Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 8 (2021) 284–298

journal homepage: www.elsevier.com/locate/inpa

Evaluation of regression algorithms for estimating leaf area index and canopy water content from water stressed rice canopy reflectance

Niranjan Panigrahi [a,b,*], *Bhabani Sankar Das* [a]

[a] *Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, West Bengal 721302, India*
[b] *School of Water, Energy and Environment, Cranfield University, Cranfield MK43 0AL, UK*

ARTICLE INFO

Article history:
Received 1 October 2019
Received in revised form
22 June 2020
Accepted 29 June 2020
Available online 6 July 2020

Keywords:
Rice canopy
Reflectance spectroscopy
Calibration
Retrieval algorithms
Partial-least-squares regression

ABSTRACT

Optical remote sensing (RS) with robust algorithms is needed for accurate assessment of crop canopy features. Despite intensive studies on algorithms, their performance using RS needs to be improved. We evaluated five different algorithms (partial-least-squares regression (PLSR), support vector regression (SVR), random forest regression (RFR), locally-weighted-PLSR ($PLSR_{LW}$) and PLSR with feature selection ($PLSR_{FS}$)) for rapid assessment of leaf area index (LAI) and canopy water content (CWC) for rice canopies using canopy reflectance spectra over visible to short-wave infrared region. Two pooled datasets of LAI (600) and CWC (480) were collected from two replicated field experiments during 2014–15 and 2015–16 rice growing season. The performance of each algorithm was evaluated using coefficient of determination (R^2). Results showed that $PLSR_{LW}$ performed more accurately than other algorithms with R^2 values 0.77 and 0.66 for LAI and CWC, respectively. We also used a bootstrapping approach to generate a kernel density estimator of root mean squared error values for each model. The results suggested that the improvement in prediction accuracy of LAI and CWC can be achieved if a suitable algorithm is selected by assigning higher weights to calibration samples, which has similar canopy structure as the test sample. Subsetting of the canopy spectral data results large error values in test dataset, therefore the use of entire season canopy spectral data should be used for model calibration.

1. Introduction

Optical remote sensing (RS) is emerging as a fast and non-destructive technique for quantifying vegetation canopy parameters [1–5]. Using RS, several vegetation biophysical and biochemical parameters may be estimated from a single spectral signature measured over the visible, near-infrared and shortwave infrared region (wavelength: 350–2500 nm) of the electromagnetic spectrum [6–7]. Fast, accurate and robust algorithms and large crop specific spectral libraries are needed for implementing RS approach [7–8]. In the last decades, several retrieval algorithms such as empirical regressions partial-least-squares regression (PLSR) [9–12]; support vector regression (SVR) [13–16]; random forest regression

* Corresponding author at: School of Water, Energy and Environment, Cranfield University, Cranfield MK43 0AL, UK.
E-mail address: Niranjan.Panigrahi@cranfield.ac.uk (N. Panigrahi).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2020.06.002

(RFR) [16–18]; and artificial neural network (ANN) [19–20]; and physical based methods such as radiative transfer models [21] have been applied to relate between reflectance spectra and in-situ vegetation data. In this study, empirical regression algorithms are focused over physical based methods because of their ill posed nature of the model inversion and less suitability for retrieval applications [5,21]. Some of the studies on empirical algorithms and their performance for prediction of vegetation canopy parameters are shown in supplementary material (Table S1). Despite using several algorithms on different vegetation parameters, the coefficients of determination (R^2) achievable are moderate to poor. Therefore, efficient algorithms that are performed well for other spectroscopic studies need to be tested for vegetation parameters.

Crop specific spectral libraries have been developed over the last decade to implement and improve the predictive power of spectral reflectance- based algorithms [22–25]. In crop specific spectral libraries, an individual crop spectral reflectance is often collected from an individual leaf or from a whole crop canopy depending on the parameter of interest. However these spectral libraries are region specific and depend on several factors such as sensors used, quality of radiation source, atmosphere, vegetation canopy, soil nutrients and water availability [26]. The prediction accuracy of a spectral model calibrated from a field campaign may be influenced by the sources of data, which may include the nature of crop and selection of sampling locations. Many crop attributes are expected to change with time and space [6]. As the crop spectral libraries are dealing with complex datasets exhibiting nonlinear attribute relationships, the algorithms based on linear transformations may not always be valid [5]. To overcome this problem algorithms such as SVR and RFR which captures non-linearity or PLSR-combined approaches such as locally-weighted-PLSR ($PLSR_{LW}$) and PLSR with feature selection ($PLSR_{FS}$) approaches are recently proposed [27–29].

Leaf area index (LAI) and canopy water content (CWC) are the main vegetation properties that can be investigated by remotely sensed spectral reflectance data. Both the vegetation variables are considered as important indicators of vegetation growth and productivity because they affect the exchanging of water and energy with the atmosphere [30–31]. Most of the present research using remotely sensed spectral reflectance data to link LAI and CWC has been conducted for assessing vegetation conditions and plant water status [31–35]. Leaf area index and canopy water content have been estimated by several algorithms for example using PLSR [6,10,17,36–42], SVR and RFR [10,18]. These studies suggest that the coefficient of determination (R^2) varied in the range of 0.62 to 0.94 for LAI and CWC with different vegetation type and environmental conditions. From the above literature, it is also evident that spectroscopic data need to be explored further and tested with newly developed algorithms and different vegetation types for prediction of LAI and CWC.

PLSR redesigned into locally-weighted PLSR ($PLSR_{LW}$) and feature selection-based PLSR ($PLSR_{FS}$) have been used in optical remote sensing applications for estimating soil properties, which improves the prediction accuracy significantly [28–29]. However these algorithms are new and yet to be tested in crop studies. Therefore, there is a need to test the aforementioned algorithms and evaluate their capability to estimate crop canopy parameters. Here, we evaluated five different algorithms for the rapid assessment of LAI and CWC for rice canopies using canopy reflectance spectra over visible to short-wave infrared regions. The suitability of each algorithm was analyzed and compared in terms of root mean squared error (RMSE) and coefficient of determination (R^2). To fully control the consistency of the measurements, a replicated field experiment was conducted with three varieties of rice and five levels of soil water potential.

2. Materials and methods

2.1. Study site and experimental details

Two year replicated field experiments were carried out on the experimental farm of Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, West Bengal (22° 18′ 51′′ N; 87° 18′ 54′′ E) during the dry seasons (December–April) of 2014–2016. Soils of the experimental plots are acidic with sandy loam in texture. Soil types are classified as Typic-Haplustalf according to the United States Department of Agriculture (USDA) system of soil taxonomy [43]. The climate in the area is humid subtropical characterized by short winter and long hot summer with an average rainfall of about 1600 mm yr^{-1} and pan evaporation of about 700 mm yr^{-1}. Thirty experimental plots (6 m × 5 m) were used for growing two drought-susceptible (Satabdi and IR36) and one drought tolerant (Vandana) rice varieties (Fig. 1). Satabdi and IR36 are medium duration (110–130 days) varieties suitable for lowland irrigated environment while Vandana is a short duration (90 days) upland variety. Rice seedlings of 28 and 18 days durations were transplanted in the puddled plots in the year 2014–2015 and 2015–2016, respectively. Five water stress treatments as soil water potential levels (−30 kPa, −50 kPa, −70 kPa, −120 kPa and −140 kPa) were applied with two replications in completely randomized block design. Water stress treatments were applied during vegetative growth stages by withholding irrigation and allowing soils to dry out naturally. Recommended fertilizers doses containing 120 kg ha^{-1} nitrogen (Urea), 50 kg ha^{-1} phosphate (P_2O_5) and 60 kg ha^{-1} potash (K_2O) were applied as basal before transplanting, 25% N at mid tillering and remaining 50% N at flowering stages. Insects and pests were controlled using recommended doses of chemical pesticides.

2.2. Field and laboratory data collection

2.2.1. Canopy spectral reflectance measurements

Canopy reflectance spectra were measured in each treatment plots at the canopy during 11 AM– 2 PM on cloud free days using a handheld spectroradiometer (Field Spec 3 FR, Analytical Spectral Devices Inc., Colorado, USA; spectral range: 300–2500 nm). Canopy spectral measurements were taken from canopy height of 0.5 m with a field of view of 8°. A spectralon white reference panel (Labsphere, Inc., Sutton, NH) was

Fig. 1 – Field view of the experimental site.

mounted on a tripod stand, which was placed between two adjacent plots. The panel was mounted at the same level of the canopy on which spectral measurements were done. Canopy spectral reflectance data were collected five times in two different locations in each plot during 2014–15 and 2015–16 growing seasons (55, 62, 66, 83 and 103 different days after sowing (DAS) in 2014–15 and 55, 73, 81, 99 and 120 different days after sowing (DAS) in 2015–16). This resulted in 600 (30 plots × 2 locations × 5 DAS × 2 growing seasons) canopy spectra and associated rice canopy parameters.

