

RESEARCH

Open Access



Estimation of area- and mass-based leaf nitrogen contents of wheat and rice crops from water-removed spectra using continuous wavelet analysis

Dong Li, Xue Wang, Hengbiao Zheng, Kai Zhou, Xia Yao, Yongchao Tian, Yan Zhu, Weixing Cao and Tao Cheng*

Abstract

Background: The visible and near infrared region has been widely used to estimate the leaf nitrogen (N) content based on the correlation of N with chlorophyll and deep absorption valleys of chlorophyll in this region. However, most absorption features related to N are located in the shortwave infrared (SWIR) region and the physical mechanism of leaf N estimation from fresh leaf reflectance spectra remains unclear. The use of SWIR region may help us reveal the underlying mechanism of casual relationships and better understand the spectral responses to N variation from fresh leaf reflectance spectra. This study combined continuous wavelet analysis (CWA) and water removal technique to improve the estimation of N content and leaf mass per area (LMA) by reducing the effect of water absorption and enhancing absorption signals in the SWIR region. The performance of the wavelet-based method was evaluated for estimating leaf N content and LMA of rice and wheat crops from fresh leaf reflectance spectra collected over a 2-year field experiment and compared with normalization difference (ND)-based spectral indices.

Results: The LMA and area-based N content (N_{area}) exhibited better correlations with the determined wavelet features derived from the water-removed (WR) spectra (LMA: $R^2 = 0.71$, N_{area} : $R^2 = 0.77$) than those from the measured reflectance (MR) spectra (LMA: $R^2 = 0.62$, N_{area} : $R^2 = 0.64$). The wavelet features performed remarkably better than the optimized ND indices for the estimations of LMA and N_{area} with MR spectra or WR spectra. Based on the best estimations of LMA and N_{area} with wavelet features from WR spectra, the mass-based N content (N_{mass}) could be retrieved with a high accuracy ($R^2 = 0.82$, RMSE = 0.32%) in the indirect way. This accuracy was higher than that for N_{mass} obtained in the direct use of a single wavelet feature ($R^2 = 0.68$, RMSE = 0.42%).

Conclusions: The enhancement of absorption features in the SWIR region through the CWA applied to water-removed (WR) spectra was able to improve the spectroscopic estimation of leaf N content and LMA as compared to that obtained with the reflectance spectra of fresh leaves. The success in estimating LMA and N with this method would advance the spectroscopic estimations of grain quality parameters for staple crops and individual dry matter constituents for various vegetation types.

Keywords: Nitrogen content, Water-removed, Wavelet analysis, Shortwave infrared

*Correspondence: tcheng@njau.edu.cn

National Engineering and Technology Center for Information Agriculture (NETCIA), Key Laboratory of Crop System Analysis and Decision Making, Ministry of Agriculture and Rural Affairs, Jiangsu Key Laboratory for Information Agriculture, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, One Weigang, Nanjing, Jiangsu 210095, China



Background

Rice and wheat are two major staple crops in the world and provide primary dietary calories and protein for the global population [1, 2]. Leaf nitrogen (N) content is an important indicator of crop photosynthetic capacity [3] and is needed by agronomists for making fertilization recommendations [4]. Quantification of leaf N content could provide valuable information for monitoring crop physiology [5], and practicing precise farming [6] so as to improve the use efficiency of nitrogen fertilizers. Remote sensing has been widely used as a non-destructive approach for estimating leaf N content in the past few decades [7–9]. The common practice is to establish linear or nonlinear relationships between leaf N content and spectral features derived from leaf reflectance spectra.

Curran [10] listed 42 absorption features caused by bending and stretching of foliar chemical bonds in the 400–2400 nm range and identified many of them related to N in the shortwave infrared region (SWIR) region (1000–2400 nm). Based on a number of N-sensitive wavelengths in the SWIR region, leaf N can be estimated accurately from the reflectance spectra of dried and ground leaves [11, 12]. However, these absorption features are masked by water absorption and hence not clearly visible in the SWIR reflectance spectra of fresh leaves, thereby leading to weaker signals of N in the entire spectra [13–15]. This has been proven by the findings that the accuracy of N estimation from the reflectance spectra of fresh leaves is lower than that of dry leaves or dried and ground leaves [11]. Therefore, the SWIR region in the reflectance spectra of fresh leaves has seldom been used for N estimation. Instead, the visible and near infrared (VNIR) region has been widely used for this purpose [8, 16, 17] because nitrogen and chlorophyll are well related [3, 18] and the latter has deep absorption valleys in this region [19, 20]. To date, most studies built N estimation models with the spectral information of fresh leaves in the VNIR region alone [21–23] or VNIR and SWIR regions [9, 24, 25]. A few studies made use of SWIR reflectance alone [12] but their focus was on the spectra of dried leaves or leaf powder, rather than fresh leaves that are easier to handle for reflectance measurements. How accurately the N content could be estimated from reflectance spectra of fresh leaves in the SWIR region alone is poorly understood.

As a group of N-containing compounds in fresh leaves, chlorophylls account for only a portion of the total nitrogen [3, 26]. The physical mechanism underlying the spectroscopic estimation of leaf N content remains unclear due to the use of indirect N-sensitive wavelengths in the VNIR region. The use of SWIR region may help us reveal the mechanism and better understand the spectral responses to N variation [10, 11]. A major problem

constraining the satisfactory estimation of N content is the effect of leaf water absorption on reflectance spectra, which was claimed to be removed to within 10% by Kokaly and Clark [11]. Recently, a few studies adopted a water removal technique originally proposed by Gao and Goetz [27] to remove the effect of water absorption so as to improve the estimation of N content [15, 28] and N to phosphorous ratio [29]. These studies compared water removal to traditional spectral transformation techniques, but did not decompose the water removal process for understanding the underlying mechanism of this technique. To make use of the spectral information in the SWIR region, they applied traditional methods such as stepwise multiple linear regression and partial least-squares regression to the spectra after the water removal process. Based on these methods, some of the selected wavelengths were not related to the absorption features of the chemical being examined [12] and the regressions were often faced with model overfitting and indirect relationships [10]. In addition, how to enhance the causal absorption features in WR spectra for the improved estimation of leaf N content is poorly understood.

Recently, continuous wavelet analysis (CWA) has been widely used to estimate chlorophyll content [30–32], water content [33, 34], dry matter content [35], and leaf area index [36] from leaf and canopy reflectance spectra. After the application of continuous wavelet transform, a reflectance spectrum is decomposed into a number of scale components, which have the same length as the reflectance spectrum and are composed of wavelet features as a function of wavelength and scale [33]. Wavelet features have been proven to be superior to vegetation indices (VIs) in the characterization of absorption by foliar chemicals in reflectance spectra [33, 35, 37]. However, none of previous studies have used CWA and water removal techniques collectively and have investigated the application of CWA to WR spectra for examining dry matter and nitrogen related absorption characteristics. In contrast to the continuum removal operation used in Schlerf et al. [15] and Ramoelo et al. [28], the continuous wavelet transform (CWT) is a linear operation and enables us to decompose the CWT of WR spectra into that of spectral addition or subtraction. Combining CWA and water removal has the potential to improve the estimation of N content by reducing the effect of water absorption and enhancing N absorption signals.

Common measures for expressing leaf N content are either area-based (N_{area} , g/m²) or mass-based (N_{mass} , %). Because of its tight correlation with photosynthetic capacity [3] and the widespread use in fertilization management [38, 39], N_{mass} has been extensively studied and estimated from remotely sensed data [7, 23–25, 40]. Both N_{area} and N_{mass} can be directly obtained

in a destructive way, but remote sensing usually works better for estimating N_{area} . The leaf biochemistry for dried ground samples is usually expressed as concentration (mass-based, mass per unit dry weight). For intact fresh leaves, the biochemistry is often expressed as content (area-based, mass per unit leaf area) [41]. The use of chemical content may be more suitable for remote sensing applications because it is a better representation of the interaction between matter and light per unit surface area. For example, the units of biochemical parameters used in PROSPECT model are all area-based [19, 42]. It may be proven by the higher correlations of many VIs with N_{area} than N_{mass} documented in Hansen and Schjoerring [22] and Jay et al. [43]. Although the interaction of leaves with light per unit leaf area is directly related to such area-based traits as N_{area} [41, 44], few studies paid attention to the estimation of N_{area} [13, 43, 45]. As pointed out by Wright et al. [46], N_{mass} and N_{area} are interconverted via leaf mass per area (LMA, g/m^2). To the best of our knowledge, only one study made use of this connection and derived N_{area} indirectly from estimated N_{mass} and LMA [25]. Based on our understanding of N_{mass} and N_{area} , this study will investigate first estimating the two area-based factors (N_{area} and LMA) and then indirectly estimating N_{mass} instead. The SWIR region encompasses all of the major absorption features of dry matter and N and may provide sufficient spectral information to estimate LMA, N_{area} and N_{mass} .

Therefore, the objectives of this study were (1) to determine the wavelet features most sensitive to LMA and N_{area} from both leaf reflectance spectra and WR spectra in the SWIR region, (2) to evaluate the feasibility of improving LMA, N_{area} and N_{mass} estimations with the integration of CWA and the water removal technique, and (3) to compare the CWA and normalization difference (ND) index approaches in the performance of LMA, N_{area} and N_{mass} estimations.

Methods

Experimental design

Four experiments were conducted at the experimental station in Rugao, Jiangsu of eastern China ($120^{\circ}45'E$ and $32^{\circ}16'N$) with two for rice and two for wheat. Both crops were chosen as they were grown in rotation in this experimental area. We intended to develop robust models for both of them despite their differences in biochemical parameters and surface properties [47]. The pooled data represented a wider range of samples and were beneficial to test the stability of the method proposed in this study. The treatment in each experiment represented variations in cultivar type, nitrogen fertilization rate and planting density with three replications. These treatments were applied to create wide ranges for N_{mass} , N_{area} and LMA, which includes the extremely high and low values. There were a total of 36 plots with the same size of $5\text{ m} \times 6\text{ m}$ for each experiment. Details about the experiments and sampling dates are shown in Table 1. Rice was transplanted and grown in a grid pattern with a plant spacing of 15 cm and two row spacings of 30 and 50 cm. Wheat was sowed in drill with different row spacings and the rows were oriented in a south-north direction. The 2-year experimental data for rice and wheat were combined to form RICE and WHEAT datasets, respectively.

Measurements of reflectance spectra and chemical constituents

Three leaves per plot were collected and each leaf made a sample. Their reflectance spectra were measured using an ASD FieldSpec Pro spectrometer (Analytical Spectral Devices, Boulder, CO, USA) assisted with a leaf clip accessory. Reflectance spectra were taken per leaf for three leaf positions and these spectra were averaged to represent the leaf sample spectrum. The spectrometer collects data at 1.4 nm and 2 nm sampling interval in the 350–1000 nm and 1000–2500 nm spectral regions, respectively. The spectral data were obtained at a 1 nm spectral interval.

Table 1 Summary of designs and ground samplings periods of the field plot experiments

Dataset	Year of data collection	Cultivar	Nitrogen rate (kg hm^{-2})	Planting density (cm)	Growth stages for sampling	Number of samples
Rice	2015	Yliangyou 1* Wuyunjing 24**	0, 100, 200, 300	$30 \times 15, 50 \times 15$	Jointing, booting, heading	95
	2016	Yliangyou 1* Wuyunjing 24**	0, 150, 300	$30 \times 15, 50 \times 15$	Jointing, booting	74
Wheat	2016	Yangmai 18*	0, 80, 150, 220	20, 30, 40	Jointing, booting, heading	127
	2017	Yangmai 15* Yangmai 16**	0, 150, 300	25, 40	Late-jointing, heading, anthesis	104

*, **Denote erect-leaf and drooping-leaf cultivars, respectively

After collection of reflectance spectra, the surface area (A , m^2) and fresh weight (FW, g) were measured immediately for every leaf sample. The leaf area (A , m^2) was determined with a HP G4050 scanner (HP Development Company, L.P., USA) and calculated as the product of pixel count per leaf and the actual area represented by each pixel. The leaf dry weight (DW, g) was measured after the leaves were dried in the oven at 80 °C for 48 h. Finally, equivalent water thickness (EWT, g/m^2), leaf mass per area (LMA, g/m^2) and area-based leaf nitrogen content (N_{area} , g/m^2) were calculated with the equations below:

$$EWT = (FW - DW)/A \quad (1)$$

$$LMA = DW/A \quad (2)$$

$$N_{area} = N_{mass} \times LMA \quad (3)$$

where mass-based leaf nitrogen content (N_{mass} , %) was determined using Kjeldahl method with SEAL AutoAnalyzer 3 HR (SEAL Analytical, Ltd., German). Note that the unit g/m^2 of EWT can be interconverted to cm ($1 g/m^2 = 10^{-4} g/cm^2 = 10^{-4} cm$) via the density of water ($1 g/cm^3$: $1 g \text{ water} = 1 cm^3 \text{ water}$) [48].

