on the contrary,  $VI \times Abd_p$  showed a negative correlation. And the correlation is numerically higher in  $VI \times Abd_p$  and yield than in  $VI \times Abd_L$  and yield. In



**FIGURE 7 |** The abundance images of (A) top layer leaf, (B) top layer panicle, (C) dry soil, (D) bottom layer leaf, (E) bottom layer panicle, and (F) wet soil.

**TABLE 4 |** The Pearson correlation coefficients of yield with VI,  $VI \times Abd_L$ ,  $VI \times Abd_p$ , and  $VI \times Abd_{L-p}$  at heading stage.

	VI	$VI \times Abd_L$	$VI \times Abd_p$	$VI \times Abd_{L-p}$
NDRE	0.697**	0.751**	−0.775**	0.770**
NDVI	0.695**	0.759**	−0.759**	0.761**
SR	0.685**	0.724**	−0.785**	0.747**
MTCI	0.661**	0.708**	−0.783**	0.736**
Cl <sub>rededge</sub>	0.656**	0.698**	−0.786**	0.725**
GNDVI	0.644**	0.753**	−0.765**	0.767**
EVI2	0.624**	0.726**	−0.752**	0.745**
EVI	0.575**	0.710**	−0.747**	0.738**
Cl <sub>green</sub>	0.565**	0.609**	−0.796**	0.643**
VARI	−0.386	−0.081	−0.734**	0.183

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

particular, VARI showed a really weak correlation with yield at heading stage ( $r$  was  $−0.386$ ), and there was even no correlation in  $VARI \times Abd_L$  and yield. However, after multiplied by  $Abd_p$ , VARI showed a relatively strong correlation ( $r$  was  $−0.734$ ).

For further analysis, regression analysis has been used, and three linear relationship were developed using 23 samples: (1) yield vs. VI, (2) yield vs.  $VI \times Abd_L$  and (3) yield vs.  $VI \times Abd_p$ . Adjusted  $R^2$ , RMSE and  $p$ -value were obtained in SPSS – **Table 5**. For all tested indices, multiplying abundance data ( $VI \times Abd_p$  and  $VI \times Abd_L$ ) significantly increased the goodness of fit with yield, and using  $VI \times Abd_p$  to regress with rice yield was more accurate than using  $VI \times Abd_L$  with lower RMSE

**TABLE 5** | Regression analysis of yield with VI, VI×Abd<sub>L</sub>, VI×Abd<sub>P</sub>, and VI×Abd<sub>L-P</sub> at heading stage.

	Adjusted R <sup>2</sup>				RMSE				p-Value			
	VI	VI×Abd <sub>L</sub>	VI×Abd <sub>P</sub>	VI×Abd <sub>L-P</sub>	VI	VI×Abd <sub>L</sub>	VI×Abd <sub>P</sub>	VI×Abd <sub>L-P</sub>	VI	VI×Abd <sub>L</sub>	VI×Abd <sub>P</sub>	VI×Abd <sub>L-P</sub>
NDRE	0.461	0.543	0.582	0.574	0.317	0.291	0.279	0.282	0.000	0.000	0.000	0.000
NDVI	0.456	0.555	0.556	0.569	0.317	0.288	0.287	0.283	0.000	0.000	0.000	0.000
SR	0.443	0.501	0.599	0.559	0.322	0.304	0.273	0.286	0.000	0.000	0.000	0.000
MTCI	0.410	0.477	0.595	0.537	0.331	0.312	0.275	0.293	0.001	0.000	0.000	0.000
CI <sub>rededge</sub>	0.403	0.463	0.600	0.534	0.333	0.316	0.273	0.294	0.001	0.000	0.000	0.000
GNDVI	0.387	0.546	0.566	0.523	0.338	0.291	0.284	0.298	0.001	0.000	0.000	0.000
EVI2	0.360	0.504	0.545	0.519	0.345	0.304	0.291	0.299	0.001	0.000	0.000	0.000
EVI	0.299	0.481	0.538	0.502	0.361	0.311	0.293	0.304	0.004	0.000	0.000	0.000
CI <sub>green</sub>	0.286	0.341	0.615	0.385	0.364	0.350	0.267	0.338	0.005	0.002	0.000	0.001
VARI	0.108	-0.041	0.518	-0.012	0.407	0.440	0.299	0.434	0.069	0.714	0.000	0.403