2.2.2. *Leaf area index and canopy water content measurements*

Leaf area index and canopy water content measurements were obtained, immediately after canopy spectral measurements. To create variations in leaf area index and canopy water content, the measurements were taken at different days after sowing (transplanting to harvest maturity) during the drying season. An indirect method was used to measure LAI [44]. In this method, all the leaves from a rice hill were harvested and three sample leaves were taken to measure leaf area using an optical planimeter (Model no. 211, Systronics, India). The total leaf area was then estimated by multiplying the resulting leaf area with the weight ratio between the total number of leaves and three selected leaves. Total ground area for each hill was calculated from the spacing used for transplanting (0.2 m × 0.2 m). Leaf area index was calculated as:

$$\mathrm{LAI}(\mathrm{m}^2\ \mathrm{m}^{-2}) = \frac{\text{Half of total leaf area}(\mathrm{m}^2)}{\text{Total ground area}(\mathrm{m}^2)} \quad (1)$$

Canopy water content was estimated by measuring the equivalent water thickness (EWT) of the leaves harvested from a rice hill. Leaves on the rice hills were collected and packed in zip lock bags for taking fresh weights (FW). Dry weights (DW) were taken after drying the leaves at 70 °C in an oven for 48 h until constant weight was reached. CWC was calculated using EWT as:

$$\mathrm{EWT}\ (\mathrm{g}/\mathrm{m}^2) = \frac{\mathrm{FW} - \mathrm{DW}}{\text{Half of total leaf area}} \quad (2)$$

$$\mathrm{CWC}\ (\mathrm{g}/\mathrm{m}^2) = \mathrm{EWT} \times \mathrm{LAI} \quad (3)$$

A pool datasets of 600 LAI and 480 CWC measurements were obtained during the two year field experiments.

2.3. *Spectral preprocessing*

Each spectrum was corrected for the sensor adjustment. Smoothing was acquired by third order Savitzky-Golay smoothing method with a span length of 9 nm to remove noise for optimizing the signal to noise ratio [45]. Normality check of each canopy attributes were performed using Kolomogorov-Smirnov test at 5% significance level. Water absorption bands over 1 350∼1 460 nm, 1 804∼1 971 nm and noisy reflectance values before 400 nm and after 2 447 nm were removed from each reflectance spectrum. The remaining 1 809 wavebands were used for analysis. The standard normal variate (SNV) transformation was used to remove the noise from the spectra.

2.4. *Algorithm descriptions*

Three types of algorithms were used in this study: a) general PLSR, SVR and RFR, b) locally-weighted-PLSR ($\mathrm{PLSR_{LW}}$), and c) combined PLSR and feature selection ($\mathrm{PLSR_{FS}}$) algorithms. In the PLSR algorithm, input features are projected onto a space such that the covariance between the features in the projected space and response variable is maximized. In the SVR approach, a function from a given training sample is estimated in such a way that the predicted response for each training data is at most away from the observed response [46]. Since training data points serve as support vectors for the estimated function, this approach is relatively less influenced by the presence of outliers in the calibration or validation datasets compared with the general PLSR. RFR is a classifier ensembling classification trees, in which, each tree gives a classification, and the tree “votes” is used to classify samples. The forest chooses the classification having the most votes. RFR is resistant to overfitting and usually performs well in problems with a low ratio of number of samples to number of features, like spectrometric data [47].

Because the modeling results from the PLSR, SVR and RFR approaches were comparable, attempts were also made to improve the performance of a PLSR model using the $\mathrm{PLSR_{FS}}$

and $PLSR_{LW}$ approaches. $PLSR_{FS}$ approach was applied by implementing an ordered predictor selection (OPS) approach in which a parsimonious set of predictors could be selected for a specific regression model. The variables selected in this approach are PLSR dependent variables (regression coefficients (β), variable importance of projection (VIP), squared residual (SqRes), net analyte signal (N)) and PLSR independent variables (correlation vector (r), biweight midcorrelation vector (bicor), mutual information based adjacency vector (AMI), signal to noise vector (StN) and covariance procedure (CovProc)). Calculation and details on the spectral variable indicators used have been described in [28]. In addition, new sets of spectral variable indicators were generated by pair-wise combination of a) PLSR-dependent indicators only (6 combinations), b) PLSR-dependent and independent indicators (20 combinations). The combined indicator was the result of element wise product of absolute values of the normalized spectral variable indicators. Thus, a total of 35 spectral variable indicators (9 individual + 26 combinations) were examined in this study. The OPS approach was applied to each of these sets to identify a parsimonious set of predictor variables such that a reduced and effective set of predictor variables may be used in the regression model. Such a feature selection approach increased the prediction accuracy of the final PLSR model [28]. The $PLSR_{FS}$ approach was implemented for all the spectra. Results corresponding to the best combination of features with highest prediction accuracy based on root mean squared error (RMSE) and complexity (in terms of number of optimum spectral variables (NSV)) used are compared. Optimum models with RMSE value within 5% proximity (in magnitude) to the lowest RMSE were selected. Among these, the model with low NSV was treated as the best model and respective variable indicator as the 'best' for both the crop parameters.

In $PLSR_{LW}$, PLSR models are calibrated by assigning weights to each calibration sample based on its similarity to the respective validation sample [29]. The similarity between calibration and validation samples was calculated using a distance metric as follows:

$$d_{c,v} = \sqrt{(x_c - x_v)^T \theta (x_c - x_v)} \quad (4)$$

where $d_{c,v}$ is the distance between c^{th} calibration and v^{th} validation samples, x_c ($p \times 1$) is the calibration sample, x_v ($p \times 1$) is the validation sample, θ is an identity matrix to calculate the similarity between observe and predictor variables, and p is the total number of predictor variables. Two approaches were used to calculate $d_{c,v}$. In the first approach, θ was calculated based on the correlation between response and predictors, abbreviated as θ_{corr}. In the second approach, θ was calculated based on the covariance between predictors and response, abbreviated as θ_{cov}:

$$\theta_{corr} = \mathrm{diag}\left\{\frac{\|X^T y\|}{\|X\|\|y\|}\right\} \quad (5)$$

$$\theta_{cov} = \frac{X^T y y^T X}{\|X^T y\|^2} \quad (6)$$

where diag{} represents the diagonal matrix, X is an ($n \times p$) matrix of predictors with the number of samples n and number of predictors p, and y is the column vector of response variable. The weights for each sample were calculated from estimated $d_{c,v}$ as follows:

$$w_{c,v} = \exp\left(-\frac{\phi d_{c,v}}{d_s}\right) \quad (7)$$

where $w_{c,v}$ is the weight of the c^{th} calibration sample for the v^{th} validation sample, ϕ is the localization parameter, and d_s is the standard deviation of distance of all the calibration samples from the v^{th} validation sample. The localization parameter ϕ is a user-defined value, which determines the extent of localization. The value of ϕ = zero gives unit weight to each calibration sample which is essentially the PLSR. $PLSR_{LW}$ approach is expected to perform better than the PLSR because of its capability to capture non-linearity in the data [48].

2.5. Spectral model development

The algorithms discussed above were used to build the spectral models between the rice canopy parameters and reflectance spectra. The residuals (significance level = 5%) between observed and predicted response variables from the principal component regression were examined for outlier removal using the rcoplot function in MATLAB [49]. After outlier removal, the remaining dataset was partitioned into a calibration (70%) and validation (30%) dataset. The partition was performed using a stratified sampling approach [50]. In this approach, the LAI and CWC datasets were sorted in an ascending order and every third sample was selected for validation. Subsequently, the number of latent variables or components that gave minimum RMSE in cross-validation was used for model calibration [51].

2.6. Evaluation of the model over-fitting

The spectral reflectance dataset was generated by measuring in situ canopy spectra at different growth stages of rice during two rice-growing seasons. In essence, such a dataset captures time series of spectra for rice. Second, canopy spectra were also from a single replicated trial. The spectral model calibrated from such a dataset requires critical validation through multi-site, multi-variety, and even multi-season trials, which was beyond the scope of this study. Nevertheless, resulting spectral model from this study was tested for overfitting using the following three simulation approaches:

(a) Cross-year validation of the spectral dataset: This exercise was done to test the model transferability by calibrating and validating with individual year dataset
(b) Deletion of spectral data for specific dates followed by partitioning remaining data into 70:30 calibration and validation samples through sorting: This exercise was done to test if a reduced dataset (by deleting data for a specific day of observation) yields acceptable performance statistics for selected models.
(c) Selection of spectral data for one sampling date from each year as validation sample and remaining as calibration sample: This exercise was done to test if a calibrated model developed from a selected set of

observation dates during the crop growth season is capable of predicting selected crop parameters for an entirely different date of observation which is not part of the training dataset at all.