The adjusted water-removal technique

The water removal technique was originally proposed by Gao and Goetz [27] with the assumption that there is a linear spectral background level and a nonlinear combination of a reflectance spectrum of leaf water and that of leaf dry matter for fresh leaves. However, the assumption of linear spectral background is only valid for small wavelength regions [27] and this technique is often performed on spectral segments, such as 1500–1780 nm and 2100–2200 nm [15, 27]. The use of narrow and discrete spectral segments may limit the utility of full-range spectral information. Therefore, this technique was adjusted in this study to obtain full-range water removed (WR) spectra by adopting PROSPECT-5B [42] for modeling leaf reflectance spectra. The input parameters of PROSPECT-5B include chlorophyll content (C_{ab}), carotenoid content (C_{xc}), EWT, LMA, and structural parameter (N_{struc}) [42]. To estimate the WR spectra, the first step was to find the best combination of input parameters for PROSPECT-5B by minimizing the following merit function:

$$\min = \sum_{\lambda=1}^n (R_{\lambda} - \widetilde{R}_{\lambda})^2 \quad (4)$$

where R_{λ} and \widetilde{R}_{λ} are the measured reflectance (MR) and modeled reflectance values of a spectrum at wavelength λ , respectively. The mathematical optimization

was performed with the ‘CONSTRAINED_MIN’ function in IDL 8.3 (Exelis Visual Information Solutions, Boulder, CO, USA), which gives the best combination of parameters that minimizes the difference between modeled and measured reflectance for the specified spectral range. Since most nitrogen and dry matter related absorption features are located in the short-wave infrared (SWIR) region [10], the spectral range 1000–2400 nm was used in this study for implementing PROSPECT model inversion as well as the determination of WR spectra. In addition, the C_{ab} and C_{xc} were fixed at 30 and 3 $\mu g/cm^2$, respectively, because these pigments have no absorption in this spectral region [42, 47].

After the PROSPECT model inversion process, the second step was to model the leaf water reflectance spectra by inputting the inverted parameters into PROSPECT-5B with zero LMA values. Since pigments have no influences on the reflectance in this spectral region, the parameters C_{ab} and C_{xc} were fixed at 30 and 3 $\mu g/cm^2$, respectively [42, 47]. As water was the only absorbing chemical left in fresh leaves for this purpose, the modeled leaf water reflectance spectra were named as water-only (WO) spectra. The residual spectra, namely the WR spectra, were estimated as below:

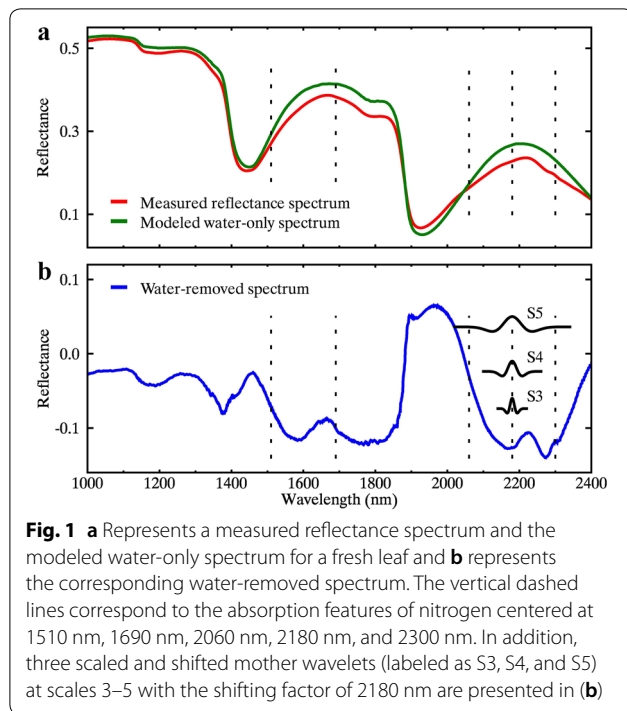
$$WR_{\lambda} = (R_{\lambda} - R_{\lambda}^{wo})/\bar{R} \quad (5)$$

where R_{λ}^{wo} was the reflectance of a WO spectrum at wavelength λ . \bar{R} was the mean reflectance of a MR spectrum over all wavelengths.

After the absorption of dry matter was removed, the amplitude of modeled leaf water reflectance increased for all wavelengths except for the strong water absorption region around 1900 nm (Fig. 1a). Many nitrogen-related absorption features that were not obvious in the MR spectrum became visible in the WR spectrum, such as the feature at 2180 nm (Fig. 1b). Note that the reflectance values have physical meanings and cannot be negative, but the amplitudes of WR spectra were relative values estimated from MR spectra and the negative MR values only meant the negative differences between a reflectance spectra and the corresponding water-only spectrum. Similar terminology is also seen in “derivative spectra”, which represents the derivative data of reflectance spectra and encompasses negative and positive values [49].

Continuous wavelet analysis (CWA)

Mathematically, continuous wavelet transform (CWT) is a linear operation that performs the convolution of reflectance spectrum with a scaled and shifted mother wavelet. The realization process is shown as below:



$$\psi_{a,b}(\lambda) = \frac{1}{\sqrt{a}} \psi\left(\frac{\lambda - b}{a}\right) \quad (6)$$

$$W_r(a, b) = \int_{-\infty}^{+\infty} r(\lambda) \psi_{a,b}(\lambda) d\lambda \quad (7)$$

where $\psi(\lambda)$ is the mother wavelet function and $\psi_{a,b}(\lambda)$ is the scaled and shifted version of $\psi(\lambda)$. $W_r(a, b)$ is the wavelet coefficient (or wavelet feature, denoted as $WF_{b,a}$) for the scaling factor a and the shifting factor b . The scaling factor indicates the width of the scaled mother wavelet, which can be comparable with the width of an absorption feature. Narrow absorption features can be captured by a low scaling factor while broad features need high scaling factors. The scaling factor used in this study is at dyadic scales 2^3 (scale 3), 2^4 (scale 4), 2^5 (scale 5), and 2^6 (scale 6). The shifting factor determines the central wavelength of shifted mother wavelet, which is used to capture the peak or valley of an absorption feature. Taking the absorption feature of nitrogen at 2180 nm as an example, the three scaled and shifted mother wavelets (Scale 3, Scale 4, and Scale 5) are displayed in Fig. 1b. The spectral width of the corresponding wavelet feature increases with the scaling factor. Different with a spectral index calculated with two discrete wavelengths, a wavelet

feature represents the information for a continuous spectral sub-region that determined by scaling and shifting factors. Although CWA is known to be resistant to signal noise, noise reduction is not the focus of this study. Rather, we took the advantage of CWA in capturing and enhancing the absorption features by nitrogen and dry matter. The multiscale property of CWA is well suited for characterizing the absorption changes that may occur in different widths over different wavelength ranges.

The one-dimensional reflectance spectra are converted into two-dimensional wavelet power (magnitude of wavelet coefficient) datasets after the application of CWT. Hence, a scalogram of coefficients of determination (R^2) was obtained at all wavelengths and scales [33]. Moreover, the regions that comprised the top 1% R^2 values were extracted from the comprised scalograms and finally the wavelet feature with the highest R^2 was selected. In this study, the mother wavelet function was determined as the second derivative of the Gaussian function because of its similarity to the shapes of leaf absorption features [35]. CWT was conducted using the ‘WV_CWT’ function in IDL 8.3 (Exelis Visual Information Solutions, Boulder, CO, USA). Since the wavelet is linearly additive [37], the wavelet power derived from a WR spectrum equals the difference in wavelet power between the MR spectrum and the WO spectrum divided by the mean of reflectance over all wavelengths as follows:

$$WP(WR_\lambda) = WP(R_\lambda) / \bar{R} - WP(R_\lambda^{wo}) / \bar{R} \quad (8)$$

where WP represents the wavelet power and other symbols are the same as in Eq. (5).

Evaluation of estimation accuracy

The widely used normalization difference (ND) indices were calculated in this study for comparison with the wavelet features in the evaluation of estimation accuracy. The ND index wavelengths were determined as the best band combination from R^2 scalograms except these combinations with difference between two wavelengths within 10 nm [50, 51]. Since the aim of this study was to build a robust model across rice and wheat crops, the most sensitive wavelet features and ND indices to LMA, N_{area} and N_{mass} were determined separately from pooled data. These spectral features were used to represent their regression relationships and generate direct estimates of those traits. Since area-based foliar traits are directly related to the interaction of foliar constituents with light and can be more easily estimated with spectral features than mass-based traits [41, 44], N_{mass} was also estimated in an indirect way as the ratio of estimated LMA and N_{area} values as below:

$$\text{Estimated } N_{\text{mass}} = \frac{\text{Estimated } N_{\text{area}}}{\text{Estimate LMA}} \times 100\% \quad (9)$$

The indirect estimation of N_{mass} , based on two wavelet features separately sensitive to N_{area} and LMA, might outperform the direct estimation with a single wavelet feature.

The estimation accuracies were evaluated with the metrics R^2 and root mean square root (RMSE):

$$R^2 = 1 - \frac{\sum_i (y_i - y'_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (10)$$

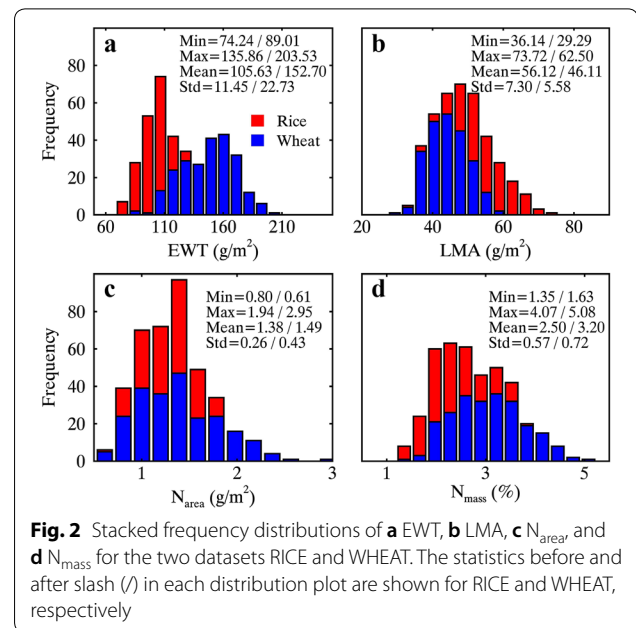
$$RMSE = \sqrt{\frac{\sum_i (y_i - y'_i)^2}{n}} \quad (11)$$

where y_i and y'_i are the measured and estimated trait values for sample i . \bar{y} is the arithmetic mean of trait and n is the number of samples. Due to the intrinsic discrepancy between rice and wheat leaves, the best feature for rice may not be exactly the same as that for wheat and vice versa. A generic model that would work best across rice and wheat samples was preferred over that specific to one crop alone. Therefore, the selected features were determined from the pooled data of rice and wheat. These evaluation metrics were calculated from two commonly used validation processes. One is a 10-fold cross validation procedure applied to the pooled data, and the other is the leave-one-out (LOO) method. In the LOO process, three sub-datasets were used for model calibration and the left one was used for validation. This validation process was repeated for the every one of the four sub-datasets (Rice 2015, Rice 2016, Wheat 2016, and Wheat 2017) in Table 1. Since many studies also used the VNIR region to calculate the Red-Edge Chlorophyll Index ($CI_{\text{red edge}}$) for leaf N content estimation [16, 52], a brief comparison between SWIR and VNIR regions was presented.

Results

Descriptive statistics of foliar traits

Differences in foliar traits between the two datasets could be observed in terms of range and magnitude (Fig. 2). Compared with RICE, the WHEAT dataset exhibited a higher mean and a wider range for EWT, N_{area} and N_{mass} ($p < 0.01$) but a lower mean and a narrower range for LMA ($p < 0.01$). In addition, N_{area} was positively correlated with N_{mass} for both RICE ($r = 0.79$, $p < 0.01$) and WHEAT ($r = 0.92$, $p < 0.01$) (Table 2). N_{area} was positively correlated with LMA for WHEAT ($r = 0.65$, $p < 0.01$) but



not correlated with EWT for both datasets. N_{mass} was correlated with LMA for RICE ($r = -0.56$, $p < 0.01$) and WHEAT ($r = 0.31$, $p < 0.01$) but correlated with EWT only for RICE ($r = -0.30$, $p < 0.01$).

As one of the most important factors in this experiment, the effect of N treatment on leaf reflectance spectra is displayed in Fig. 3 by taking the WHEAT 2017 samples as an example. The N_{mass} , N_{area} and LMA values for WHEAT 2017 increased with the N fertilization rate, and EWT was almost invariant for N1 and N2 (Fig. 3a–d). The leaf reflectance spectra under three N treatments exhibited marginal differences in the SWIR region (Fig. 3e). This phenomenon was expected for fresh leaves, which demonstrated the need of water removal technique to remove the water absorption and CWA to enhance the absorption features of nitrogen and dry matter.