and higher Adjusted R<sup>2</sup> values. Among the Abd<sub>L</sub> multiplied indices (VI × Abd<sub>L</sub>), NDVI × Abd<sub>L</sub> had the best goodness of fit with yield (Adjusted R<sup>2</sup> was 0.555), while CI<sub>green</sub> × Abd<sub>P</sub> for Abd<sub>P</sub> multiplied indices (Adjusted R<sup>2</sup> was 0.615). Whereas the saturation phenomenon of NDVI, the GNDVI × Abd<sub>L</sub> was also taken into account. As shown in **Figure 5C**, the linear fitting result of yield and these three indices (NDVI × Abd<sub>L</sub>, GNDVI × Abd<sub>L</sub> and CI<sub>green</sub> × Abd<sub>P</sub>) was acceptable with R<sup>2</sup> above 0.56, and CI<sub>green</sub> × Abd<sub>P</sub> produced the highest R<sup>2</sup> (R<sup>2</sup> was 0.633). However, the saturation phenomenon of NDVI was still existed in NDVI × Abd<sub>L</sub>, and the CI<sub>green</sub> × Abd<sub>P</sub> data concentrated near 0 which caused a phenomenon similar to saturation. There was no obvious saturation in the fit plot of yield and GNDVI × Abd<sub>L</sub>.

The Abd<sub>P</sub> multiplied indices (VI × Abd<sub>P</sub>) could produce higher R<sup>2</sup> when regressed with yield, but was subjected to saturation (the value was closed to 0). As for Abd<sub>L</sub> multiplied indices, they were not susceptible to saturation although their R<sup>2</sup> was relatively lower. And **Figure 5B** showed that Abd<sub>L</sub> related positively with yield but Abd<sub>P</sub> showed a negatively correlation. Therefore, the difference of Abd<sub>L</sub> and Abd<sub>P</sub> (Abd<sub>L-P</sub>) was next used to incorporate with VI. In this way, the correlation analysis and regression analysis were also applied to the relationship of VI × Abd<sub>L-P</sub> and yield – **Tables 4, 5**. Compared with VI, VI × Abd<sub>L-P</sub> acquired better goodness of fit when regressed with yield, Adjusted R<sup>2</sup> could reach 0.574 and RMSE below 0.282 (NDRE × Abd<sub>L-P</sub>). And for the best two indices (NDRE × Abd<sub>L-P</sub> and GNDVI × Abd<sub>L-P</sub>), there was no obvious saturation existing in fit plot – **Figure 5D**. Then on the base of these, leave one out cross-validation approach was utilized to obtain the final yield estimation model. The specific estimation formulas and the goodness of fit between measured yield and estimated yield was shown in **Figure 5E**.

## DISCUSSION

The primary purpose of this study was to improve the accuracy of rice yield estimation at heading stage based on the UAV data.

Vegetation index (VI) and SMA were incorporated to construct new yield estimation algorithm.

In this paper, the same two sets of data were acquired at booting and heading stage respectively, including ground measured data (LAI × SPAD) and UAV data. At these two stages, rice gradually completed the transformation from vegetative growth to reproductive growth. The interval between each two stages was around 2 weeks, and rice had similar growth status except the influence of panicle. At booting stage, the rice plant flourished but with no panicle appearing in canopy. On the contrary, the panicle gradually emerged in canopy at heading stage with the growth of rice plant. At these two stages, the changes of rice leaves were not obvious – **Figure 3**.