Case 1: A cross year validation of the dataset was performed. During the process, the model was tested for individual year. In this case, model was developed and tested with individual year data. Model simulations are performed for the five modeling approaches (PLSR, RFR, SVR, $PLSR_{FS}$, and $PLSR_{LW}$) used in this study.

Case 2: Deletion of spectral data for specific dates followed by partitioning remaining data into 70:30 calibration and validation samples through sorting: In this case, spectral data for a specific DAS were completely removed from each year's modeled data. The remaining data was sorted in ascending order with every third sample used as validation data and remaining samples as calibration data. Simulations were performed separately for each growing season. Thus, out of 300 samples (5 sampling DAS × 60 samples per sampling) in a growing season, 60 samples were removed. Remaining 240 samples were divided into 160 calibration samples and 80 validation samples. We also merged data for both the growing seasons and followed the same procedure to have a calibration dataset of 320 samples and validation dataset of 160 samples by removing 120 samples from the total of 600 samples. Five regression algorithms (PLSR, SVR, RFR, $PLSR_{LW}$ and $PLSR_{FS}$) were evaluated for each of these three datasets (2014–15 growing season, 2015–16 growing season and pooled data). These combinations led to 15 modeling scenarios as summarized in Table 1.

Case 3: Selection of spectral data for one sampling date from each year as validation sample and remaining as calibration sample: In this case, data for a selected DAS were used as validation data and the remaining were used to train a chemometric model as shown for scenarios b to f in Table 2. The data used for the validation data was indeed completely censored from the training dataset. Simulations were performed with all the five regression algorithms (PLSR, RFR, SVR, $PLSR_{FS}$, and $PLSR_{LW}$).

2.7. Assessment of model performance

The accuracies of the developed algorithms were evaluated using the coefficient of determination (R^2), root-mean-squared error (RMSE).

$$R^2 = 1 - \frac{SS_{error}}{SS_{total}} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(Y_i - Y_i\right)^2} \tag{9}$$

where SS_{error} is the sum of the square of the errors and SS_{total} is the sum of the square of the observed response variable. N is the total number of samples, Y_i, and Y_i are actual and predicted response variable, respectively, for i^{th} sample. Also, a significance test was carried out to check if algorithms were statistically different or not. For checking the robustness of regression algorithms, a bootstrapping approach was also implemented on the validation dataset using 1000 bootstrap events. The RMSEs were calculated on 1000 bootstrap samples of the validation dataset for both LAI and CWC and a two-sample *t*-test was carried at 5% significance level [52]. Our null hypothesis was that the RMSEs obtained by the models came from independent datasets from normal distributions with equal means but unknown variances and the alternate hypothesis was that it came from two distributions with unequal means. All the necessary data processing, calibration sampling and regression modeling were executed in the MATLAB (R2012a, The Mathworks, USA) environment.

3. Results and discussion

3.1. Descriptive statistics of LAI and CWC

Descriptive statistics of LAI and CWC are shown in Table 3, which suggests that the average values for both the rice parameters were within their expected range. Similar findings were reported for LAI [53–55] and CWC [56–57]. The experimental protocol ensured a wide range of variation in LAI ($m^2\ m^{-2}$) and CWC ($g\ m^{-2}$). LAI varied from 1.21 $m^2\ m^{-2}$ to 8.59 $m^2\ m^{-2}$ with an average of 4.05 $m^2\ m^{-2}$. Similarly, CWC varied from 61.9 $g\ m^{-2}$ to 1 130.7 $g\ m^{-2}$ with an average value of 382.4 $g\ m^{-2}$. Wide variability in LAI and CWC may be the result of reduced leaf sizes and changes in canopy architecture during water stress. Other factors such as different varieties, growth stage and soil health may have been responsible for such variability. The coefficient of variation (CV) values for LAI and CWC data collected for both the season were found to be higher (CV = 44.9% and 45.6%, respectively), which suggests that the data generated in this study has desirable variability to develop regression models. The frequency distributions of both vegetation parameters were relatively less skewed and flattened. The Kolomogorov-Smirnov test at the 5% significance level suggested that both the vegetation parameters followed normal distribution. Hence, these two crop parameters were estimated without applying any transformation on them.

3.2. Rice canopy spectral reflectance

Large variability in the range of observed canopy variables could potentially influence the canopy reflectance. Fig. 2(a) shows the average canopy spectra (n = 600) along with their variations. The boxplots with median, lower and upper quartile values of spectral reflectance over visible (VIS; wavelength: 350∼700 nm), near infrared (NIR; wavelength: 700∼1 300 nm), first half of shortwave infrared band (SWIR1; wavelength: 1500–1803 nm), and second half of shortwave infrared band (SWIR2; wavelength: 1 971∼2 440 nm) are shown in Fig. 2(b). These figures illustrate the influence of canopy traits and leaf optical properties in terms of high reflectance in the NIR and SWIR1 and low reflectance in the VIS and SWIR2 regions. Surface soil characteristics in the experimental site were relatively similar across 30 different plots (supplementary material, Fig. S1). Thus, the variability of the canopy spectra can be attributed to variations in the canopy architecture and their optical properties. Similar assumptions are also reported in literature [30,58].

Table 1 – Modeling scenarios obtained by considering spectral data for selected days after sowing (DAS) for model calibration and validation; deleted DAS values were completely removed from modeling dataset thereby reducing the size of the modeled data.

Modeling scenario	Calibration	Validation	Deleted DAS
	DAS for 2014 growing season		
1	62, 66, 83 &103	62, 66, 83 &103	55
2	55, 66, 83 & 103	55, 66, 83 & 103	62
3	55, 62, 83 & 103	55, 62, 83 &103	66
4	55, 62, 66 & 103	55, 62, 66 & 103	83
5	55, 62, 66 & 83	55, 62, 66 & 83	103
	DAS for 2015 growing season		
6	73, 81, 99 & 120	73, 81, 99 & 120	55
7	55, 81, 99 &120	55, 81, 99 &120	73
8	55, 73, 99 & 120	55, 73, 99 &120	81
9	55, 73, 81 & 120	55, 73, 81 &120	99
10	55, 73, 81 & 99	55, 73, 81 & 99	120
	DAS for pooled samples		
11	62, 66, 83 &103 73, 81, 99 & 120	62, 66, 83 &103 73, 81, 99 & 120	55 and 55
12	55, 66, 83 & 103 55, 81, 99 &120	55, 66, 83 & 103 55, 81, 99 &120	62 and 73
13	55, 62, 83 & 103 55, 73, 99 &120	55, 62, 83 & 103 55, 73, 99 &120	66 and 81
14	55, 62, 66 & 103 55, 73, 81 &120	55, 62, 66 & 103 55, 73, 81 &120	83 and 99
15	55, 62, 66 & 83 55, 73, 81 & 99	55, 62, 66 & 83 55, 73, 81 & 99	103 and 120

Table 2 – Modelling scenarios obtained by considering spectral data for selected days after sowing (DAS) for model calibration and validation.

Modeling Scenario	Collection dates (DAS) for spectra in 2014 and 2015 growing season	
	Calibration	Validation
Scenario with Un-censored Data		
a	2014 season: 55, 62, 66, 83 &103 2015 season: 55, 73, 81, 99 & 120	Validation data obtained by sorting followed by partitioning and all the 600 samples are used.
Scenarios with Censored Data		
b	2014 season: 62, 66, 83 &103 2015 season: 73, 81, 99 & 120	2014 season: 55 2015 season: 55
c	2014 season: 55, 66, 83 & 103 2015 season: 55, 81, 99 &120	2014 season: 62 2015 season: 73
d	2014 season: 55, 62, 83 & 103 2015 season: 55, 73, 99 & 120	2014 season: 66 2015 season: 81
e	2014 season: 55, 62, 66 & 103 2015 season: 55, 73, 81 & 120	2014 season: 83 2015 season: 99
f	2014 season: 55, 62, 66 & 83 2015 season: 55, 73, 81 & 99	2014 season: 103 2015 season: 120

3.3. *Prediction of LAI and CWC using different modelling algorithms*

Fig. 3 and Table 4 list the validation R^2, RMSE values of LAI and CWC using different algorithms in the pooled dataset. LAI was predicted with higher accuracy than CWC for the algorithms used. Canopy spectral reflectance is functionally linked with LAI [36,59], which has led to the development of several retrieval algorithms for LAI from multispectral data. Similarly, among algorithms, PLSR combined approaches of $PLSR_{LW}$ and $PLSR_{FS}$ performed better than those of PLSR, SVR and RFR algorithms for LAI (R^2 = 0.77 and 0.73) and CWC (R^2 = 0.67 and 0.63), respectively. Scatter plot between observed and predicted values of LAI and CWC values suggested that the $PLSR_{LW}$ and $PLSR_{FS}$ algorithms may be used as robust algorithms for the estimation of both the vegetation

Table 3 – Descriptive statistics of leaf area index and canopy water content.