Relationships of N_{area} with optimal wavelet features and ND indices

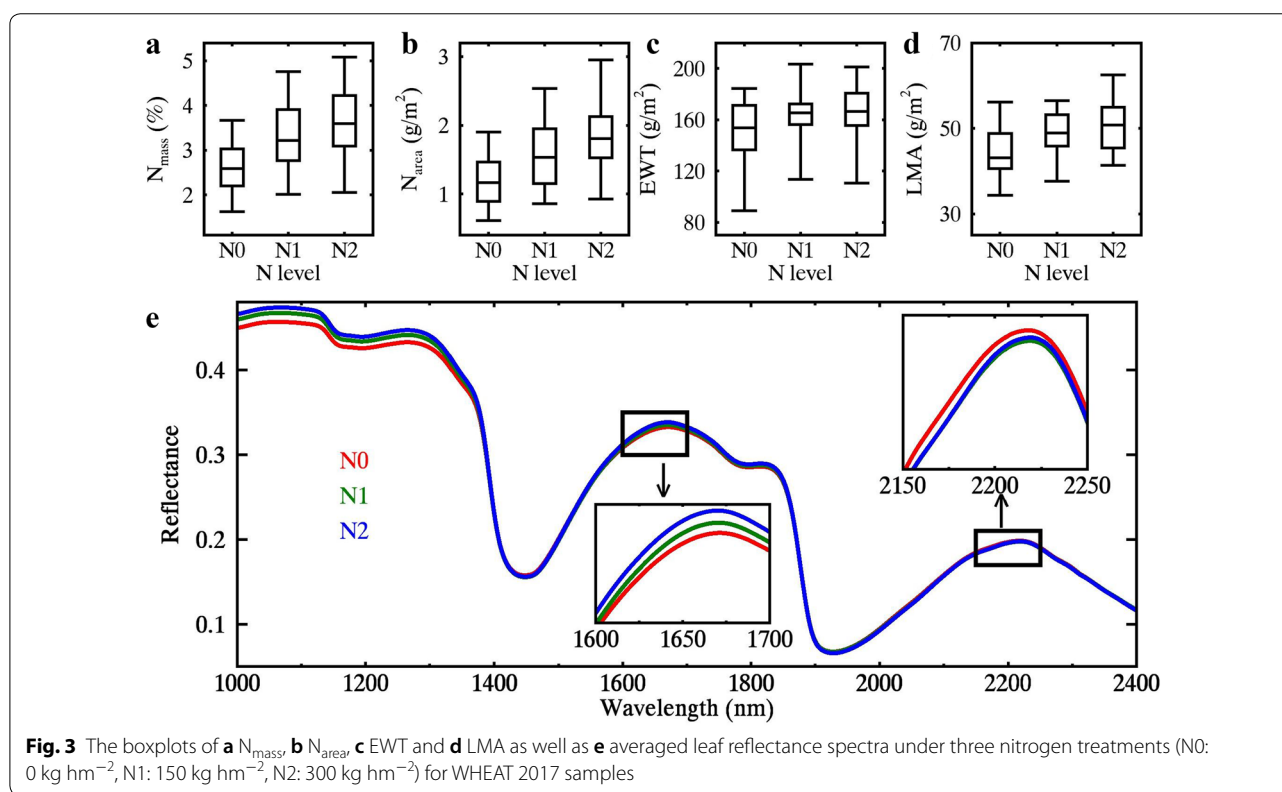
The highest R^2 values between N_{area} and wavelet features for pooled data were 0.64 and 0.77 for MR and WR spectra, respectively (Fig. 4a, b). The two feature regions determined for MR spectra were centered at 2060 nm and 2180 nm, and both of them matched up with the absorptions of protein and nitrogen [10]. As for the WR spectra, a similar feature region with longer wavelengths was found, and a minor one exhibited much shorter wavelengths than the counterpart for the MR spectra.

The R^2 contour maps for the relationship between N_{area} and ND indices are shown in Fig. 4c, d for MR (maximum

Table 2 Pearson’s correlation matrix for EWT, LMA, N_{area} and N_{mass} (top-right triangle for RICE and bottom-left triangle for WHEAT)

	EWT	LMA	N_{area}	N_{mass}
EWT		0.56**	0.06	-0.30**
LMA	0.08		0.04	-0.56**
N_{area}	0.05	0.65**		0.79**
N_{mass}	0.01	0.31**	0.92**	

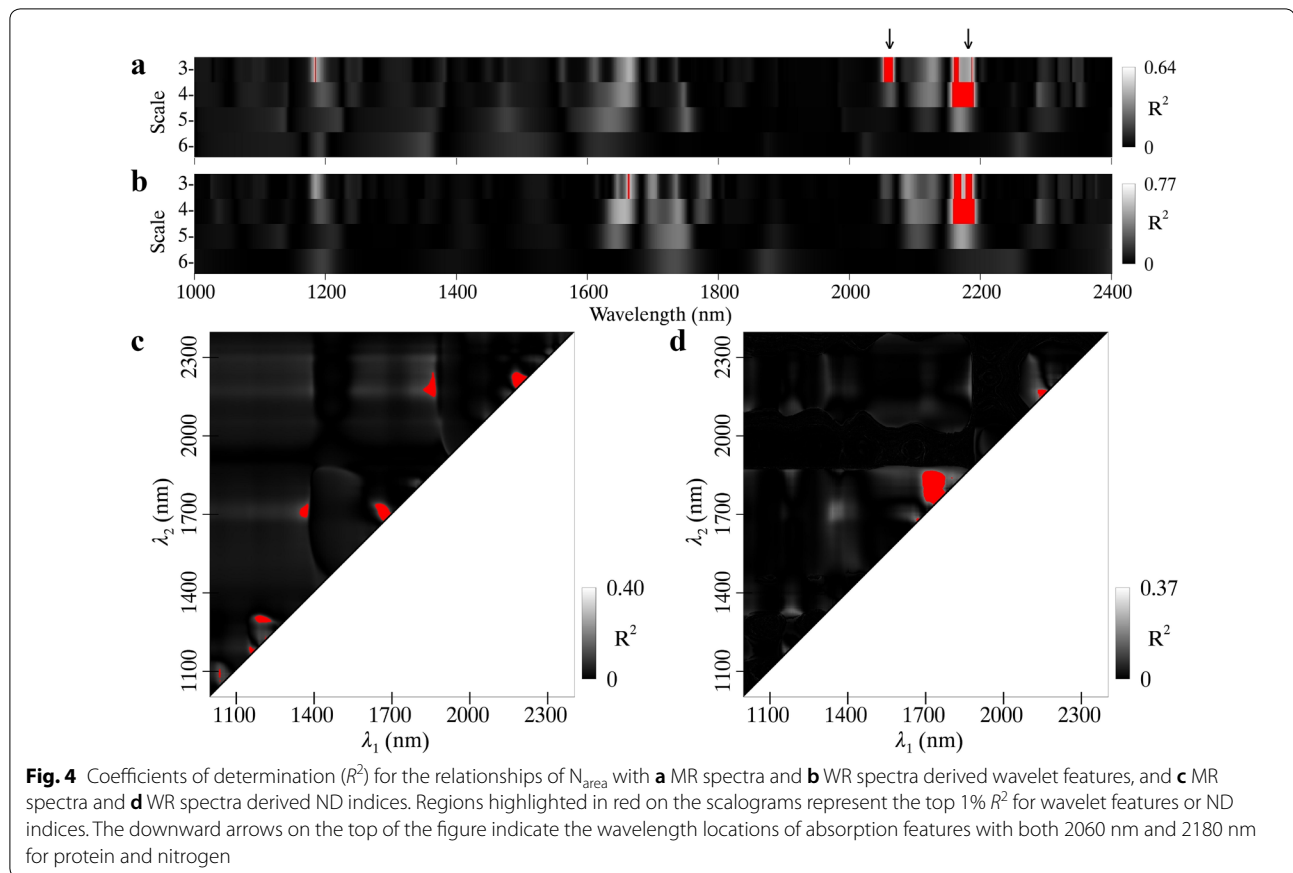
**Significant correlations with $p < 0.01$



$R^2 = 0.40$) and WR (maximum $R^2 = 0.37$) spectra, respectively. Seven wavelength regions most sensitive to N_{area} were observed for the MR spectra and five of them were close to the diagonal line, which indicates the presence of similar wavelengths in an ND index. As for the WR spectra, the majority of wavelength combinations were located in the 1700–1800 nm region in which both λ_1 and λ_2 fall.

Figure 5 shows the relationships of N_{area} with four optimal spectral features ($WF_{2060,3}$, $WF_{2181,4}$, $ND_{2192,2202}$ and $ND_{1749,1773}$) determined from the top 1% correlations in Fig. 4. The two optimal wavelet features ($WF_{2060,3}$ and

$WF_{2181,4}$) used the spectral information over two spectral sub-regions that matched well with the absorption features centered at 2060 nm and 2180 nm (Fig. 4a, b). Nevertheless, the two optimal ND indices ($ND_{2192,2202}$ and $ND_{1749,1773}$) used the spectral information at four individual wavelengths 1749 nm, 1773 nm, 2192 nm and 2202 nm (Fig. 4c, d). All of the relationships for WHEAT, RICE and pooled data were linear. After the application of the WR process, the $N_{area} \sim WF$ correlation exhibited a pronounced improvement (R^2 increased from 0.71 to 0.83) over that from the MR spectra for the WHEAT dataset but a slight increase for RICE dataset (Fig. 5a,



b). Overall, this caused an increase of R^2 from 0.64 to 0.77 for the pooled data. For $N_{\text{area}} \sim \text{ND}$ correlations, the combination of an increase for RICE and a decrease for WHEAT resulted in an overall decrease in R^2 (Fig. 5c, d).

Because of the linearly additive property of CWT, the final regressions from the WR spectra could be decomposed into intermediate regressions to understand how the effect of water absorption was removed. Taking the N_{area} estimation as an example, Fig. 6 shows the intermediate and final regressions based on Eq. (8) to illustrate the advantages of the water removal technique. N_{area} was well correlated to the $\text{WF}_{2181,4}$ derived from the MR spectra (Fig. 6a) but not to the $\text{WF}_{2181,4}$ derived from the WO spectra for both RICE and WHEAT datasets (Fig. 6b). The $\text{WF}_{2181,4} \sim N_{\text{area}}$ relationship in the MR spectra was weaker than that in the WR spectra (Fig. 6c, d). After the removal of water absorption information in the WO spectra (Fig. 6e), the WR-derived $\text{WF}_{2181,4}$ that equaled the difference between MR-derived and WO-derived $\text{WF}_{2181,4}$ values contained no more water information ($p < 0.001$) (Fig. 6f). Clearly, the WR-derived $\text{WF}_{2181,4}$ showed improved relationships with N_{area} with increases

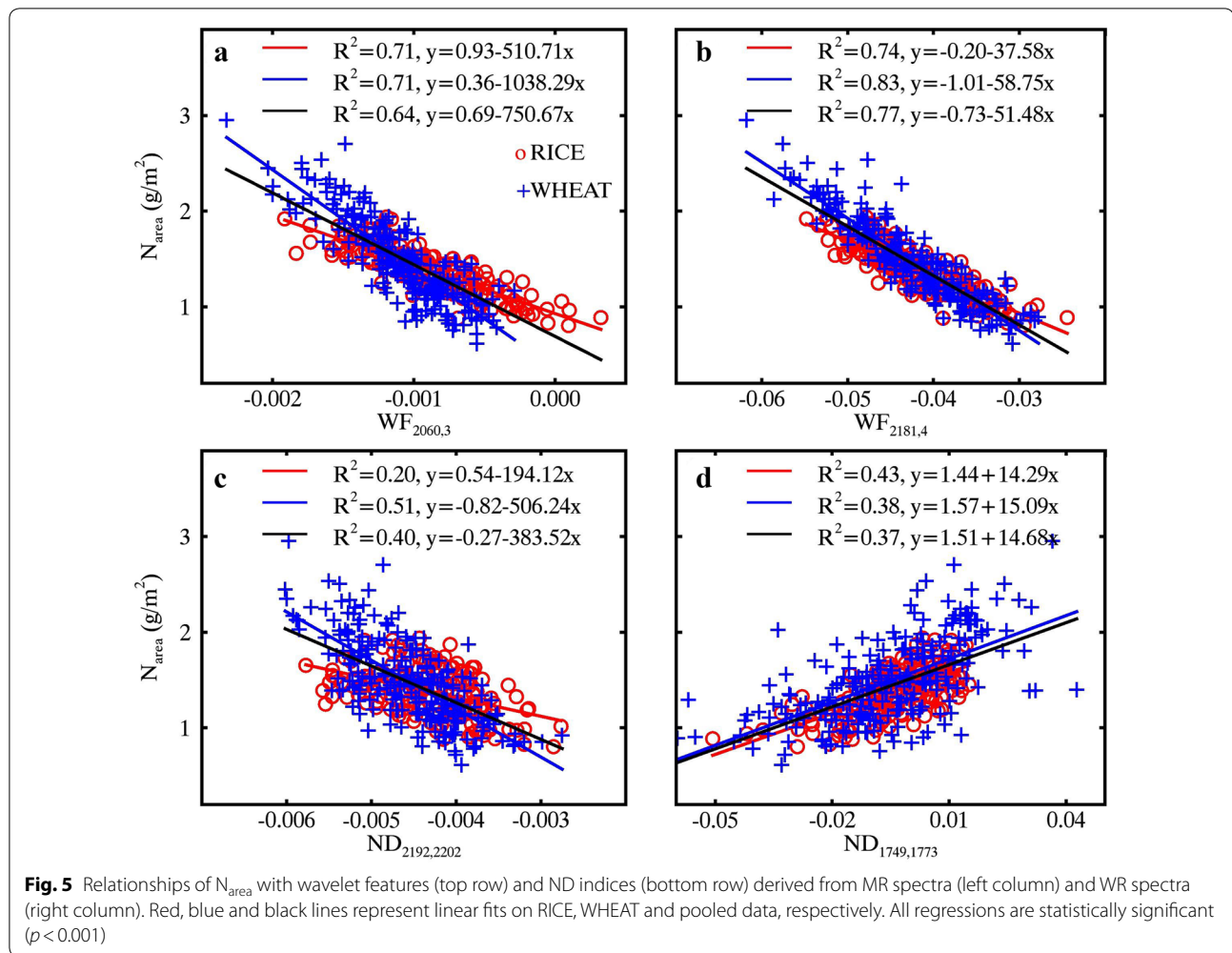
of R^2 values by 0.23 and 0.18 for RICE and WHEAT datasets, respectively.