First of all, the ground measured LAI × SPAD was correlated with rice yield directly at the two individual stages. Before regression analysis, LAI × SPAD data and yield data have passed the normal distribution test. As shown in **Figure 5A**, the goodness of fit between LAI × SPAD and yield at booting stage was better than that at heading stage. And this result revealed that the booting stage had better predictive ability for rice yield based on ground measured data. The reason for this was probably that the panicle began to emerge at heading stage. Sun et al. (2010) found that the appearance of panicle may lead to the changes of canopy spectral reflectance and thus the predictive ability for yield decreased during the later stages of rice growth. In our study, the LAI × SPAD was calculated from ground measured LAI and SPAD, and the LAI data was obtained by a Sunscan canopy analysis system. According to the principle of the used SunScan instrument, each organ of rice plant (including leaf, panicle, and stem) contributed to the LAI value. Compared with booting stage, the LAI value of heading stage contained extra panicle information. Actually, the LAI × SPAD which derived from green leaves was associated with rice yield (Liu et al., 2017). Therefore, the booting stage had better predictive ability for rice yield than heading stage based on ground measured LAI × SPAD.

The plot-level VI derived from UAV data was also correlated with rice yield. We chose 10 widely used VIs which were successfully applied in estimating vegetation grow-related parameters such as chlorophyll content, LAI, vegetation fraction and grain yield. The similar result was found that the Pearson

correlation coefficient ( $r$ ) between VI and yield at booting stage was mostly higher than that at heading stage – **Table 3**. And the correlation analysis was also applied to analyze the relationship between VI and ground measured LAI  $\times$  SPAD. Although the correlation between VI and LAI  $\times$  SPAD had a bit difference at these two stages, most VIs were well-correlated with LAI  $\times$  SPAD ( $r$  above 0.75). Moreover, the correlation of VI and yield was consistent with the correlation of VI and LAI  $\times$  SPAD – **Figure 6**. This result revealed that the VI which had a better good correlation with LAI  $\times$  SPAD probably produced a higher  $r$  value with yield. The LAI  $\times$  SPAD was, so to speak, a bridge between VI and yield. VI derived from UAV data had correlation with LAI  $\times$  SPAD and thus had a potential to estimate yield. As shown in **Table 3**, the satisfactory correlation could be found between VI and LAI  $\times$  SPAD at both two stages, but the LAI  $\times$  SPAD correlated VIs showed obviously different correlation with yield at booting and heading stage. As mentioned above, the ground measured LAI  $\times$  SPAD contained the extra panicle information at heading stage. In this way, the VI, correlated with LAI  $\times$  SPAD, may also be affected by panicle emergence which lead to the decrease of yield estimation based on UAV data at heading stage.

At heading stage, the plot-level VI was calculated from mixed components including different proportion of panicles, using VI alone for yield regression may introduce unexpected uncertainties. Therefore, the SMA was utilized to improve the predictive ability of VI in rice yield estimation at heading stage.

At this stage, the rice plant had really luxuriant growth and the paddy field was almost covered by rice plants. The canopy of rice was comprised of leaf and panicle – **Figure 3**. And the background of field was mainly soil (wet soil and dry soil), because the water was drained off. In addition, the field management of rice was strict and there were no other plants affecting the growth of rice such as weed. Therefore, six endmembers were selected, including top layer leaf (TL), bottom layer leaf (BL), top layer panicle (TP), bottom layer panicle (BP), dry soil (DS), and wet soil (WS). Based on fully constrained least squares linear SMA, the abundance images of six mainly endmembers in paddy field were derived. Of the six abundance images we obtained, the brightness of dry and wet soil abundance images was relatively low. And compared to panicle abundance image, the leaf abundance image was obviously brighter both in top layer and bottom layer. This was consistent with the actual situation of paddy field. At heading stage, rice leaves grew lushly and the leaves occupied the largest proportion in paddy field. Due to the different nitrogen treatments applied in 24 plots, the growth situation of rice plant in different plots significantly varied from each other even at the same growth stage. It is observed that there was obvious difference among different plots in leaf abundance images and panicle abundance images – **Figure 7**. In consideration of endmember selection, the sum of panicle abundance in top layer and bottom layer was calculated as panicle abundance ( $Abd_p$ ), and the sum of leaf abundance in top layer and bottom layer was calculated as leaf abundance ( $Abd_L$ ). The result of analysis indicated that neither  $Abd_L$  nor  $Abd_p$  followed normal distribution and  $Abd_L$  concentrated near 1, while  $Abd_p$  concentrated near 0. This numerical distribution characteristic was in accordance with the sum-to-one constraint condition