Vegetation parameters	N	Min	Max	Mean	CV	Skewness	Kurtosis
LAI (m^2/m^2)	600	1.21	8.59	4.05	44.90	0.11	−1.31
CWC (g/m^2)	480	61.87	1 130.7	382.43	45.60	−0.06	0.57

n, Sample size; Min, Minimum; Max, Maximum; Mean; CV, Coefficient of variation; Skewness; Kurtosis; LAI, Leaf area index; CWC, Canopy water content.

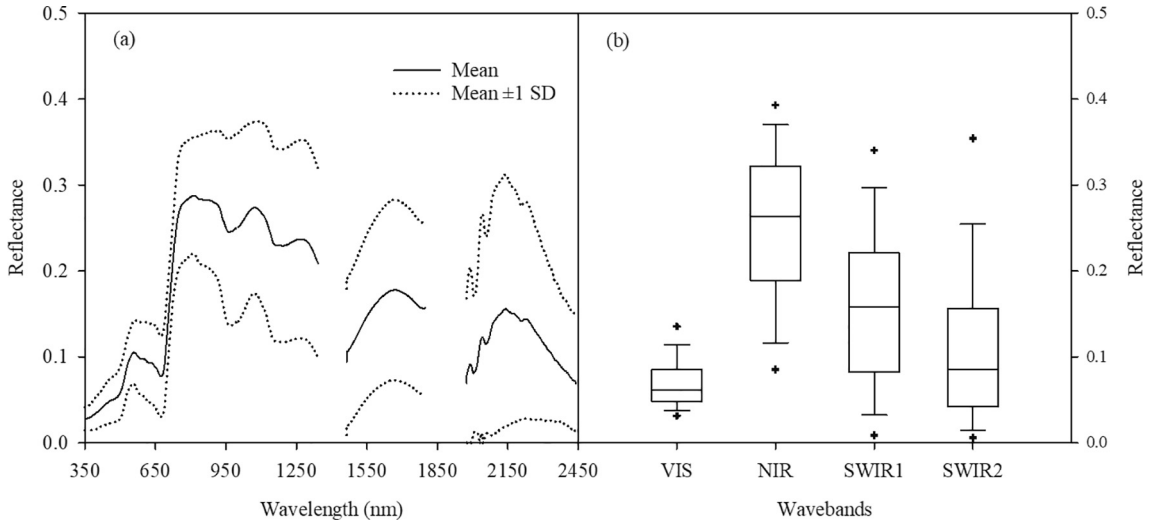

Fig. 2 – Average spectral reflectance (n = 600) with their variation (a) and the boxplots with median, lower and upper quartile values of spectral reflectance over visible (VIS; wavelength: 350∼700 nm), near-infrared (NIR; wavelength: 700∼1 300 nm), first half of shortwave infrared band (SWIR1; wavelength: 1 500∼1 803 nm), and second half of shortwave infrared band (SWIR2; wavelength: 1 971∼2 440 nm).

parameters (Fig. 3). In general, there was no improvement in prediction accuracy upon using PLSR, SVR and RFR, which may be due to the poor waveband selection and inconsistent weight adjustment to minimize the estimation error [8,60].

Estimated RMSE values of our study are similar to or better than those reported for LAI prediction on rice canopies with RMSE values in the range of 0.85–1.04 using MLR, PLSR and LS-SVM models [60]. In contrast, lower RMSE values (0.23∼0.41) were observed for vegetation water content (VWC) using canopy spectra from pot experiments and PLSR and artificial neural network algorithms [41]. Resulting RMSE values in our study for LAI ranged from (0.85∼0.94) for $PLSR_{LW}$ and $PLSR_{FS}$ to as high as (0.97∼1.05) for PLSR, SVR and RFR approaches. Similarly, RMSE values for CWC ranged from 96.25 to 99.92 for $PLSR_{LW}$ and $PLSR_{FS}$ to as high as (101.2∼105.3) for the remaining algorithms. These results suggest that LAI and CWC can be estimated from canopy spectral data with reasonable accuracy using $PLSR_{LW}$ and $PLSR_{FS}$ algorithms.

3.4. *Assessment of the model performance and their significance test*

The RMSE values were computed for both the vegetation parameters (Table 4). Results show that RMSE reduction in LAI improved the performance of proposed $PLSR_{LW}$ algorithm about 15.3%, compared to the poor performing model SVR. Similar results are also observed in CWC with a reduction in RMSE of about 9.4%, However, the $PLSR_{FS}$ improves the model performance about 4.3% and 5.5% for LAI and CWC, respectively. These results are shown in Figs. 4 and 5. Left panels in these figures show the kernel-smoothed density functions and the right panels show the resulting boxplots for possible error distributions in using the calibrated model for these two parameters. In the $PLSR_{FS}$ approach, the percent difference in RMSE values between full-spectrum model and optimum models using different variable indicators for both crop parameters are shown in Fig. 6. The baseline value of zero corresponds to the RMSE of full-spectrum model. The negative and positive bars represent the improvement and deterioration (possibly, information loss) in the prediction accuracy accomplished with the OPS approach, respectively. The best variable indicators identified for LAI is CovProc. However, the OPS approach performed by β-StN and N-StN appears to be the best variable indicators for CWC. The performance of best model was found to be superior to full-spectrum model for both the crop parameters. The percent decrease in RMSE value attained using OPS approach was found to be highest for LAI (6.37%). About 4% decrease in RMSE was noted for CWC.

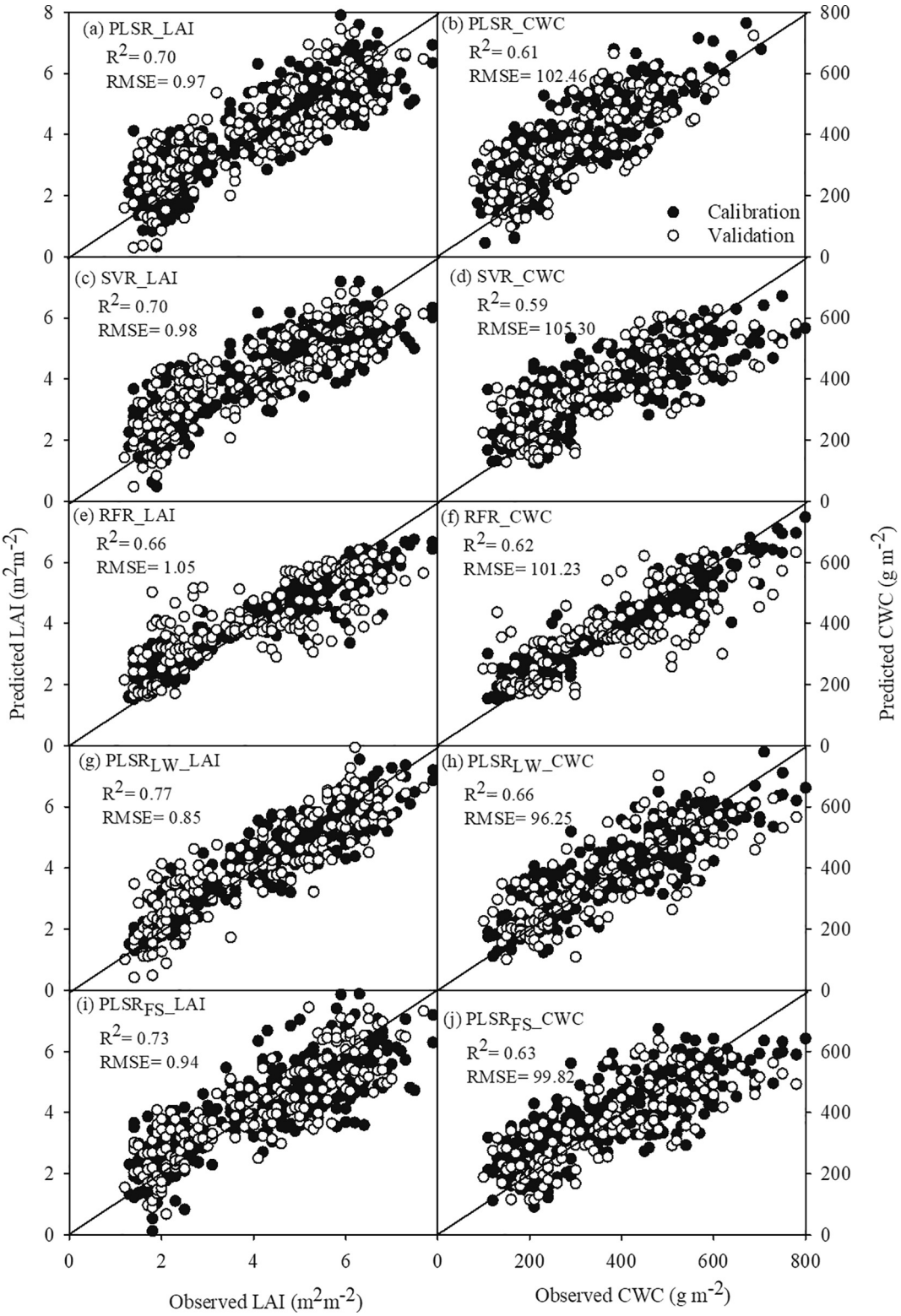

Fig. 3 – Scatter plots between observed and predicted values of leaf area index (LAI) and canopy water content (CWC).