Relationships of N_{mass} with optimal wavelet features and ND indices

The highest R^2 between N_{mass} and wavelet features for pooled data were 0.71 and 0.68 for MR and WR spectra, respectively (Fig. 7a, b). The three feature regions determined for MR spectra were centered at 2050 nm, 2120 nm and 2170 nm, all of which corresponded to the absorption wavelengths of protein and nitrogen [10] but showed an offset about 10 nm. As for the WR spectra, only one feature region sensitive to N_{mass} was found and its center wavelength (2110 nm) was 20 nm off the known absorption feature.

The R^2 contour maps for the relationships between N_{mass} and ND indices are shown in Fig. 7c, d for MR (maximum $R^2 = 0.32$) and WR (maximum $R^2 = 0.37$) spectra, respectively. Regardless of MR and WR spectra, wavelength regions sensitive to N_{area} were dispersed on the contour map and a majority of them were close to the diagonal line.

Figure 8 shows the linear relationships of N_{mass} with four optimal spectral features ($\text{WF}_{2053,3}$, $\text{WF}_{2113,4}$,



$ND_{2281,2291}$ and $ND_{2147,2161}$) determined from the top 1% regions in Fig. 7. With the WR technique, the $N_{\text{mass}} \sim WF$ correlation exhibited a decrease over that from the MR spectra for the individual datasets and for the pooled data (Fig. 8a, b). $N_{\text{area}} \sim ND$ correlations increased from MR spectra to WR spectra but the R^2 values were still substantially lower than those for the $N_{\text{mass}} \sim WF$ correlations (Fig. 8c, d).

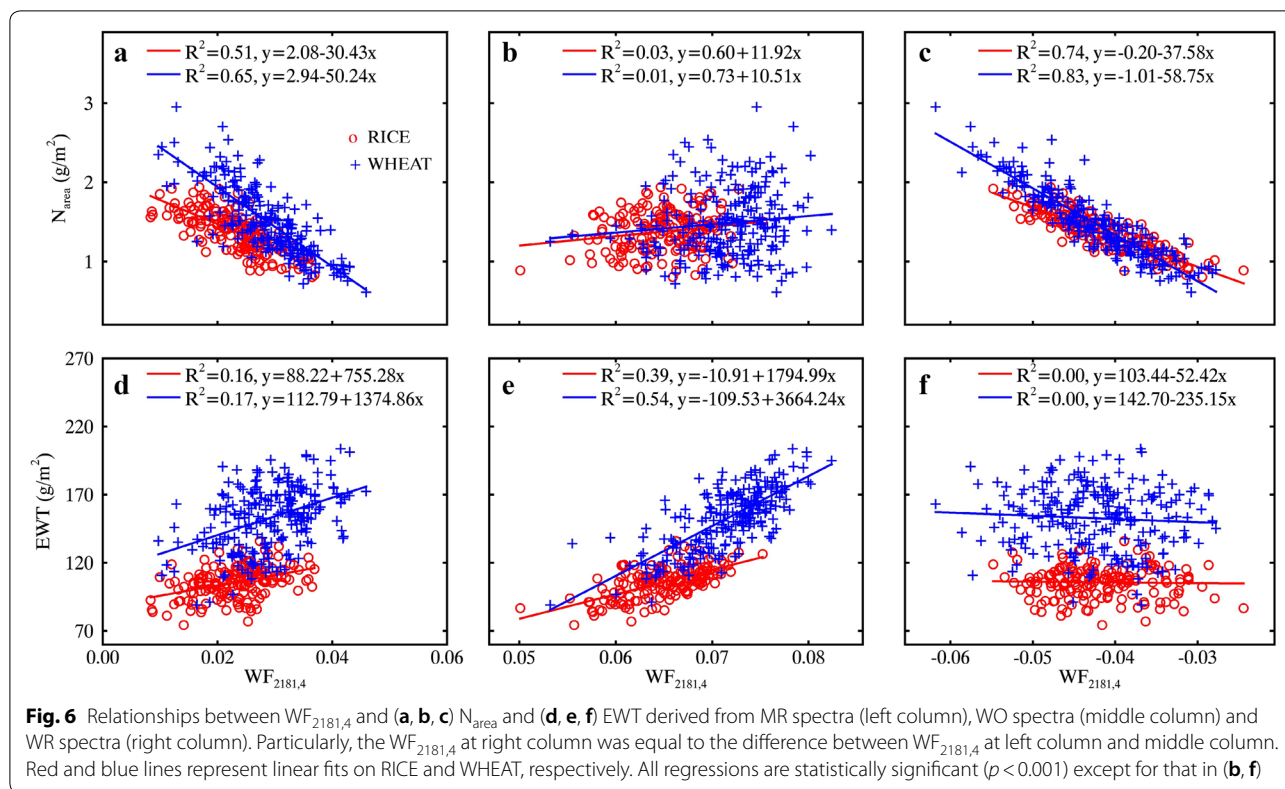
Relationships of LMA with optimal wavelet features and ND indices

The highest R^2 between LMA and wavelet features for pooled data were 0.62 and 0.71 for MR and WR spectra, respectively (Fig. 9a, b). The two feature regions determined for MR spectra were centered at 2210 nm and 2270 nm. The former did not coincide with any known absorption feature but the latter matched up with the absorption features of cellulose, sugar and starch [10]. After the water removal process, a similar feature region was found with longer wavelengths and a narrower range

of scales. In addition, a new feature region with shorter wavelengths was found and exhibited a wavelength centered at 1580 nm that was associated with the absorptions of starch and sugar [10].

The R^2 contour maps for the correlations between LMA and ND indices for pooled data are shown in Fig. 9c, d for MR (maximum $R^2 = 0.60$) and WR (maximum $R^2 = 0.60$) spectra, respectively. Three wavelength regions most sensitive to LMA for the MR spectra were represented by center wavelengths λ_1 at 1450 nm, 1870 nm and 2230 nm, and λ_2 at 2240 nm. As for the WR spectra, the most significant wavelength combinations were located in the regions with both λ_1 and λ_2 shorter than 1800 nm. The ND index regions representing the top 1% R^2 with LMA were generally more scattered than the wavelet feature regions, and therefore more difficult to relate with major absorption features documented in the literature.

Figure 10 shows the relationships of LMA with four optimal spectral features ($WF_{2273,4}$, $WF_{1578,4}$, $ND_{2236,2247}$



and $ND_{1716,1727}$) determined as the best candidates from the top 1% correlations in Fig. 9. These relationships were linear for WHEAT, RICE and pooled data. With the WR technique, the LMA ~ WF correlation exhibited a pronounced improvement (R^2 increased from 0.45 to 0.66) over that from the MR spectra for the RICE dataset but did not change for WHEAT (Fig. 10a, b). This caused an increase of R^2 from 0.62 to 0.71 for the pooled data. For LMA ~ ND correlations, the combination of a decrease for RICE and an increase for WHEAT led to no change in R^2 for the pooled data (Fig. 10c, d).

Assessment for the estimations of LMA, N_{area} and N_{mass}

Table 3 shows the estimation accuracy for LMA, N_{area} and N_{mass} over the pooled data with the 10-fold cross-validation. In particular, N_{mass} was estimated with the determined wavelet features as shown in Fig. 10 (the direct way) and with the ratio of estimated N_{mass} and LMA values (the indirect way). In all cases, wavelet features outperformed ND indices substantially in the estimations of N_{area} and N_{mass} for both MR and WR spectra. For the estimation of N_{area} , the accuracy with the best correlated wavelet features was improved from MR spectra ($R^2=0.64$ and $RMSE=0.22$ g/m²) to WR spectra ($R^2=0.77$ and $RMSE=0.18$ g/m²), but that with ND indices even decreased slightly. For the direct

estimation of N_{mass} with the best correlated wavelet features and ND indices, there was no significant difference in estimation accuracy before and after the application of water removal. As the ratio of wavelet-derived N_{area} and LMA, N_{mass} could be better estimated from WR spectra ($R^2=0.82$ and $RMSE=0.32\%$) than from MR spectra ($R^2=0.71$ and $RMSE=0.40\%$). As can be seen in Fig. 11, the scatter plot of estimated versus measured N_{mass} is closer to the 1:1 line, especially for samples in the WHEAT dataset. This indirect estimation of N_{mass} from WR spectra was also more accurate than the direct estimation. The improvement in the indirect estimation of N_{mass} from MR spectra to WR spectra was not seen for ND indices. The performance of these wavelet features derived from the SWIR region ($R^2=0.77$ and 0.82 for N_{area} and N_{mass} , respectively) was better than CI_{red} edge from the VNIR region ($R^2=0.76$ and 0.71 for N_{area} and N_{mass} , respectively), especially for N_{mass} .

In addition to the 10-fold cross validation, the LOO validation accuracies are listed in Table 4. Generally, the estimation accuracy from the LOO process was lower. However, both wavelet features and ND indices derived from the WR spectra still produced higher R^2 for LMA, N_{area} and the ratio of N_{area} to LMA (N_{mass} in an indirect way) than those from the MR spectra. As for N_{mass} estimation, the indirect way with wavelet features ($R^2=0.79$

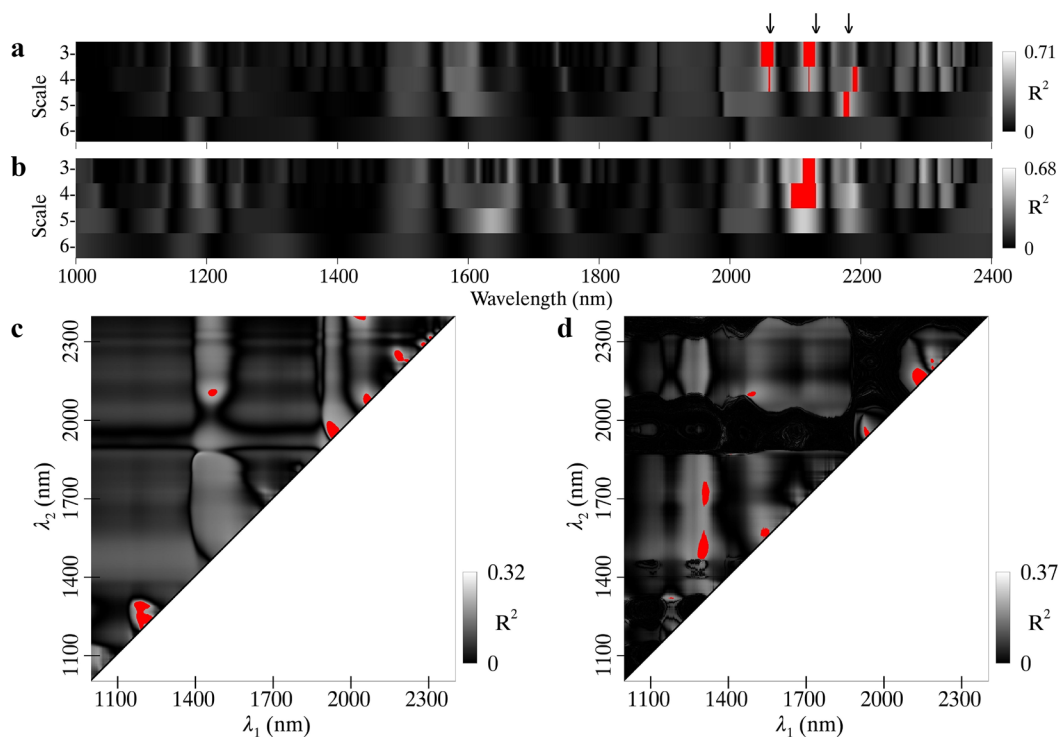


Fig. 7 Coefficients of determination (R^2) for the relationships of N_{mass} with **a** MR spectra and **b** WR spectra derived wavelet features and **c** MR spectra and **d** WR spectra derived ND indices. Regions highlighted in red on the scalograms represent the top 1% R^2 for wavelet features or ND indices. The downward arrows on the top of the figure indicate the wavelength locations of absorption features with both 2060 nm and 2180 nm for protein and nitrogen and 2130 nm for protein

and $RMSE=0.34\%$) still worked better than the direct way ($R^2=0.61$ and $RMSE=0.46\%$).