in fully constrained least squares linear SMA. Accordingly, the abundance data was not adequate for yield estimation due to its aggregation effect – **Figure 5B**.

Note that, the LAI  $\times$  SPAD of green leaf was associated with rice yield closely and the VI obtained from UAV data had a good correlation with LAI  $\times$  SPAD at booting and heading stage. However, the LAI  $\times$  SPAD derived from VI contained extra panicle information at heading stage which decreased the predictive ability of VI in rice yield. The abundance data, indeed, implies the proportion information of different components in rice field. In this case,  $Abd_L$  was incorporated with VI to extract the portion of green leaf in total LAI  $\times$  SPAD (contained panicle portion) and thus the VI  $\times$   $Abd_L$  was used to estimate rice yield at heading stage. For comparison purpose, the relationship of VI  $\times$   $Abd_p$  and yield was also analyzed. The result of correlation analysis revealed that after multiplied by  $Abd_L$  or  $Abd_p$ , most VIs produced relatively higher  $r$  values with yield than VI alone – **Table 3**. And the VI  $\times$   $Abd_L$  was correlated positively with yield, while VI  $\times$   $Abd_p$  showed a negative correlation. According to the sum-to-one constraint condition of fully constrained SMA, the correlation of VI  $\times$   $Abd_L$  and yield was opposite to that of VI  $\times$   $Abd_p$  and yield. Obviously, the VI  $\times$   $Abd_L$ , reflected the LAI  $\times$  SPAD contributed by green leaf, had a positive correlation with yield. The result of analysis also revealed that VI  $\times$   $Abd_p$  exhibited a better correlation with yield than VI  $\times$   $Abd_L$  did. However, as shown in **Figure 5C**, VI  $\times$   $Abd_p$  concentrated on 0 and the aggregation effect of VI  $\times$   $Abd_p$  produced a relatively higher  $R^2$ . The reason was that  $Abd_p$  was close to 0 and the value of VI  $\times$   $Abd_p$  deeply depended on  $Abd_p$ . Although there was no aggregation phenomenon in VI  $\times$   $Abd_L$ , the saturation of VI itself still existed in its corresponding VI  $\times$   $Abd_p$  ( $NDVI$  and  $NDVI \times Abd_L$ ).