Bootstrapped RMSE results suggest that the RFR model leads to maximum error in predicting the LAI in the validation datasets while the local modeling approach with PLSR performs the best with minimum RMSE and low interquartile range for the resulting RMSE values. For the CWC, PLSR, SVR and RFR models show similar performance while the $PLSR_{LW}$ again showing minimum RMSE value with low narrow interquartile range. In both the cases of LAI and CWC, $PLSR_{FS}$

Table 4 – Regression statistics for prediction of LAI and CWC.

Vegetation parameters	PLSR		SVR		RFR		$PLSR_{LW}$		$PLSR_{FS}$	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
LAI (m^2/m^2)	0.70	0.97	0.70	0.98	0.66	1.05	0.77	0.85	0.73	0.94
CWC (g/m^2)	0.61	102.46	0.59	105.30	0.62	101.23	0.66	96.25	0.63	99.82

LAI, Leaf area index; CWC, Canopy water content; R^2, Coefficient of determination; RMSE, Root mean squared error; PLSR, Partial least squared regression; SVR, Support vector regression; RFR, Random forest regression; $PLSR_{LW}$, Partial least squared regression- locally weighted; $PLSR_{FS}$, Partial least squared regression- feature selection.

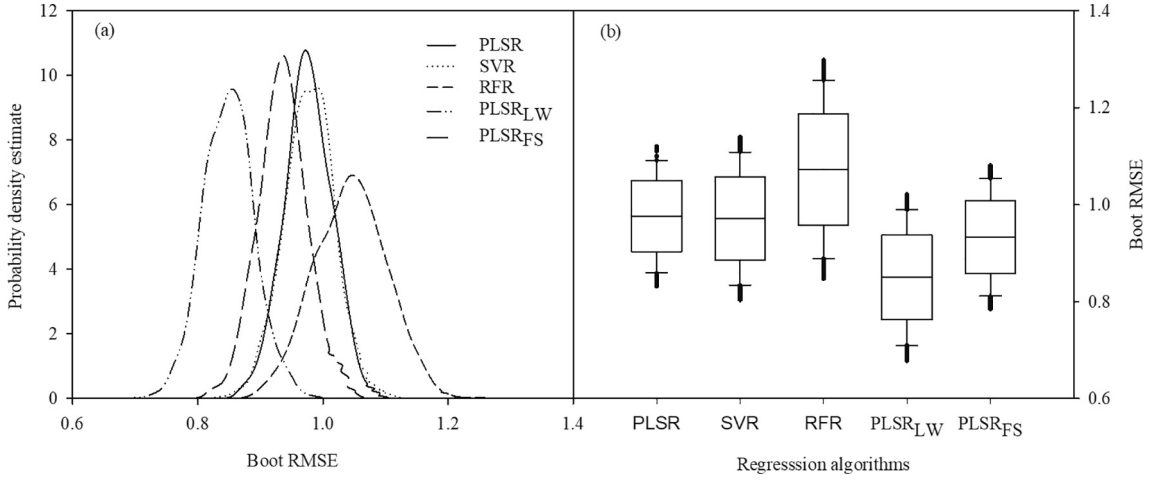

Fig. 4 – Distribution of (a) probability density function (pdf) estimated by kernel density estimator of the 1000 RMSE values and (b) the boxplots with median, lower quartile and upper quartile from all the boot strapped calibration approaches used for estimating LAI.

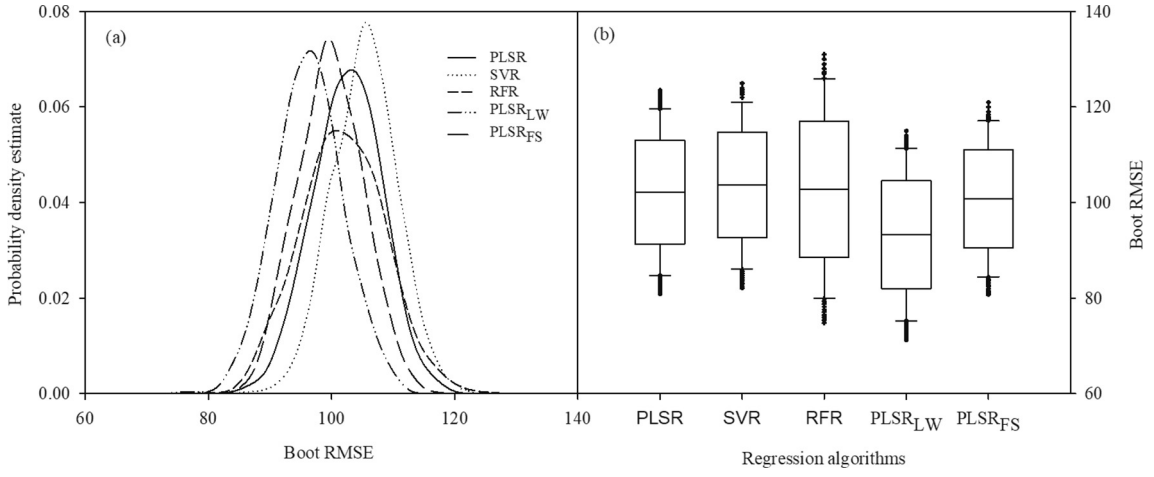

Fig. 5 – Distribution of (a) probability density function (pdf) estimated by kernel density estimator of the 1 000 RMSE values and (b) the boxplots with median, lower quartile and upper quartile from all the boot strapped calibration approaches used for estimating CWC.

was the second best performing spectral aalgorithm among the five algorithms tested in this study. Therefore, it may be concluded that these modified PLSR-based approaches and, specifically, the locally-weighted PLSR model appears to be a promising approach for estimating LAI and CWC from field-measured canopy spectra.

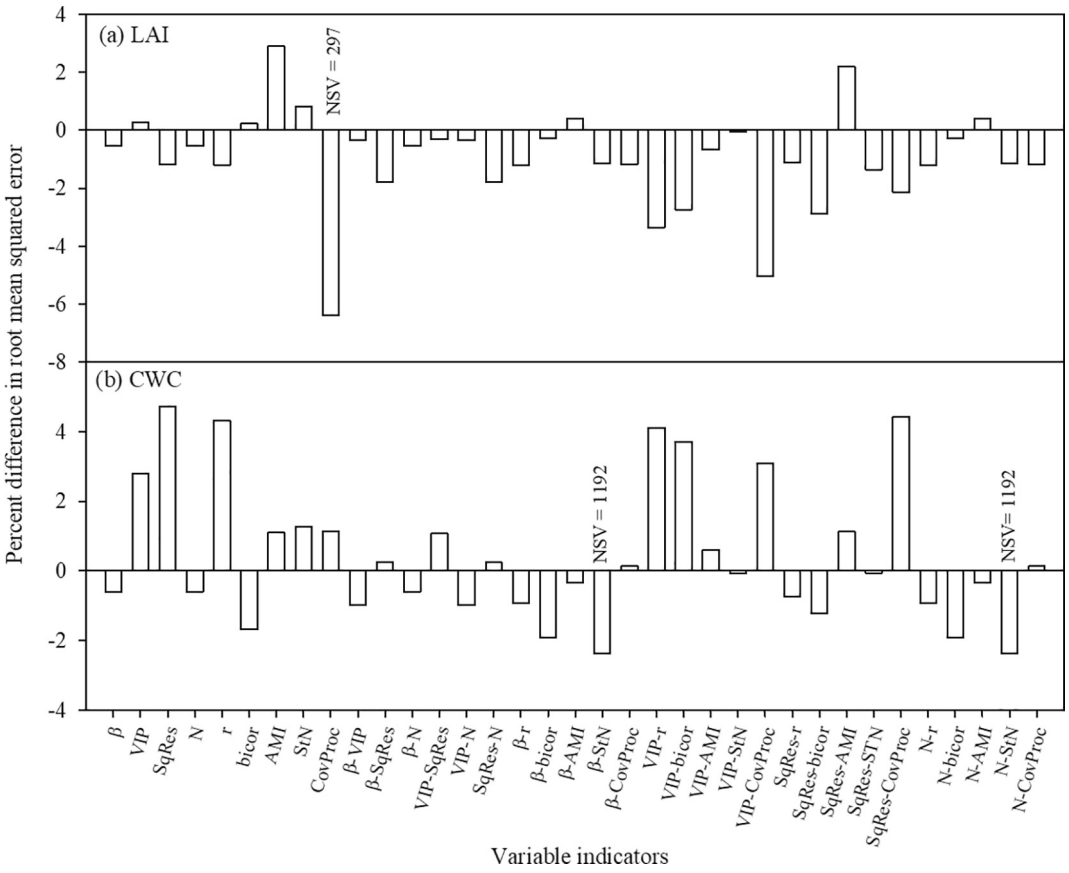

Fig. 6 – Comparison of RMSE of full-spectrum with optimum spectral models using variable indicators used (β: regression coefficient, VIP: variable influence on projection, SqRes: squared residual vector, r: pearson correlation coefficients, bicor: biweight mid correlation vector, AMI: mutual information based adjacency vector, StN: signal to noise ratio, CovProc: covariance procedure and their pairwise combinations). NSV: number of spectral variables.