Discussion

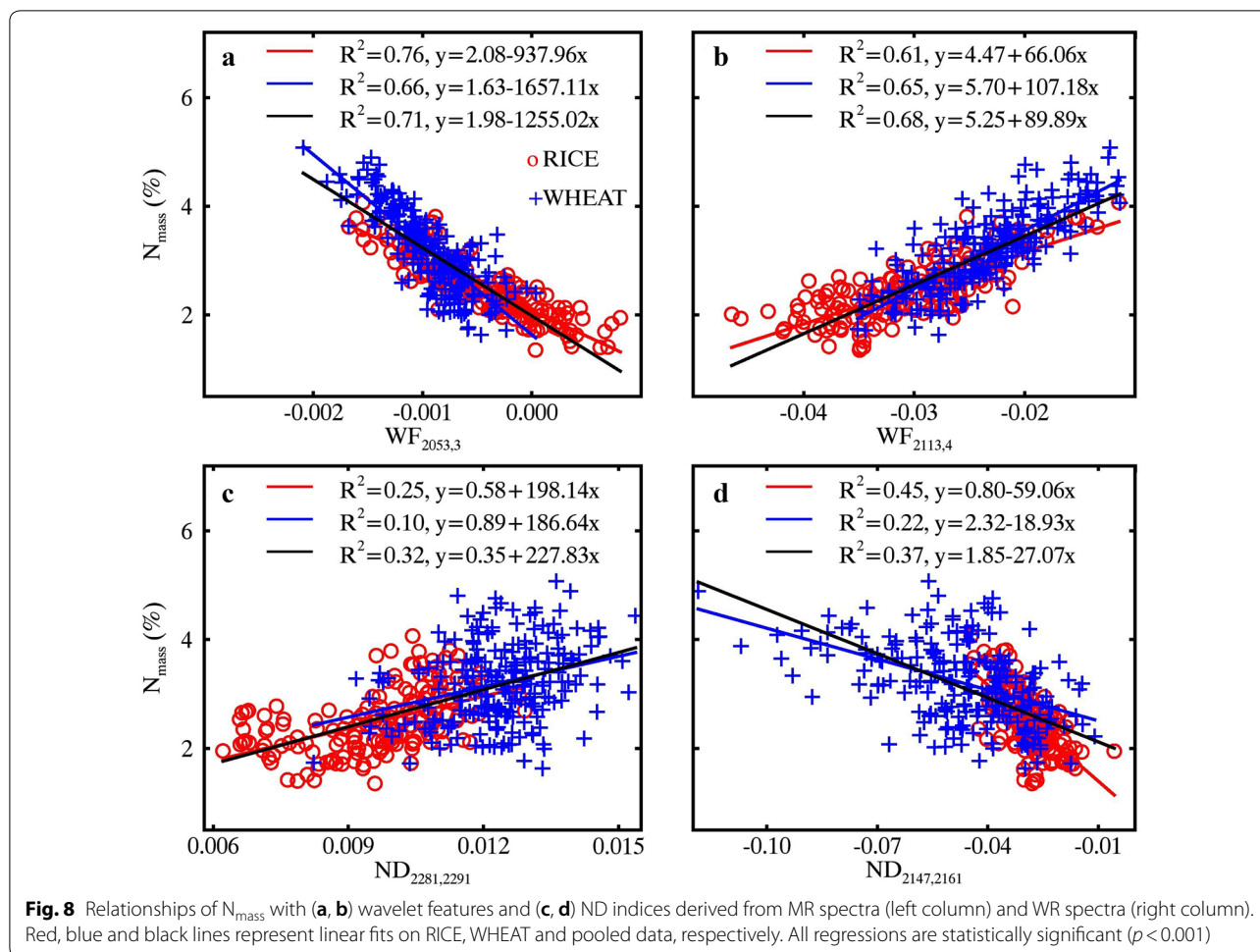
Difference in sensitive spectral features derived from MR and WR spectra

When comparing the spectral features specifically sensitive to LMA, N_{area} and N_{mass} , we found that the difference in wavelet feature was less significant between MR and WR spectra than that in ND index. The highlighted feature regions for ND indices were generally more dispersed across the scalograms, which complicated the selection of the best ND indices and was also seen in the studies by le Maire et al. [51] and Wang et al. [53]. Two dispersed index regions with even similar R^2 values would lead to substantially different wavelength combinations, not to mention many of them as highlighted in Figs. 4, 7 and 9. The sensitive wavelet features were mostly concentrated in no more than two feature regions, from which it would be easier to select the best wavelet features in less than three scales and wavelength ranges. The mismatch of the wavelengths selected by ND indices with the absorption centers suggested the weakness of ND indices and the strength of wavelet features in capturing the

changes in nitrogen absorption features. This might be attributed to the different calculation principles for ND indices [54] and wavelet features [33]. The former was determined by the reflectance values at two wavelengths and was sensitive to the amplitude of reflectance spectra, but the latter mainly captured the shape of a spectral segment [35]. Before and after water removal, the amplitude of spectra changed but the shapes of absorption features retained and became even more obvious. Therefore, common sensitive wavelet features can be observed between MR and WR spectra, but not for ND indices.

Necessity of removing the effect of water absorption with CWA

The absorption of leaf water is strong in the SWIR region [19, 42, 48], hence the subtle absorption features of such dry matter constituents as protein and nitrogen are masked in the reflectance spectra of fresh leaves (Fig. 1). Many studies have shown that this masking effect reduces the estimation accuracy of these chemical constituents [14, 15, 55]. This is supported by our findings that the estimations of LMA and N_{area} were significantly improved through the wavelet-based method after the effect of water absorption had been removed.



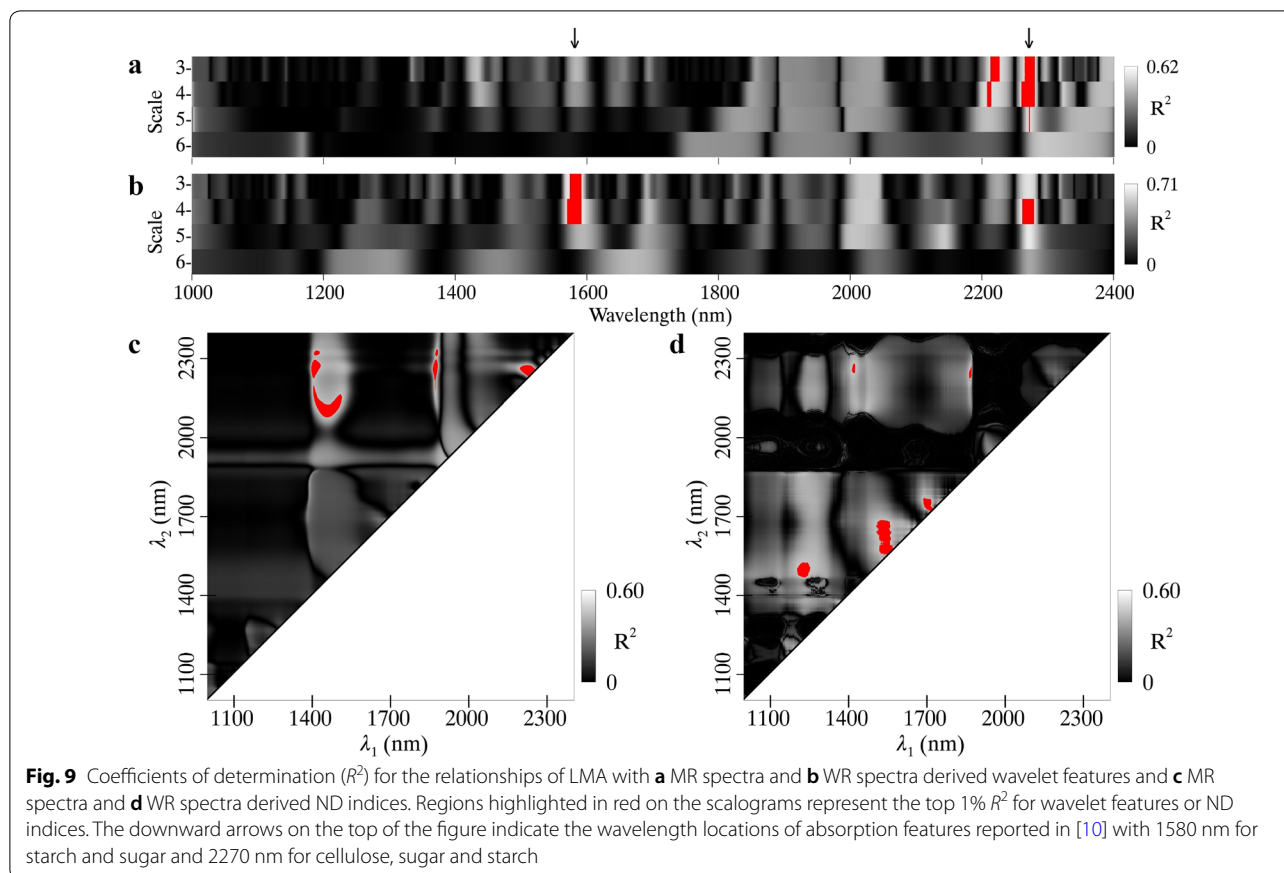
Particularly, Kokaly and Clark [11] modeled the estimation accuracy of leaf chemistry under eight levels of water content in percentage of fresh weight and found that the accuracy decreased rapidly when leaf water content exceeded 30%. To date, the effect of water absorption on ND indices and wavelet features is still poorly understood. Here we present an exploration of their performance for N_{area} estimation under different water content levels in EWT. To avoid the confounding factors of leaf surface property and internal structure, the WHEAT dataset with a wider range of EWT was selected to generate three subsets with EWT levels of 130–140 g/m² (W1), 150–160 g/m² (W2), and 170–180 g/m² (W3).

The optimal index $ND_{2192,2202}$ was well correlated to N_{area} only at the lowest EWT level (Fig. 12a), but $WF_{2060,3}$ bore stronger relationships with N_{area} at every EWT level (Fig. 12b). In contrast to the close relationships of N_{area} and $ND_{2192,2202}$ under three EWT levels, the $N_{\text{area}} \sim WF_{2060,3}$ relationship was sensitive to EWT level.

This is probably because $WF_{2060,3}$ made use of the spectral information in the 2020–2100 nm range [33], which is the right wing of the water absorption valley centered at 1950 nm [19] and whose shape changes as a function of water content [11]. In spite of the tight $N_{\text{area}} \sim WF_{2060,3}$ relationships, the sensitivity of $WF_{2060,3}$ to EWT reduced the model transferability, e.g., from subset W1 to subset W3. Therefore, it was essential to remove water absorption for building a robust model that can be directly applied to datasets with different EWT levels.

How did the water removal technique improve the estimations of N_{area} and N_{mass} ?

The accuracies of LMA and N_{area} estimations using wavelet features were improved when the effect of water absorption was removed with the water removal technique. Some recent studies used WR spectra through spectral indices or partial least squares regression to improve the estimations of LMA and N_{mass} [15,



28, 29, 56], but they did not provide explanations as to why those improvements were obtained. This study reported the application of CWA to the WR spectra for estimating foliar traits by removing the effect of water absorption and enhancing the absorption features of foliar traits. The better $WF_{2181,4} \sim N_{\text{area}}$ relationships in the WR spectra than in the MR spectra (Fig. 6a, c) could be explained by the information on EWT in $WF_{2181,4}$ which was derived from the MR spectra (Fig. 6d) but not from the WR spectra (Fig. 6f).

With regard to the water removal technique for N_{mass} , the wavelet-based direct way did not improve the estimation accuracy but the indirect way exhibited a significant improvement (Table 3). This could be attributed to the improved estimations of both LMA and N_{area} . Kokaly and Clark [11] suggested any remote sensing algorithm for retrieving foliar chemicals from fresh leaf spectra should remove the influence of water absorption. Wang et al. [56] used the normalized dry matter index (NDMI) from WR spectra to estimate the LMA but only obtained slight improvement over MR spectra as for the use of ND indices in this study (Fig. 10c, d). The wavelet features determined for LMA estimation by Cheng et al. [35] from a wide range of plant species

performed better than NDMI, but they might still suffer from the effect of water absorption and be hard to be applied to other datasets with different water content levels. Therefore, the difficulty in LMA estimation hampered the indirect estimation of N_{mass} due to the dominant role of LMA as the denominator in Eq. (9). For the first time, this study obtained improved estimations for LMA as well as N_{area} by applying CWA to the WR spectra and eventually increased the estimation accuracy for N_{mass} in the indirect way.