The  $Abd_L$  multiplied indices (VI  $\times$   $Abd_L$ ) and  $Abd_p$  multiplied indices (VI  $\times$   $Abd_p$ ) both had its advantages and disadvantages, VI  $\times$   $Abd_p$  could produce higher  $r$  but subjected to aggregation, while VI  $\times$   $Abd_L$  was not affected by aggregation but produced relatively lower  $r$ . Therefore,  $Abd_L$  and  $Abd_p$  were combined together to give full play to the advantages of the two. In consideration of that  $Abd_L$  correlated positively with yield and  $Abd_p$  correlated negatively with yield – **Figure 5B**, the difference of  $Abd_L$  and  $Abd_p$  was calculated as  $Abd_{L-p}$  and the relationship between VI  $\times$   $Abd_{L-p}$  and yield was analyzed. The regression analysis showed that the goodness of fit in VI  $\times$   $Abd_{L-p}$  and yield was better than that in VI and yield (with higher Adjusted  $R^2$ ) – **Table 5**. However, not all VIs had better regression results with yield after multiplied by  $Abd_{L-p}$  (such as  $CI_{green} \times Abd_{L-p}$  and  $VARI \times Abd_{L-p}$ ). And the VI  $\times$   $Abd_{L-p}$  was influenced by the VI itself, if the VI had a poor correlation with yield, the corresponding VI  $\times$   $Abd_{L-p}$  may also show a bad correlation with yield. And the result of regression analysis in VI  $\times$   $Abd_{L-p}$  and yield was better than that in VI  $\times$   $Abd_L$  and yield but worse than that in VI  $\times$   $Abd_p$  and yield. It was important to note that there was no obvious aggregation effect in VI  $\times$   $Abd_{L-p}$  – **Figure 5D**. The  $Abd_{L-p}$  multiplied indices (VI  $\times$   $Abd_{L-p}$ ) combined the advantages of VI  $\times$   $Abd_L$  and VI  $\times$   $Abd_p$  and performed a better goodness of fit with no obvious aggregation effect in the relationship of VI  $\times$   $Abd_{L-p}$  and yield. Among all

the tested  $VI \times Abd_{L-P}$ , two indices which had highest Adjusted  $R^2$  were selected ( $NDRE \times Abd_{L-P}$  and  $GNDVI \times Abd_{L-P}$ ) and then leave one out cross-validation approach was utilized to obtain the final yield estimation model – **Figure 5E**. The result proved that  $NDRE$  and  $GNDVI$  multiplied by  $Abd_{L-P}$  could accurately estimate the rice yield at heading stage with  $R^2$  reaching 0.6 and estimation errors below 10%.

In this study, we developed a new approach to estimate rice yield at heading stage using the integration of VI and abundance information retrieved from the UAV image. The approach was simple but it given some significant enlightenment for the yield estimation of grain crop like rice. In the application of RS, the impact of spectral mixture must be considered and this is also important to UAV RS technology which has a really high spatial resolution. And the SMA is a good way to get rid of the influence of different spectral components contained in remotely sensed images. Although the endmembers proposed in our approach were limited to rice yield estimation at heading stage, this work may offer a theoretical framework for yield estimation in grain crops which have obvious grain with significantly different spectra from their leaves. In the future study, we will try to apply this approach to satellite data and in other crops.

## CONCLUSION

In this study, we developed an approach to improve the estimation of rice yield at heading stage using UAV-based Vegetation Index and abundance data. Compared with booting stage, a relatively weaker relationship between VI and rice yield was found at heading stage. The reason was the uneven emergence of rice panicle at heading stage which caused the decrease of predictive ability of VI for rice yield. In order to improve the accuracy of yield estimation at heading stage, a fully

constrained least squares linear spectral mixture method was used to eliminate the influence of the panicle appearance on yield estimation. The abundance images of six mainly endmembers in paddy field was produced based on the six-band UAV image and ground measured spectra, including top layer leaf, bottom layer leaf, top layer panicle, bottom layer panicle, dry soil, and wet soil. The integration of plot-level VI and abundance information can estimate rice yield more accurately than using VI alone. Among the test VIs,  $NDRE$ , and  $GNDVI$  multiplied by the difference of leaf and panicle abundance were the most accurate for yield estimation in rice under different nitrogen fertilizer treatment with estimation errors below 10%.

## AUTHOR CONTRIBUTIONS

SF conceived of the research ideas and built the infrastructure for the study site to make this research possible. SW provided rice yield and advised on data collection. YG and YP designed the experiments in detail and provided valuable guidance on data analysis. RZ and XW provided important insights and suggestions on this research from the perspective of agronomists. BD performed the majority of the data processing and provided the writing of this paper. All authors read and approved the final manuscript and significant contributions to this manuscript.