3.5. *Assessment of simulations designed for testing model over-fitting*

Simulation 1: In addition to the pooled dataset prediction performances, a cross year validation of the datasets were tested and the results shown in supplementary material (Table S.2). The model accuracies in terms RMSE are within (0.73∼1.04) for LAI and (79.04∼124.64) for CWC. These results are similar in range to the pooled datasets shown in Table 4 and Fig. 3. These results show that pooled datasets yields similar model performances with individual year dataset. In addition to the LAI and CWC, the modelling results for EWT are shown in supplementary material (Table S.2). The model accuracies are within the range of (14.13∼18.11). These results show that modelling EWT with all the five algorithms do not provide reasonable performance.

Simulation 2: Deletion of spectral data for specific dates followed by partitioning remaining data into 70:30 calibration and validation samples through sorting: Simulation results are examined using R^2 statistics (supplementary material, Fig. S.2) for illustrating over-fitting issue. For LAI, the R^2 values for the validation datasets ranged from 0.38 to 0.90 for RFR model to as high as 0.61–0.90 for the $PLSR_{LW}$ modeling approach. When all the data were considered, the R^2 value ranged from 0.60 to 0.88 (as shown in Table 4). We also repeated this exercise for CWC and observed similar performance statistics. These results show that a reduced sample size (derived by removing data from specific dates) yields similar model performance as was shown for the whole dataset.

Simulation 3: Selection of spectral data for one sampling date from each year as validation sample and remaining as calibration sample: Observed vs. predicted LAI and CWC values for the calibration and validation datasets are shown along the 1:1 line for the best model ($PLSR_{LW}$) in Figs. 7 and 8, respectively. These figures show that the observed and predicted values for the deleted DASs (validation data) generally fall on the 1:1 line except for the first modeling scenario b (Table 2). Resulting RMSE values for the validation datasets are generally larger for LAI and are often lower (scenario c, e and f in supplementary material (Fig. S.3)) for the CWC values than those obtained when whole datasets were considered in modeling. Similar results were obtained for remaining four modeling approaches (PLSR, RFR, SVR, $PLSR_{FS}$). Thus, the censored data shows similar model performance to those of uncensored data.

Figures in supplementary material (Fig. S.3 and Fig. S.4) were generated by bootstrapping procedure on the validation datasets outlined for the six modeling scenarios in Table 2 for

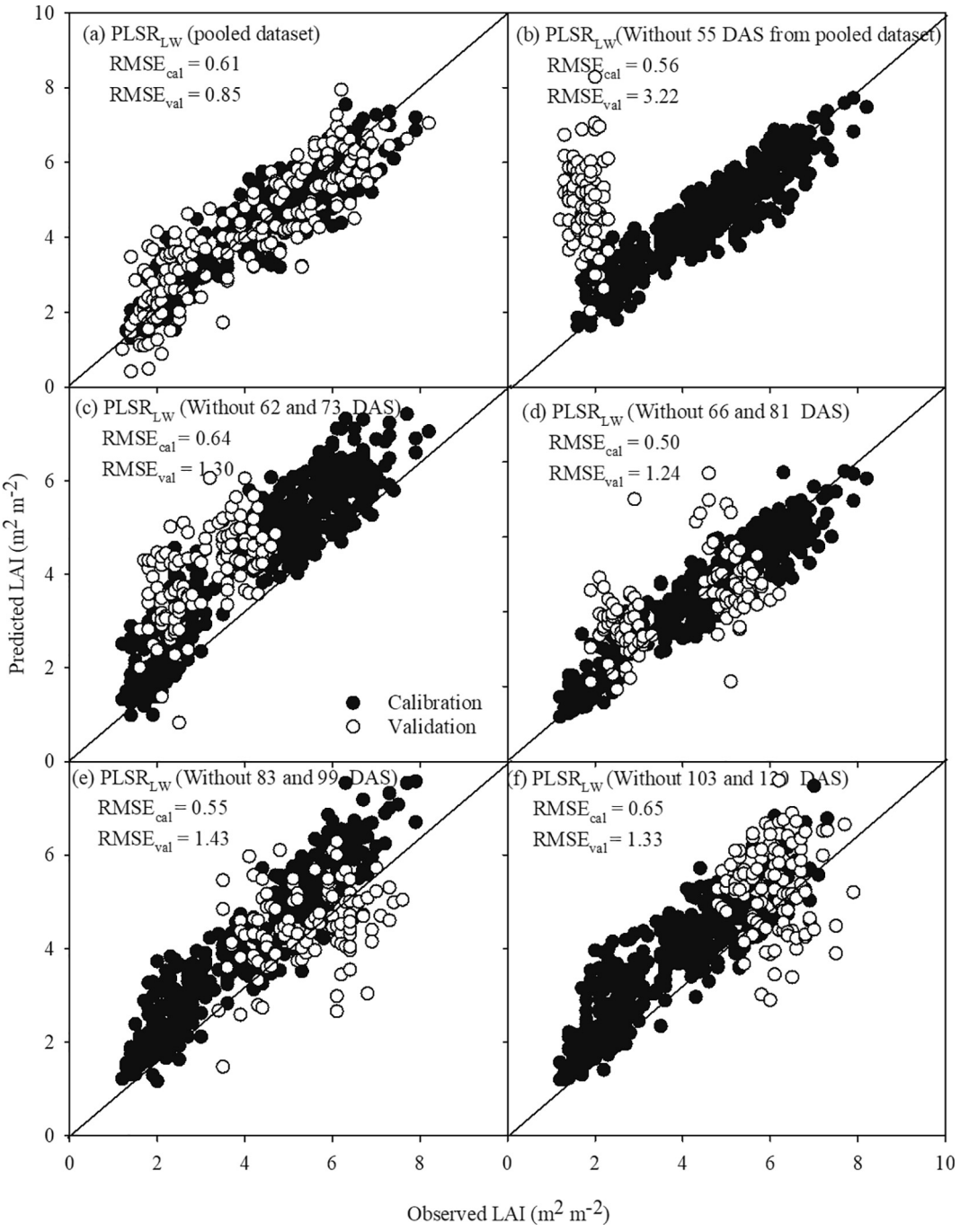

Fig. 7 – Scatter plot between calibration and validation datasets of LAI (m^2 m^{-2}) for $PLSR_{LW}$ model with removal of a specific date (DAS) from each year dataset.

LAI and CWC, respectively. The median values, lower and upper quartile values for the estimated validation RMSE values scenarios c to scenario e are similar to those of scenario a. As expected, Subsetting of the spectral data for the first DAS (scenario b) and last DAS (scenario e) generally leads to larger RMSE values because the training dataset does not include low and high LAI values. These results further support our observations made from Figs. 7 and 8. These simulation analysis suggest that the model calibration resulted reported in this study may be free from over-fitting issues. Nevertheless, the method developed herein requires, at least, multi-site assessment.