The close fits of data points to the linear regression and the 1:1 lines in Figs. 5 and 11 suggested the best models worked well for the whole ranges of N_{area} and N_{mass} . Although the rice and wheat plants in this experiment were deliberately treated in the field, the data were collected in natural light conditions and the models should be applicable to plants growing in the same condition and without any N treatment. To take advantage of the proposed strategy of combining CWA and water removal, it is imperative to have a dataset exhibiting weak covariance of EWT with LMA and N_{area} . Generally, this applies to leaf level but may not apply to canopy level where the traits are determined as the products of leaf traits and leaf area index and high correlations often exist

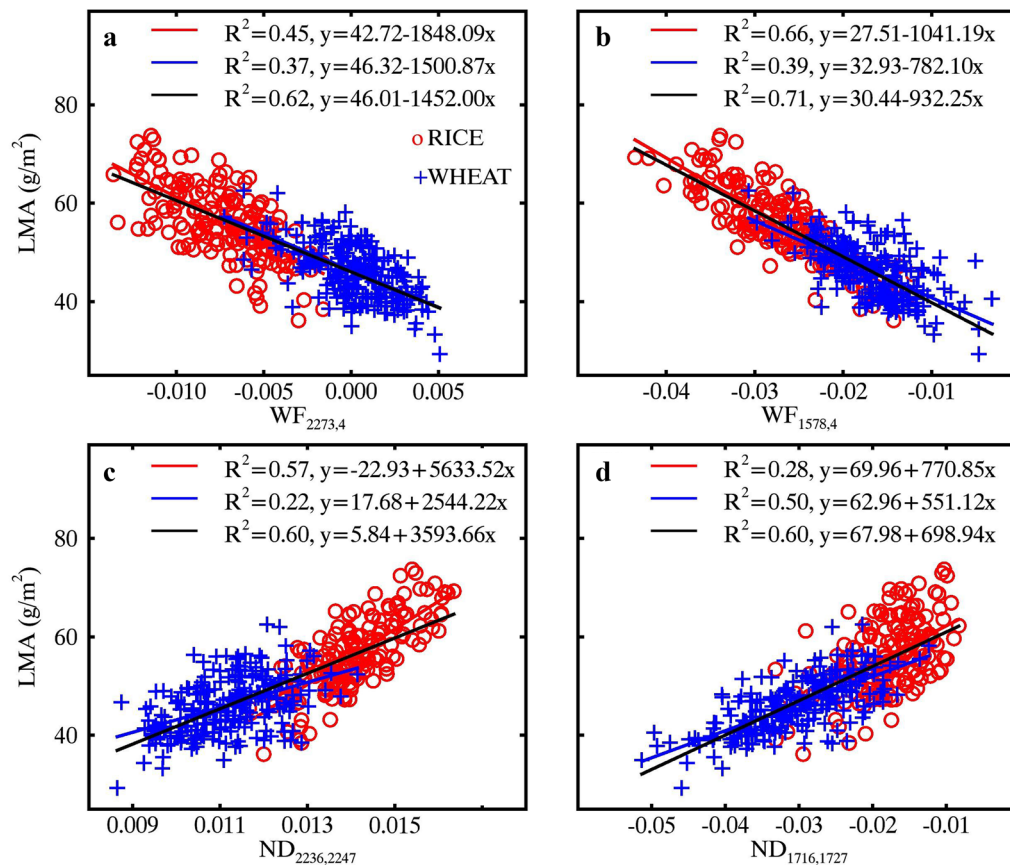


Fig. 10 Relationships of LMA with (a, b) wavelet features and (c, d) ND indices derived from MR spectra (left column) and WR spectra (right column). Red, blue and black lines represent linear fits on RICE, WHEAT and pooled data, respectively. All regressions are statistically significant ($p < 0.001$)

Table 3 Assessment of LMA, N_{area} and N_{mass} estimations generated with the direct and indirect ways

Dependent variables	Explanatory variables	MR spectra		Explanatory variables	WR spectra	
		R^2	RMSE		R^2	RMSE
LMA	WF _{2273,4}	0.62	4.95	WF _{1578,4}	0.70	4.38
	ND _{2236,2247}	0.60	5.12	ND _{1716,1727}	0.59	5.13
N_{area}	WF _{2060,3}	0.64	0.22	WF _{2181,4}	0.77	0.18
	ND _{2192,2202}	0.39	0.29	ND _{1749,1773}	0.36	0.30
N_{mass} -direct	WF _{2053,3}	0.71	0.40	WF _{2113,4}	0.68	0.42
	ND _{2283,2286}	0.31	0.62	ND _{2153,2154}	0.36	0.60
N_{mass} -indirect	WF _{2273,4} , WF _{2060,3}	0.71	0.40	WF _{1578,4} , WF _{2181,4}	0.82	0.32
	ND _{2236,2247} , ND _{2192,2202}	0.31	0.62	ND _{1716,1727} , ND _{1749,1773}	0.33	0.61

The R^2 and RMSE values were obtained using a 10-fold cross-validation

Note the units of RMSE were g/m^2 for LMA and N_{area} and % for N_{mass}

between EWT_{canopy} and other traits. In that case, it is difficult to find a wavelet feature insensitive to EWT_{canopy} but sensitive to another canopy trait. Therefore, the water removal procedure would have limited contribution to

the improvement of N_{mass} or N_{area} estimation. To test the suitability of our strategy at canopy level, one would need to avoid crop datasets with high correlations of canopy EWT and nitrogen related traits. Alternatively,

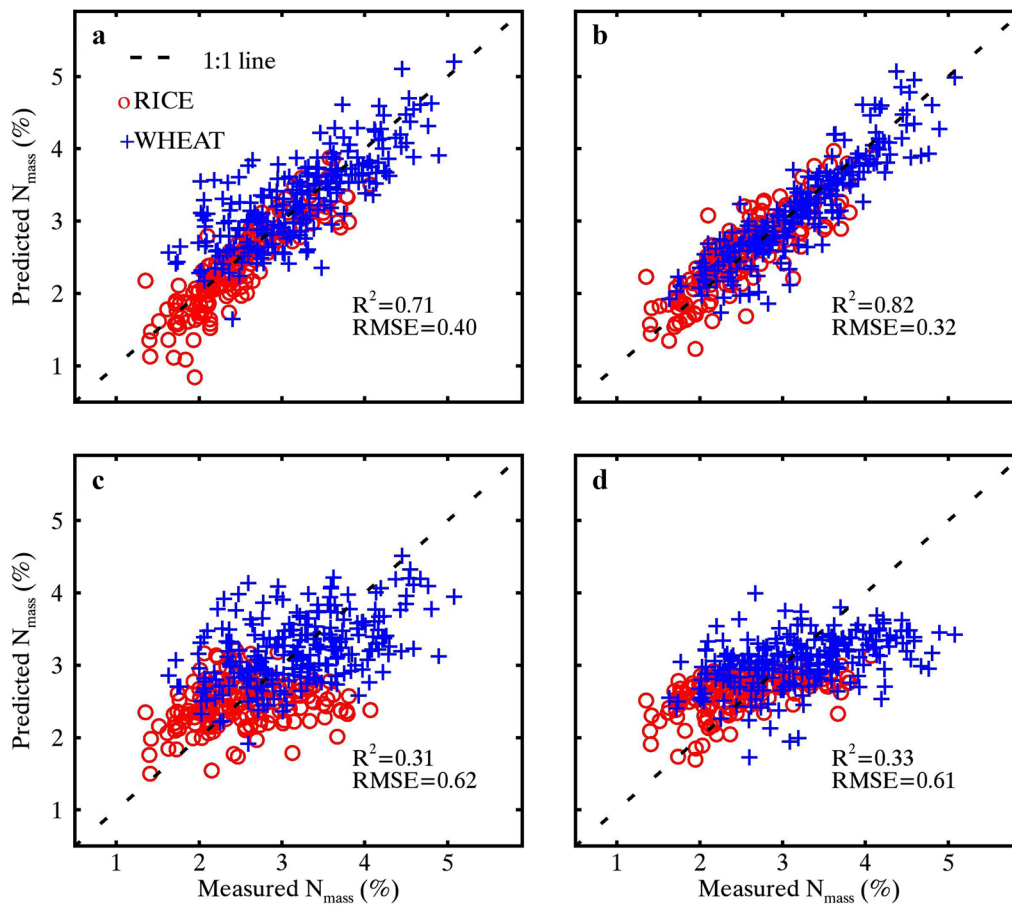


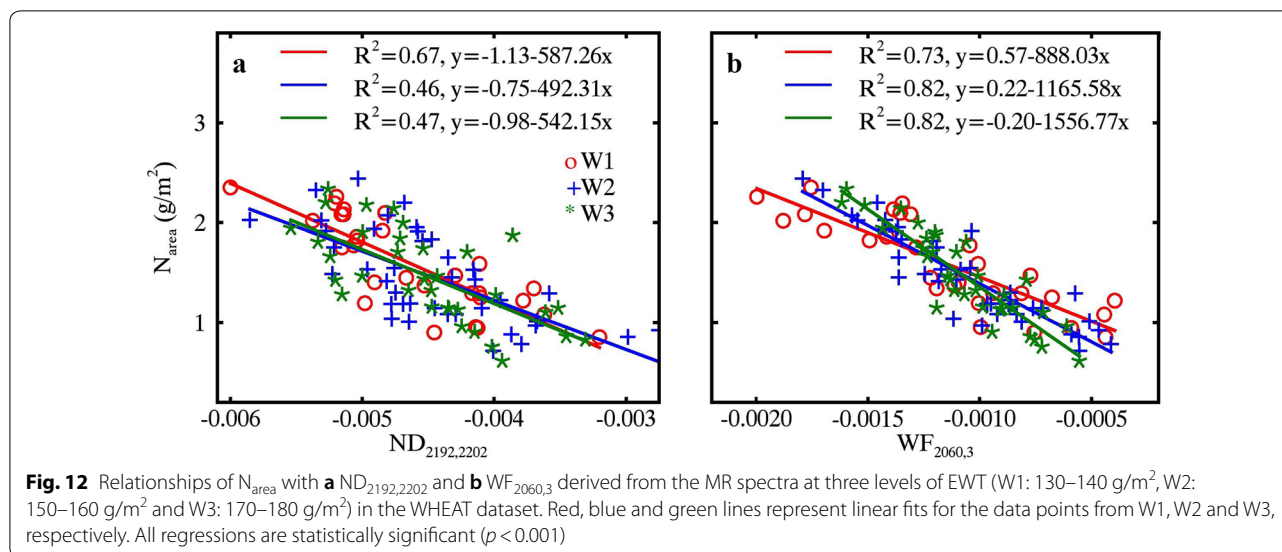
Fig. 11 The scatter plots of measured and estimated N_{mass} derived as the ratio of LMA and N_{area} estimated from (a, b) wavelet features and (c, d) ND indices from MR spectra (leaf column) and WR spectra (right column). The LMA and N_{area} estimations were generated from the 10-fold cross validation process

Table 4 Assessment of LMA, N_{area} and N_{mass} estimations generated with the direct ways and the indirect way

Dependent variables	Explanatory variables	MR spectra		Explanatory variables	WR spectra	
		R^2	RMSE		R^2	RMSE
LMA	WF _{2273,4}	0.60	5.09	WF _{1578,4}	0.63	4.88
	ND _{2236,2247}	0.46	5.93	ND _{1716,1727}	0.50	5.71
N_{area}	WF _{2060,3}	0.56	0.25	WF _{2181,4}	0.72	0.19
	ND _{2192,2202}	0.25	0.32	ND _{1749,1773}	0.32	0.31
N_{mass} -direct	WF _{2053,3}	0.65	0.44	WF _{2113,4}	0.64	0.44
	ND _{2283,2286}	0.29	0.63	ND _{2153,2154}	0.14	0.69
N_{mass} -indirect	WF _{2273,4} , WF _{2060,3}	0.61	0.46	WF _{1578,4} , WF _{2181,4}	0.79	0.34
	ND _{2236,2247} , ND _{2192,2202}	0.01	0.74	ND _{1716,1727} , ND _{1749,1773}	0.26	0.64

The R^2 and RMSE values were obtained using a leave-one-out validation

Not the units of RMSE were g/m^2 for LMA and N_{area} and % for N_{mass}



the strategy could be applied to natural plant types such as trees [15] and grasses [28, 29]. Although this research was conducted at leaf level, it represented a significant advance in the mechanistic understanding of water removal technique for removing water signals and enhancing nitrogen and dry matter signals in fresh leaves. The leaf-level analysis concentrated on chemical signals and laid the foundation of water removal application for scaling up to the canopy level. To better understand the mechanism at canopy level, the strategy should be examined in detail by considering other external factors such as canopy structure [57, 58], soil background [59, 60], solar zenith angle [61, 62], view zenith angle [7, 43], and diffuse/direct light condition [63]. Finally, the signal to noise ratio should also be accounted for in case the nitrogen-related second overtones are suppressed by the spectral noise from instruments and atmospheric interference [10].