## FUNDING

This research was supported by the National Natural Science Foundation of China (41771381) and the Fundamental Research Funds for the Central Universities (2042017kf0236).

## REFERENCES

- Aasen, H., and Bolten, A. (2018). Multi-temporal high-resolution imaging spectroscopy with hyperspectral 2D imagers – From theory to application. *Remote Sens. Environ.* 205, 374–389. doi: 10.1016/j.rse.2017.10.043
- Becker-Reshef, I., Vermote, E., Lindeman, M., et al. (2010). A generalized regression based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* 114, 1312–1323. doi: 10.1016/j.rse.2010.01.010
- Behrens, T., Müller, J., and Diepenbrock, W. (2006). Utilization of canopy reflectance to predict properties of oilseed rape (*Brassica napus* L.) and barley (*Hordeum vulgare* L.) during ontogenesis. *Eur. J. Agron.* 25, 345–355. doi: 10.1016/j.eja.2006.06.010
- Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., et al. (2012). Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5, 354–379. doi: 10.1109/JSTARS.2012.2194696
- Broge, N. H., and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* 76, 156–172. doi: 10.1016/S0034-4257(00)00197-8
- Candiago, S., Remondino, F., De Giglio, M., et al. (2015). Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sens.* 7, 4026–4047. doi: 10.3390/rs70404026
- Dash, J., and Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* 25, 5403–5413. doi: 10.1080/0143116042000274015
- Dwyer, J. L., Kruse, F. A., and Lefkoff, A. B. (1995). Effects of empirical versus model-based reflectance calibration on automated analysis of imaging spectrometer data: a case study from the drum mountains, Utah. *Photogramm. Eng. Remote Sens.* 61, 1247–1254.
- Fang, S., Tang, W., Peng, Y., et al. (2016). Remote estimation of vegetation fraction and flower fraction in oilseed rape with unmanned aerial vehicle data. *Remote Sens.* 8, 416. doi: 10.3390/rs8050416
- Feret, J. B., François, C., Asner, G. P., Gitelson, A. A., Martin, R. E., Bidel, L. P. R., et al. (2011). PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens. Environ.* 112, 3030–3043. doi: 10.1016/j.rse.2008.02.012
- Fielding, A. H., and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49. doi: 10.1017/S0376892997000088
- Kohavi, R. (1995). The power of decision tables. in: *Machine Learning: ECML-95*. Berlin: Springer, 174–189. doi: 10.1007/3-540-59286-5\_57
- Gausman, H. W., Allen, W. A., and Cardenas, R. (1969). Reflectance of cotton leaves and their structure 1. *Remote Sens. Environ.* 1, 19–22. doi: 10.1016/S0034-4257(69)90055-8
- Gholizadeh, H., Rahman, A. F., and Rahman, A. F. (2015). *Comparing the Performance of Multispectral Vegetation Indices and Machine-Learning*