3.6. *Computational efficiency of algorithms*

Computational efficiency of algorithms were also compared based on their execution time to run each algorithm. Among the non-parametric algorithms, SVR, RFR and neural network

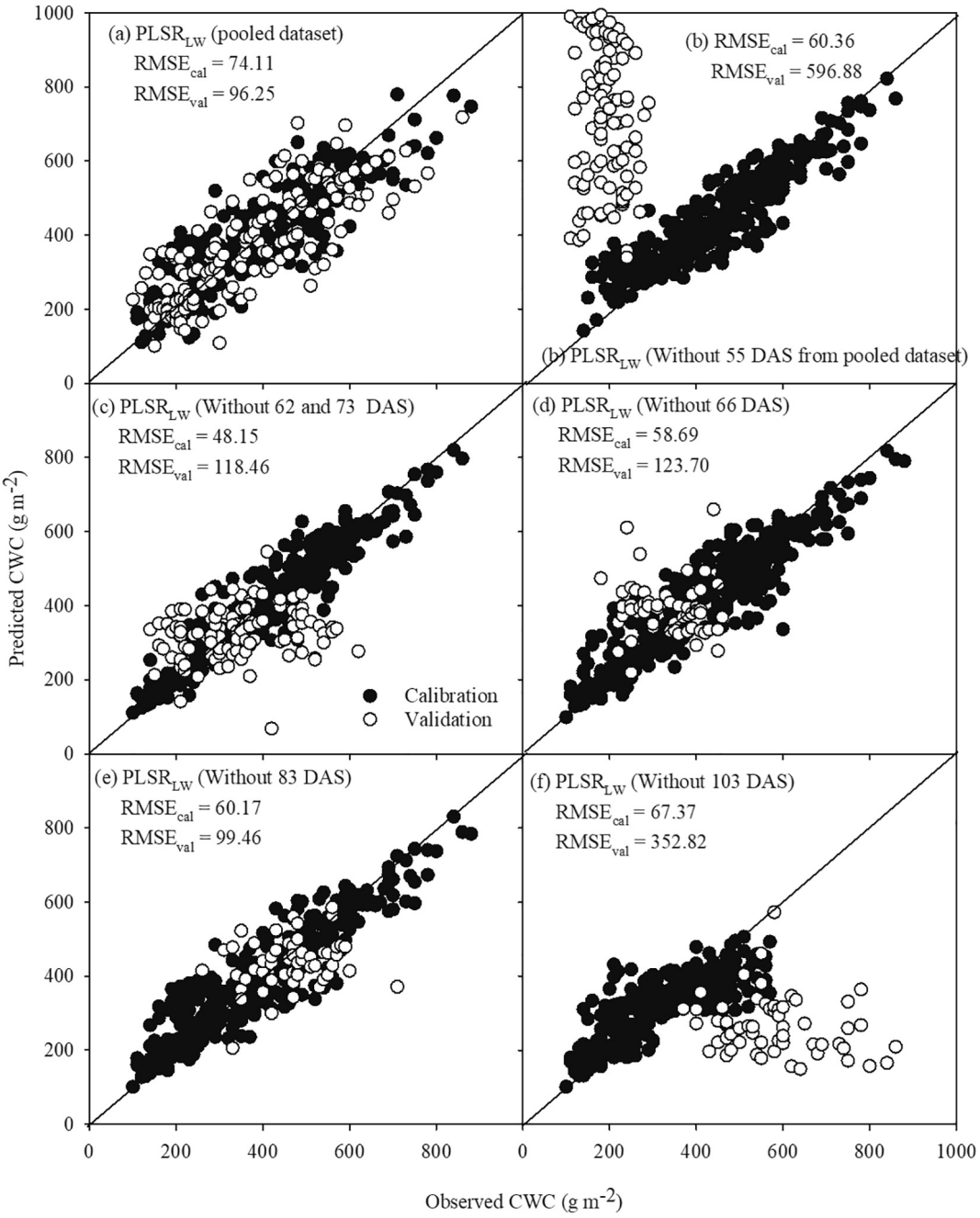

Fig. 8 – Scatter plot between calibration and validation datasets of CWC (g m^{-2}) for PLSR$_{LW}$ model with removal of a specific date (DAS) from each year dataset.

were comparable, but no strong preference was found for any specific model. In general, PLSR$_{LW}$ had higher computational time requirement for executing the algorithm, which may be due to the different distance calculations for each iteration to optimize the parameters [29]. Similarly, PLSR combined with feature selection approach is also computationally intensive, as it take extra time to select the best feature space from a set of 35 different variable indicators consisting of regression coefficient (ß), VIP and mutual information-based adjacency factors and their combinations [28]. High computational time requirement for local and the variable indicators-based approaches may be the reason for their minimal use in remote sensing literature. With the advancement of fast computational technology, it may be possible to use these algorithms for real-time estimation of vegetation canopy parameters.

4. Conclusions

In this study, optical remote sensing with algorithms were tested to determine the most efficient empirical model for the assessment of LAI and CWC for dry rice canopies. Reflec-

tance spectroscopic approach proved to be a valuable tool. Results showed that locally- weighted PLSR performed accurately for predicting LAI and CWC with R^2 values 0.77 and 0.66 respectively. The improved performance of $PLSR_{LW}$ appears to be influenced by the choice of suitable weighing methods to reduce the estimation error in the visible and near infrared space and its projected spaces. Moreover, the selection of an appropriate distance measure for constructing the weighing scheme appears to be a crucial step in $PLSR_{LW}$ for improving its efficacies in modelling crop canopy parameters. This study proved that locally weighted PLSR, which was developed based on assigning higher weighing to similar variables based on their distance metrics is a fast and robust approach for remote quantification of LAI and CWC on rice canopies. A specific limitation of this study is that only a single experimental field was used in testing our approach. Variations in soil and crop characteristics require that the present approach be evaluated at different locations and for different cropping conditions. Therefore, more varieties and wide ranges of crop parameters can be studied from remote sensing data using the $PLSR_{LW}$ algorithm in future. A further limitation of this study is the background soil reflectance, which was not accounted in modeling canopy spectra for LAI and CWC estimation. Although there may be small variation in soil properties in this experimental plots, changes in soil wetness may lead to substantial variation in spectral reflectance values specifically during the early growing season when much of the soil surface is not covered by the crop canopy.

Declaration of Competing Interest

The author declare that there is no conflict of interest.

Acknowledgements

The authors acknowledge the Ministry of Human Resource Development, Government of India for providing assistantship to the senior author. Also, we would like to acknowledge Dr. M.C Sarathjith and Abhinav Gupta for providing the Matlab code of partial-least-squares combined feature selection and locally-weighted partial-least-squares regression.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.inpa.2020.06.002. [61,62,63,64,65,66,67,68,69,70,71,72,73].

REFERENCES

[1] 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.
[2] Clevers JG, Kooistra L, Schaepman ME. Estimating canopy water content using hyperspectral remote sensing data. Int J Appl Earth Obs 2010;12(2):119–25.
[3] Ustin SL, Gamon JA. Remote sensing of plant functional types. New Phytol 2010;186(4):795–816.
[4] Song C. Optical remote sensing of forest leaf area index and biomass. Prog Phys Geog 2013;37(1):98–113.
[5] Verrelst J, Camps-Valls G, Muñoz-Marí J, Rivera JP, Veroustraete F, Clevers JG, et al. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. ISPRS J Photogramm 2015;108:273–90.
[6] 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 Obs 2008;10(3):358–73.
[7] Darvishzadeh R, Skidmore A, Schlerf M, Atzberger C, Corsi F, Cho M. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. ISPRS J Photogramm Sens 2008;63(4):409–26.
[8] Verrelst J, Malenovský Z, Van der Tol C, Camps-Valls G, Gastellu-Etchegorry, JP, Lewis, P, North P, Moreno J. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surv Geophys. 2018; 1–41.
[9] De Jong SM, Pebesma EJ, Lacaze B. Above-ground biomass assessment of Mediterranean forests using airborne imaging spectrometry: the DAIS Peyne experiment. Int J Remote Sens. 2003;24(7):1505–20.
[10] Nguyen HT, Lee BW. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. Eur J Agron. 2006;24(4):349–56.
[11] Atzberger C, Jarmer T, Schlerf M, Kötz B, Werner W. Spectroradiometric determination of wheat bio-physical variables: Comparison of different empirical-statistical approaches. In Rem Sens in Trans, Proc. 23rd EARSeL symposium, Belgium 2003; p. 463-470.
[12] Cho MA, Skidmore A, Corsi F, Van Wieren SE, Sobhan I. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. Int J Appl Earth Obs 2007;9 (4):414–24.
[13] Borin A, Ferrao MF, Mello C, Maretto DA, Poppi RJ. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. Anal Chim Acta 2006;579(1):25–32.
[14] Thissen U, Pepers M, Üstün B, Melssen WJ, Buydens LM. Comparing support vector machines to PLS for spectral regression applications. Chemometr Intell Lab 2004;73 (2):169–79.
[15] Yu H, Lin H, Xu H, Ying Y, Li B, Pan X. Prediction of enological parameters and discrimination of rice wine age using least-squares support vector machines and near infrared spectroscopy. J Agr Food Chem 2008;56(2):307–13.
[16] Siegmann B, Jarmer T. Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. Int J Remote Sens 2015;36(18):4519–34.
[17] Zhou X, Zhu X, Dong Z, Guo W. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. The Crop J 2016;4(3):212–9.
[18] Zhang XM, He GJ, Zhang ZM, Peng Y, Long TF. Spectral-spatial multi-feature classification of remote sensing big data based on a random forest classifier for land cover mapping. Cluster Comput 2017;20(3):2311–21.
[19] Jensen RR, Hardin PJ, Hardin AJ. Estimating urban leaf area index (LAI) of individual trees with hyperspectral data. Photogramm Eng Remote Sens 2012;78(5):495–504.
[20] Neinavaz E, Skidmore AK, Darvishzadeh R, Groen TA. Retrieval of leaf area index in different plant species using thermal hyperspectral data. ISPRS J Photogramm 2016;119:390–401.