Physical interpretations of determined wavelet features

Low-scale wavelet features are able to capture the local shape of absorption features associated with biochemical parameters while high-scale ones characterize the broad continuum in leaf reflectance spectra [30, 33, 35]. The matching wavelet features could be found for N_{area} and LMA but not for N_{mass} , which might suggest that the area-based expressions (N_{area} and LMA) were better representations of the interaction between matter and light per unit surface area [41].

The absorption feature at 2270 nm caused by bending and stretching of chemical bonds in cellulose, sugar and

starch does not obviously appear in the reflectance spectra (Figs. 1, 13a), but it is well represented in the wavelet power spectra of MR-derived $WF_{2273,4}$ (Fig. 13c). Since LMA is equivalent to the dry matter content by definition, we could relate spectral variation to dry matter absorption from the perspective of spectroscopy of foliar chemistry as done in previous studies [35]. The amplitude of reflectance in the 2200–2350 nm range increases with LMA (Fig. 13a), which contradicts the common sense that the reflectance should decrease with dry matter absorption. This might be attributed to the leaf internal scattering as well as the confounding effect of water absorption on leaf reflectance, which makes it difficult to derive the LMA with traditional techniques, such as ND index [55, 64]. In contrast, the depth of absorption center at 2270 nm in the wavelet power spectra follows the trend of LMA variation (Fig. 13c). Moreover, the absorption feature at 1580 nm by starch and cellulose are invisible in the MR spectra (Fig. 1), but it appears clearly in the WR spectra after water removal (Fig. 13b). The change in the amplitude at WR spectra-derived $WF_{2273,4}$ in the wavelet power spectra matches with the variation in LMA (Fig. 13d).

Similar to dry matter, the absorption of nitrogen at 2060 nm and 2180 nm [10] could be well characterized by the optimal N_{area} sensitive wavelet features $WF_{2060,3}$ (MR spectra) and $WF_{2180,4}$ (WR spectra) (Fig. 14). The close matches in wavelength between the determined wavelet features and the documented absorption centers documented [10–12] convince us of the physical interpretations underlying the empirical WF - N_{area} relationships.

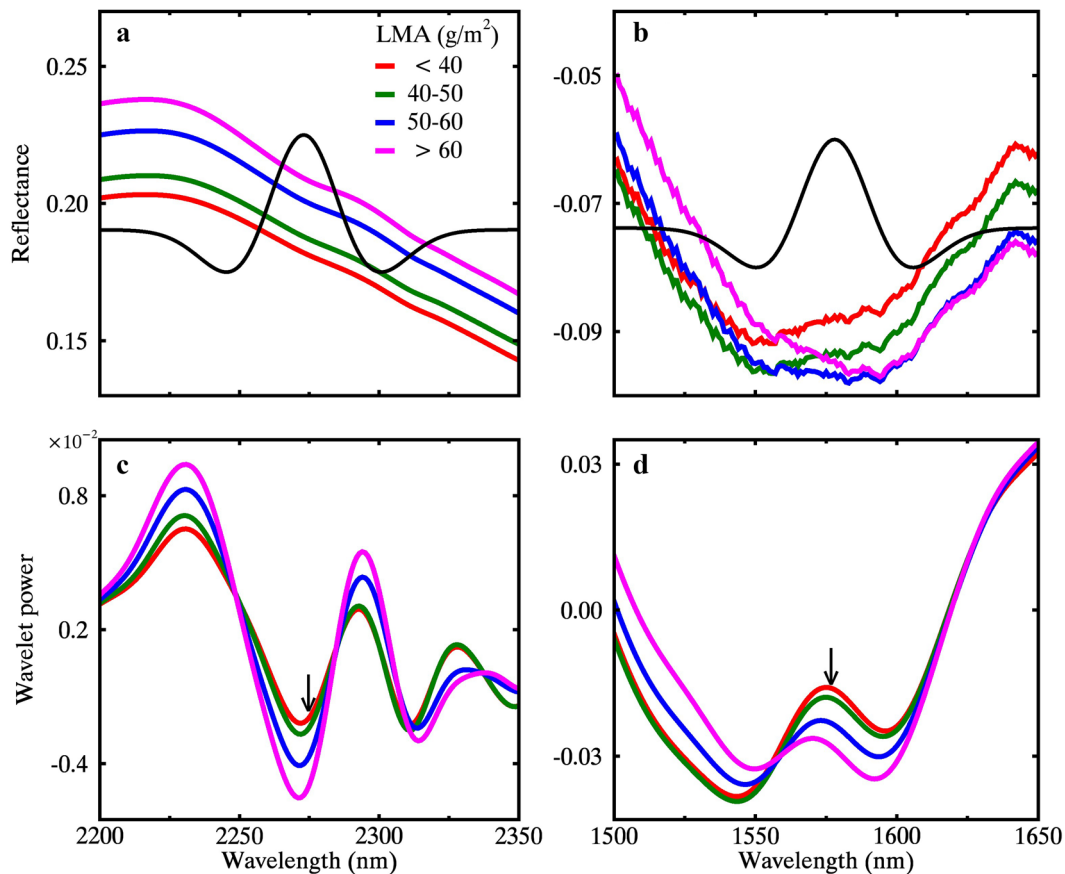


Fig. 13 Effect of LMA on spectral variation between 2200 and 2350 nm (leaf panel) and 1500–1650 nm (right panel). **a, c** Represent the MR spectra and the corresponding wavelet power spectra at scale 4. **b, d** Represent the WR spectra and the corresponding wavelet power at scale 4. The location and width of the Gaussian-shaped wavelets for the best LMA-sensitive wavelet features $WF_{2273,3}$ and $WF_{1578,4}$ from MR and WR spectra are presented in **(a)** and **(b)**, respectively. The black arrows in **(c)** and **(d)** point to the wavelength locations of $WF_{2273,3}$ and $WF_{1578,4}$, respectively

Conclusions

The enhancement of absorption features of nitrogen in the SWIR region would help improve the spectroscopic estimation of leaf nitrogen content from the SWIR reflectance spectra of fresh leaves. This study evaluated the applicability of the water-removal technique for increasing the estimation accuracies of LMA, N_{area} and N_{mass} with CWA and spectral index methods. After removing water absorption, the estimation accuracies for LMA and N_{area} were significantly improved with the use of individual wavelet features but not with ND indices. The two wavelet features $WF_{1578,4}$ and $WF_{2181,4}$ derived from the WR spectra produced the best estimations for LMA ($R^2=0.70$, $\text{RMSE}=4.38 \text{ g/m}^2$) and N_{area} ($R^2=0.77$, $\text{RMSE}=0.18 \text{ g/m}^2$), respectively. Compared to that from the MR spectra, the estimation of N_{mass} from the WR spectra only changed marginally with individual wavelet features (MR spectra: $R^2=0.71$, $\text{RMSE}=0.40\%$; WR spectra: $R^2=0.68$, $\text{RMSE}=0.42\%$). However, a notable

improvement was obtained when indirectly deriving N_{mass} as a ratio of N_{area} to LMA (MR spectra: $R^2=0.71$, $\text{RMSE}=0.40\%$; WR spectra: $R^2=0.82$, $\text{RMSE}=0.32\%$). Generally, the determined wavelet features performed remarkably better than the optimized ND indices for the estimations of all the three traits.

The results demonstrated the feasibility of improving the estimation of N_{area} and N_{mass} from the SWIR reflectance spectra by applying CWA to WR spectra. This procedure could suppress the effect of water absorption and enhance the absorption features of foliar chemicals. The characterization of nitrogen absorption variation in the SWIR region with wavelet features offers physical interpretations for the direct detection of leaf nitrogen content and complements the indirect detection of nitrogen content via the use of chlorophyll absorption bands in the red edge region. The integration of CWA and water removal provides a new insight

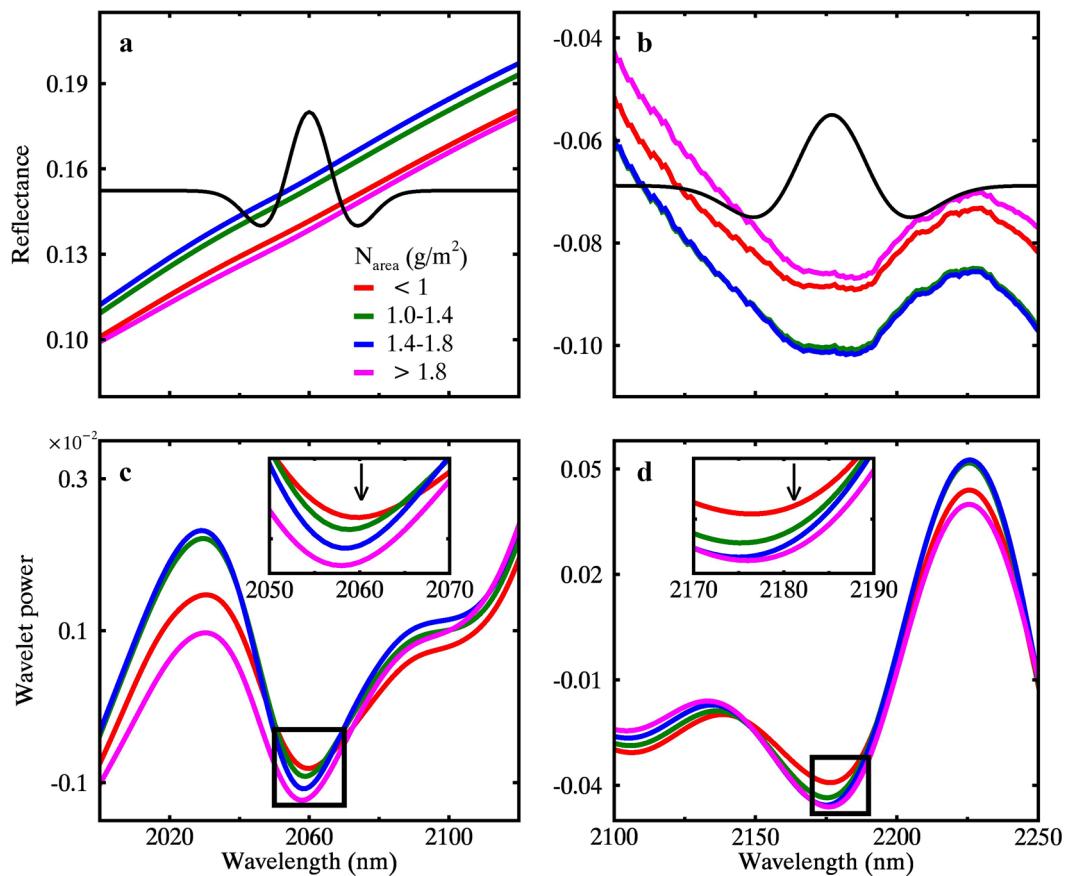


Fig. 14 Effect of N_{area} on spectral variation between 2000 and 2120 nm (leaf panel) and 2100–2250 nm (right panel). **a, c** Represent the MR spectra and the corresponding wavelet power spectra at scale 3. **b, d** Represent the WR spectra and the corresponding wavelet power at scale 4. The location and width of the Gaussian-shaped wavelets for the best N_{area} -sensitive wavelet features $WF_{2060,3}$ and $WF_{2181,4}$ from MR and WR spectra are presented in **(a)** and **(b)**, respectively. The black arrows in **(c)** and **(d)** point to the wavelength location of the $WF_{2060,3}$ and $WF_{2181,4}$, respectively

for better understanding spectral responses to the variation in LMA and leaf nitrogen content and has the potential for applications to other crops and different vegetation types. The technique developed in this study could be useful for high throughput phenotyping of leaf nitrogen related traits and the non-destructive detection of nitrogen stress in crops.

Abbreviations

N_{area} : area-based leaf nitrogen content; N_{mass} : mass-based leaf nitrogen content; LMA: leaf mass per area; WR: water-removed; CWA: continuous wavelet analysis; WF: wavelet feature; ND: normalized difference; VNIR: visible and near infrared; SWIR: shortwave infrared; R^2 : coefficient of determination; RMSE: root mean square error.