- Algorithms for Remote Estimation of Chlorophyll Content: A Case Study in the Sundarbans Mangrove Forest*. Abingdon: Taylor & Francis, Inc.
- Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58, 289–298. doi: 10.1016/S0034-4257(96)00072-7
- Gitelson, A. A., Kaufman, Y. J., Stark, R., et al. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80, 76–87. doi: 10.1016/S0034-4257(01)00289-9
- Gitelson, A. A., Keydan, G. P., and Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophys. Res. Lett.* 33, 431–433. doi: 10.1029/2006GL026457
- Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., and Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 32, 93–114. doi: 10.1029/2005GL022688
- Glenn, F., Daniel, R., and Garry, O. (2010). Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—the canopy chlorophyll content index (CCCI). *Field Crops Res.* 116, 318–324. doi: 10.1016/j.fcr.2010.01.010
- Gong, Y., Duan, B., Fang, S., Zhu, R., Wu, X., Ma, Y., et al. (2018). Remote estimation of rapeseed yield with unmanned aerial vehicle (UAV) imaging and spectral mixture analysis. *Plant Methods* 14, 70. doi: 10.1186/s13007-018-0338-z
- Hansen, P. M., and Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 86, 542–553. doi: 10.1016/S0034-4257(03)00131-7
- Harrell, D. L., Tubaña, B. S., Walker, T. W., et al. (2011). Estimating rice grain yield potential using normalized difference vegetation index. *Agron. J.* 103:1717. doi: 10.2134/agronj2011.0202
- Hatfield, J. L., Gitelson, A. A., Schepers, J. S., and Walthall, C. L. (2008). Application of spectral remote sensing for agronomic decisions. *Agron. J.* 100, 117–131. doi: 10.2134/agronj2006.0370c
- Heinz, D. C. (2001). Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* 39, 529–545. doi: 10.1109/36.911111
- Holzman, M. E., Rivas, R., and Piccolo, M. C. (2014). Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. *Int. J. Appl. Earth Obs. Geoinf.* 28, 181–192. doi: 10.1016/j.jag.2013.12.006
- Jhan, J. P., Rau, J. Y., and Huang, C. Y. (2016). Band-to-band registration and ortho-rectification of multilens/multispectral imagery: a case study of MiniMCA-12 acquired by a fixed-wing UAS. *ISPRS J. Photogramm. Remote Sens.* 114, 66–77. doi: 10.1016/j.isprsjprs.2016.01.008
- Jiang, Z., Huete, A. R., Didan, K., et al. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 112, 3833–3845. doi: 10.1016/j.rse.2008.06.006
- Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 50, 663–666. doi: 10.2307/1936256
- Kira, O., Linker, R., and Gitelson, A. (2015). Non-destructive estimation of foliar chlorophyll and carotenoid contents: focus on informative spectral bands. *Int. J. Appl. Earth Obs. Geoinf.* 38, 251–260. doi: 10.1016/j.jag.2015.01.003
- Lablert, A. S., Goforth, M. A., Steele, C. M., et al. (2011). Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments. *Remote Sens.* 3, 2529–2551. doi: 10.3390/rs3112529
- Li, F., Miao, Y., Feng, G., Yuan, F., Yue, S., Gao, X., et al. (2014). Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Res.* 157, 111–123. doi: 10.1016/j.fcr.2013.12.018
- Liu, H. Q., and Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* 33, 457–465. doi: 10.1109/36.377946
- Liu, X., Zhang, K., Zhang, Z., Cao, Q., Lv, Z., Yuan, Z., et al. (2017). Canopy chlorophyll density based index for estimating nitrogen status and predicting grain yield in rice. *Front. Plant Sci.* 8:1829. doi: 10.3389/fpls.2017.01829
- Lobell, D. B., and Asner, G. P. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sens. Environ.* 93, 412–422. doi: 10.1016/j.rse.2004.08.002
- Peng, Y., Gitelson, A. A., and Sakamoto, T. (2013). Remote estimation of gross primary productivity in crops using MODIS 250 m data. *Remote Sens. Environ.* 128, 186–196. doi: 10.1016/j.rse.2012.10.005
- Pu, H., Chen, Z., Wang, B., et al. (2014). Constrained least squares algorithms for nonlinear unmixing of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* 53, 1287–1303. doi: 10.1109/TGRS.2014.2336858