Authors' contributions

DL and TC designed the experiment. DL, XW, HZ, and KZ performed the experiment. DL, XY, YT, YZ, WC, and TC performed the overall data analysis. DL and TC drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Min Jia, Ning Lu, Xiao Zhang, Chen Zhou, Yong Liu, Ni Wang, Xue Ma, Xiaoqing Xu and Chaojie Niu for their help in the data collection. The authors also thank two anonymous reviewers for their detailed suggestions for improving the manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This work was supported by the National Key R&D Program (2016YFD0300601), the Fundamental Research Funds for the Central Universities (KYLH201603), the National Natural Science Foundation of China (31470084, 31725020, 31671582), Jiangsu Distinguished Professor Program,

and the Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 1 April 2018 Accepted: 22 August 2018

Published online: 29 August 2018

References

- Seck PA, Diagne A, Mohanty S, Wopereis MCS. Crops that feed the world 7: rice. *Food Secur.* 2012;4(1):7–24.
- Shiferaw B, Smale M, Braun HJ, Duveiller E, Reynolds M, Muricho G. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Secur.* 2013;5(3):291–317.
- Evans JR. Photosynthesis and nitrogen relationships in leaves of C_3 plants. *Oecologia.* 1989;78(1):9–19.
- Ladha JK, Krupnik TJ, Six J, Kessel CV, Pathak H. Efficiency of fertilizer nitrogen in cereal production: retrospects and prospects. *Adv Agron.* 2005;87:85–156.
- Gamon JA, Surfus JS. Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* 2010;143(1):105–17.
- Haboudane D, Miller JR, Tremblay N, Zarco-Tejada PJ, Dextraze L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens Environ.* 2002;81(2–3):416–26.
- He L, Song X, Feng W, Guo B-B, Zhang Y-S, Wang Y-H, Wang C-Y, Guo T-C. Improved remote sensing of leaf nitrogen concentration in winter wheat using multi-angular hyperspectral data. *Remote Sens Environ.* 2016;174:122–33.
- Cho MA, Skidmore AK. A new technique for extracting the red edge position from hyperspectral data: the linear extrapolation method. *Remote Sens Environ.* 2006;101(2):181–93.
- Huang Z, Turner BJ, Dury SJ, Wallis IR, Foley WJ. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens Environ.* 2004;93(1–2):18–29.
- Curran PJ. Remote sensing of foliar chemistry. *Remote Sens Environ.* 1989;30(3):271–8.
- Kokaly RF, Clark RN. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens Environ.* 1999;67(3):267–87.
- Kokaly RF. Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. *Remote Sens Environ.* 2001;75(2):153–61.
- Pacheco-Labrador J, Gonzalez-Cascon R, Pilar Martin M, Riano D. Understanding the optical responses of leaf nitrogen in Mediterranean Holm oak (*Quercus ilex*) using field spectroscopy. *Int J Appl Earth Observ Geoinf.* 2014;26:105–18.
- Yoder BJ, Pettigrew-Crosby RE. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. *Remote Sens Environ.* 1995;53(3):199–211.
- Schlerf M, Atzberger C, Hill J, Buddenbaum H, Werner W, Schüler G. Retrieval of chlorophyll and nitrogen in Norway spruce (*Picea abies* L. Karst.) using imaging spectroscopy. *Int J Appl Earth Observ Geoinf.* 2010;12(1):17–26.
- Cao Q, Miao Y, Feng G, Gao X, Li F, Liu B, Yue S, Cheng S, Ustin SL, Khosla R. Active canopy sensing of winter wheat nitrogen status: an evaluation of two sensor systems. *Comput Electron Agric.* 2015;112:54–67.
- Wang Y, Wang D, Shi P, Omasa K. Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. *Plant Methods.* 2014;10:36.
- Schlemmer M, Gitelson A, Schepers J, Ferguson R, Peng Y, Shanahan J, Rundquist D. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *Int J Appl Earth Observ Geoinf.* 2013;25:47–54.
- Jacquemoud S, Baret F. PROSPECT: a model of leaf optical properties spectra. *Remote Sens Environ.* 1990;34(2):75–91.
- le Maire G, François C, Dufrêne E. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sens Environ.* 2004;89(1):1–28.
- Lamb DW, Steyn-Ross M, Schaare P, Hanna MM, Silvester W, Steyn-Ross A. Estimating leaf nitrogen concentration in ryegrass (*Lolium* spp.) pasture using the chlorophyll red-edge: theoretical modelling and experimental observations. *Int J Remote Sens.* 2002;23(18):3619–48.
- Hansen PM, Schjoerring JK. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens Environ.* 2003;86(4):542–53.
- Feng W, Yao X, Zhu Y, Tian YC, Cao W. Monitoring leaf nitrogen status with hyperspectral reflectance in wheat. *Eur J Agron.* 2008;28(3):394–404.
- Lepine LC, Ollinger SV, Ouimette AP, Martin ME. Examining spectral reflectance features related to foliar nitrogen in forests: implications for broad-scale nitrogen mapping. *Remote Sens Environ.* 2016;173:174–86.
- Ecarnot M, Compan F, Roumet P. Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer. *Field Crops Research.* 2013;140:44–50.
- Hikosaka K, Terashima I. Nitrogen partitioning among photosynthetic components and its consequence in sun and shade plants. *Funct Ecol.* 1996;10:335–43.
- Gao BC, Goetz AFH. Extraction of dry leaf spectral features from reflectance spectra of green vegetation. *Remote Sens Environ.* 1994;47(3):369–74.
- Ramoelo A, Skidmore AK, Schlerf M, Mathieu R, Heitkonig IMA. Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations. *ISPRS J Photogramm Remote Sens.* 2011;66(4):408–17.
- Ramoelo A, Skidmore AK, Schlerf M, Heitkonig IMA, Mathieu R, Cho MA. Savanna grass nitrogen to phosphorus ratio estimation using field spectroscopy and the potential for estimation with imaging spectroscopy. *Int J Appl Earth Observ Geoinf.* 2013;23:334–43.
- Blackburn GA, Ferwerda JG. Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sens Environ.* 2008;112(4):1614–32.
- Li D, Cheng T, Zhou K, Zheng H, Yao X, Tian Y, Zhu Y, Cao W. WREP: a wavelet-based technique for extracting the red edge position from reflectance spectra for estimating leaf and canopy chlorophyll contents of cereal crops. *ISPRS J Photogramm Remote Sens.* 2017;129:103–17.
- Wang HF, Huo ZG, Zhou GS, Liao QH, Feng HK, Wu L. Estimating leaf SPAD values of freeze-damaged winter wheat using continuous wavelet analysis. *Plant Physiol Biochem.* 2016;98:39–45.
- Cheng T, Rivard B, Sánchez-Azofeifa A. Spectroscopic determination of leaf water content using continuous wavelet analysis. *Remote Sens Environ.* 2011;115(2):659–70.
- Cheng T, Riano D, Ustin SL. Detecting diurnal and seasonal variation in canopy water content of nut tree orchards from airborne imaging spectroscopy data using continuous wavelet analysis. *Remote Sens Environ.* 2014;143:39–53.
- Cheng T, Rivard B, Sánchez-Azofeifa AG, Féret J-B, Jacquemoud S, Ustin SL. Deriving leaf mass per area (LMA) from foliar reflectance across a variety of plant species using continuous wavelet analysis. *ISPRS J Photogramm Remote Sens.* 2014;87:28–38.
- Huang Y, Tian QJ, Wang L, Geng J, Lyu CG. Estimating canopy leaf area index in the late stages of wheat growth using continuous wavelet transform. *J Appl Remote Sens.* 2014;8:083517.
- Rivard B, Feng J, Gallie A, Sanchez-Azofeifa A. Continuous wavelets for the improved use of spectral libraries and hyperspectral data. *Remote Sens Environ.* 2008;112(6):2850–62.
- Houlès V, Guérif M, Mary B. Elaboration of a nitrogen nutrition indicator for winter wheat based on leaf area index and chlorophyll content for making nitrogen recommendations. *Eur J Agron.* 2007;27(1):1–11.
- Filella I, Serrano L, Serra J, Penuelas J. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Sci.* 1995;35(5):1400–5.
- Shi T, Wang J, Liu H, Wu G. Estimating leaf nitrogen concentration in heterogeneous crop plants from hyperspectral reflectance. *Int J Remote Sens.* 2015;36(18):4652–67.

41. Datt B. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a + b, and total carotenoid content in Eucalyptus leaves. *Remote Sens Environ.* 1998;66(2):111–21.
42. Féret J-B, François C, Asner GP, Gitelson AA, Martin RE, Bidel LPR, Ustin SL, le Maire G, Jacquemoud S. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens Environ.* 2008;112(6):3030–43.
43. Jay S, Maupas F, Bendoula R, Gorretta N. Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Res.* 2017;210:33–46.
44. Danson FM, Bowyer P. Estimating live fuel moisture content from remotely sensed reflectance. *Remote Sens Environ.* 2004;92(3):309–21.
45. Wang B, Chen J, Ju W, Qiu F, Zhang Q, Fang M, Chen F. Limited effects of water absorption on reducing the accuracy of leaf nitrogen estimation. *Remote Sens.* 2017;9(3):291.
46. Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JH, Diemer M. The worldwide leaf economics spectrum. *Nature.* 2004;428(6985):821.
47. Li D, Cheng T, Jia M, Zhou K, Lu N, Yao X, Tian Y, Zhu Y, Cao W. PROCWT: coupling PROSPECT with continuous wavelet transform to improve the retrieval of foliar chemistry from leaf bidirectional reflectance spectra. *Remote Sens Environ.* 2018;206:1–14.
48. Ceccato P, Gobron N, Flasse S, Pinty B, Tarantola S. Designing a spectral index to estimate vegetation water content from remote sensing data: part 1. Theoretical approach. *Remote Sens Environ.* 2002;82(2–3):188–97.
49. Tsai F, Philpot W. Derivative analysis of hyperspectral data. *Remote Sens Environ.* 1998;66(1):41–51.
50. Inoue Y, Guerif M, Baret F, Skidmore A, Gitelson A, Schlerf M, Darvishzadeh R, Olioso A. Simple and robust methods for remote sensing of canopy chlorophyll content: a comparative analysis of hyperspectral data for different types of vegetation. *Plant Cell Environ.* 2016;39(12):2609–23.
51. le Maire G, François C, Soudani K, Berveiller D, Pontailier J-Y, Bréda N, Genet H, Davi H, Dufréne E. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sens Environ.* 2008;112(10):3846–64.
52. Schlemmer MR, Francis DD, Shanahan JF, Schepers JS. Remotely measuring chlorophyll content in corn leaves with differing nitrogen levels and relative water content. *Agron J.* 2005;97(1):106–12.
53. Wang L, Qu JJ, Hao X, Hunt ER Jr. Estimating dry matter content from spectral reflectance for green leaves of different species. *Int J Remote Sens.* 2011;32(22):7097–109.
54. Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring vegetation systems in the great plains with Erts. *NASA Spec Publ.* 1974;351:309–17.
55. Curran PJ, Dungan JL, Peterson DL. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: testing the Kokaly and Clark methodologies. *Remote Sens Environ.* 2001;76(3):349–59.
56. Wang L Jr, ER H Jr, Qu JJ, Hao X, Daughtry CST. Estimating dry matter content of fresh leaves from the residuals between leaf and water reflectance. *Remote Sens Lett.* 2011;2(2):137–45.
57. Knyazikhin Y, Schull MA, Stenberg P, Mottus M, Rautiainen M, Yang Y, Marshak A, Latorre Carmona P, Kaufmann RK, Lewis P, et al. Hyperspectral remote sensing of foliar nitrogen content. *Proc Natl Acad Sci USA.* 2013;110(3):E185–92.
58. Niemann KO, Quinn G, Goodenough DG, Visintini F, Loos R. Addressing the effects of canopy structure on the remote sensing of foliar chemistry of a 3-dimensional, radiometrically porous surface. *IEEE J Sel Top Appl Earth Observ Remote Sens.* 2012;5(2):584–93.
59. Darvishzadeh R, Skidmore A, Atzberger C, van Wieren S. Estimation of vegetation LAI from hyperspectral reflectance data: effects of soil type and plant architecture. *Int J Appl Earth Observ Geoinf.* 2008;10(3):358–73.
60. Yu K, Lenz-Wiedemann V, Chen X, Bareth G. Estimating leaf chlorophyll of barley at different growth stages using spectral indices to reduce soil background and canopy structure effects. *ISPRS J Photogramm Remote Sens.* 2014;97:58–77.
61. Middleton EM. Solar zenith angle effects on vegetation indices in tallgrass prairie. *Remote Sens Environ.* 1991;38(91):45–62.
62. Brede B, Suomalainen J, Bartholomeus H, Herold M. Influence of solar zenith angle on the enhanced vegetation index of a Guyanese rainforest. *Remote Sens Lett.* 2015;6(12):972–81.
63. Ishihara M, Inoue Y, Ono K, Shimizu M, Matsuura S. The impact of sunlight conditions on the consistency of vegetation indices in croplands-effective usage of vegetation indices from continuous ground-based spectral measurements. *Remote Sens.* 2015;7(10):14079–98.
64. Grossman YL, Ustin SL, Jacquemoud S, Sanderson EW, Schmuck G, Verdebout J. Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sens Environ.* 1996;56(3):182–93.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