- Rahman, A., Khan, K., Krakauer, N. Y., et al. (2012). Use of remote sensing data for estimation of aman rice yield. *Int. J. Agric. For.* 2, 101–107. doi: 10.5923/j.ijaf.20120201.16
- Ray, S. S., Jain, N., Miglani, A., Singh, J. P., Singh, A. K., Panigrahy, S., et al. (2010). Defining optimum spectral narrow bands and bandwidths for agricultural applications. *Curr. Sci.* 98, 1365–1369.
- Rouse, J. W. J., Haas, R. H., Schell, J. A., et al. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* 351:309.
- Sakamoto, T., Gitelson, A. A., and Arkebauer, T. J. (2014). Near real-time prediction of US corn yields based on time-series MODIS data. *Remote Sens. Environ.* 147, 219–231. doi: 10.1016/j.rse.2014.03.008
- Sakamoto, T., Shibayama, M., Kimura, A., et al. (2011). Assessment of digital camera-derived vegetation indices in quantitative monitoring of seasonal rice growth. *ISPRS J. Photogramm. Remote Sens.* 66, 872–882. doi: 10.1016/j.isprsjprs.2011.08.005
- Somers, B., Asner, G. P., Tits, L., et al. (2011). Endmember variability in spectral mixture analysis: a review. *Remote Sens. Environ.* 115, 1603–1616. doi: 10.1016/j.rse.2011.03.003
- Sun, H., Li, M. Z., Zhao, Y., Zhang, Y. E., Wang, X. M., and Li, X. H. (2010). The spectral characteristics and chlorophyll content at winter wheat growth stages. *Spectrosc. Spec. Anal.* 30:192.
- Sun, L., Gao, F., Anderson, M. C., et al. (2017). Daily mapping of 30 m LAI and NDVI for Grape yield prediction in California Vineyards. *Remote Sens.* 9:317. doi: 10.3390/rs9040317
- Thenkabail, P. S., Lyon, J. G., and Huete, A. (2011). *Hyperspectral Remote Sensing of Vegetation*. Boca Raton: CRC Press, 1943–1961. doi: 10.1201/b11222
- Thenkabail, P. S., Smith, R. B., and Pauw, E. D. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* 71, 158–182. doi: 10.1016/S0034-4257(99)00067-X
- Tucker, C. J., Holben, B. N., Elgin, J. H. Jr., and McMurtrey, J. E. III (1980). Relationship of spectral data to grain yield variation. *Photogramm. Eng. Remote Sens.* 46, 657–666.
- Turner, D., Lucieer, A., Malenovsky, Z., et al. (2014). Spatial co-registration of ultrahigh resolution visible, multispectral and thermal images acquired with a micro-UAV over Antarctic moss beds. *Remote Sens.* 6, 4003–4024. doi: 10.3390/rs6054003
- Verrelst, J., Camps-Valls, G., Muñoz-Mari, J., PabloRiveraa, J., Veroustraeteb, F., Clevers, J. G. P. W., et al. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. *ISPRS J. Photogramm. Remote Sens.* 108, 273–290. doi: 10.1016/j.isprsjprs.2015.05.005
- Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., and Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sens. Environ.* 115, 3468–3478. doi: 10.1016/j.rse.2011.08.010
- Woolley, J. T. (1971). Reflectance and transmittance of light by leaves. *Plant Physiol.* 47, 656–662. doi: 10.1104/pp.47.5.656
- Wu, C., Niu, Z., Tang, Q., and Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agric. For. Meteorol.* 148, 1230–1241. doi: 10.1016/j.agrformet.2008.03.005
- Xiao, X., Boles, S., Frohling, S., Salas, W., Moore, B. III, Li, C., et al. (2002). Observation of flooding and rice transplanting of paddy rice fields at the site

- to landscape scales in China using VEGETATION sensor data. *Int. J. Remote Sens.* 23, 3009–3022. doi: 10.1080/01431160110107734
- Zarcotejada, P. J., Guilléncliment, M. L., Hernándezclemente, R., et al. (2013). Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). *Agric. For. Meteorol.* 171–172, 281–294. doi: 10.1016/j.agrformet.2012.12.013
- Zhou, X., Zheng, H. B., Xu, X. Q., et al. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.* 130, 246–255. doi: 10.1016/j.isprsjprs.2017.05.003

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

Copyright © 2019 Duan, Fang, Zhu, Wu, Wang, Gong and Peng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.