[21] Darvishzadeh R, Skidmore A, Schlerf M, Atzberger C. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. Rem Sens Environ 2008;112(5):2592–604.
[22] Agapiou A, Alexakis DD, Sarris A, Hadjimitsis DG. Evaluating the potentials of Sentinel-2 for archaeological perspective. Remote Sens 2014;6(3):2176–94.
[23] Veraverbeke S, Gitas I, Katagis T, Polychronaki A, Somers B, Goossens R. Assessing post-fire vegetation recovery using red–near infrared vegetation indices: accounting for background and vegetation variability. ISPRS J Photogramm 2012;68:28–39.
[24] Tits L, Somers B, Coppin P. The potential and limitations of a clustering approach for the improved efficiency of multiple endmember spectral mixture analysis in plant production system monitoring. IEEE T Geosci Remote 2012;50 (6):2273–86.
[25] Wang FM, Huang JF, Wang XZ. Identification of optimal hyperspectral bands for estimation of rice biophysical parameters. J Integr Plant Biol 2008;50(3):291–9.
[26] Goel NS. Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. Remote Sens Rev 1988;4(1):1–212.
[27] Nocita M, Stevens A, Toth G, Panagos P, van Wesemael B, Montanarella L. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. Soil Biol Biochem 2014;68:337–47.
[28] Sarathjith MC, Das BS, Wani SP, Sahrawat KL. Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. Geoderma 2016;267:1–9.
[29] Gupta A, Vasava HB, Das BS, Choubey AK. Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. Geoderma 2018;325:59–71.
[30] Gitelson AA, Viña A, Arkebauer TJ, Rundquist DC, Keydan G, Leavitt B. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys Res Lett 2003:30(5).
[31] Penuelas J, Gamon JA, Griffin KL, Field CB. Assessing community type, plant biomass, pigment composition, and photosynthetic efficiency of aquatic vegetation from spectral reflectance. Remote Sens Environ 1993;46(2):110–8.
[32] Hunt Jr ER, Rock BN. Detection of changes in leaf water content using near-and middle-infrared reflectances. Remote Sens Environ 1989;30(1):43–54.
[33] Jackson TJ, Chen D, Cosh M, Li F, Anderson M, Walthall C, et al. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. Remote Sens Environ 2004;92(4):475–82.
[34] Tucker CJ. Remote sensing of leaf water content in the near infrared. Remote Sens Environ 1980;10(1):23–32.
[35] Claudio HC, Cheng Y, Fuentes DA, Gamon JA, Luo H, Oechel W, et al. Monitoring drought effects on vegetation water content and fluxes in chaparral with the 970 nm water band index. Remote Sens Environ 2006;103(3):304–11.
[36] Haboudane D, Miller JR, Pattey E, Zarco-Tejada PJ, Strachan IB. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. Remote Sens Environ 2004;90(3):337–52.
[37] Pimstein A, Karnieli A, Bonfil D. Wheat and maize monitoring based on ground spectral measurements and multivariate data analysis. J Appl Remote Sens 2007;1(1) 013530.
[38] Dreccer MF, Barnes LR, Meder R. Quantitative dynamics of stem water soluble carbohydrates in wheat can be monitored in the field using hyperspectral reflectance. Field Crop Res 2014;159:70–80.
[39] Jarmer T. Spectroscopy and hyperspectral imagery for monitoring summer barley. Int J Remote Sens 2013;34 (17):6067–78.
[40] Kiala Z, Odindi J, Mutanga O. Potential of interval partial least square regression in estimating leaf area index. S Afr J Sci 2017;113(9–10):1–9.
[41] Mirzaie M, Darvishzadeh R, Shakiba A, Matkan AA, Atzberger C, Skidmore A. Comparative analysis of different uni-and multi-variate methods for estimation of vegetation water content using hyper-spectral measurements. Int J Appl Earth Obs 2014;26:1–11.
[42] Li L, Cheng YB, Ustin S, Hu XT, Riaño D. Retrieval of vegetation equivalent water thickness from reflectance using genetic algorithm (GA)-partial least squares (PLS) regression. Adv Space Res 2008;41(11):1755–63.
[43] Staff S. Keys to soil taxonomy. (Eighth edition): United States Department of Agriculture, Soil Conservation Service: Washington, DC. 1998.
[44] Daughtry CS. Direct measurements of canopy structure. Remote Sens Rev 1990;5(1):45–60.
[45] Savitzky A, Golay MJ. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 1964;36 (8):1627–39.
[46] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004;14(3):199–222.
[47] Breiman L. Manual on setting up, using, and understanding random forests v3.1: Statistics Department University of California Berkeley, CA, USA, 2002; 1–58.
[48] Hazama K, Kano M. Covariance-based locally weighted partial least squares for high-performance adaptive modeling. Chemometr Intell Lab 2015;146:55–62.
[49] Santra P, Sahoo RN, Das BS, Samal RN, Pattanaik AK, Gupta VK. Estimation of soil hydraulic properties using proximal spectral reflectance in visible, near-infrared, and shortwave-infrared (VIS–NIR–SWIR) region. Geoderma 2009;152 (3):338–49.
[50] Viscarra Rossel RA, Lark RM. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. Eur J Soil Sci 2009;60(3):453–64.
[51] Rossel RV, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 2006;131(1):59–75.
[52] Efron B, Tibshirani R. Improvements on cross-validation: the 632+ bootstrap method. J Am Stat Assoc 1997;92(438):548–60.
[53] Yang X, Huang J, Wu Y, Wang J, Wang P, Wang X, et al. Estimating biophysical parameters of rice with remote sensing data using support vector machines. Sci China Life Sci 2011;54(3):272–81.
[54] Wang FM, Huang JF, Zhou QF, Wang XZ. Optimal waveband identification for estimation of leaf area index of paddy rice. J Zhejiang Univ Sci B 2008;9(12):953–63.
[55] Aschonitis VG, Papamichail DM, Lithourgidis A, Fano EA. Estimation of leaf area index and foliage area index of rice using an indirect gravimetric method. Commun Soil Sci Plan 2014;45(13):1726–40.
[56] Zhang J, Wu J, Zhou L. Deriving vegetation leaf water content from spectrophotometric data with orthogonal signal correction-partial least square regression. Int J Remote Sens 2011;32(22):7557–74.
[57] Pasqualotto N, Delegido J, Van Wittenberghe S, Verrelst J, Rivera JP, Moreno J. Retrieval of canopy water content of different crop types with two new hyperspectral indices: Water absorption area index and depth water index. Int J Appl Earth Obs 2018;67:69–78.
[58] Jackson RD, Pinter Jr PJ. Spectral response of architecturally different wheat canopies. Remote Sens Environ 1986;20 (1):43–56.

[59] Asner GP. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens Environ 1998;64 (3):234–53.
[60] Wang FM, Huang JF, Lou ZH. A comparison of three methods for estimating leaf area index of paddy rice from optimal hyperspectral bands. Precis Agric 2011;12(3):439–47.
[61] Rodríguez-Pérez JR, Ordóñez C, González-Fernández AB, Sanz-Ablanedo E, Valenciano JB, Marcelo V. Leaf water content estimation by functional linear regression of field spectroscopy data. Biosyst Eng 2018;165:36–46.
[62] Liu M, Liu X, Ding W, Wu L. Monitoring stress levels on rice with heavy metal pollution from hyperspectral reflectance data using wavelet-fractal analysis. Int J Appl Earth Obs 2011;13(2):246–55.
[63] Shi T, Liu H, Wang J, Chen Y, Fei T, Wu G. Monitoring arsenic contamination in agricultural soils with reflectance spectroscopy of rice plants. Environ Sci Technol 2014;48 (11):6264–72.
[64] Zhang F, Zhou G. Estimation of canopy water content by means of hyperspectral indices based on drought stress gradient experiments of maize in the north plain China. Remote Sens 2015;7(11):15203–23.
[65] Ali AM, Darvishzadeh R, Skidmore AK, van Duren I, Heiden U, Heurich M. Estimating leaf functional traits by inversion of PROSPECT: Assessing leaf dry matter content and specific leaf area in mixed mountainous forest. Int J Appl Earth Obs 2016;45:66–76.
[66] Darvishzadeh R, Skidmore AK, Mirzaie M, Atzberger C, Schlerf M. Fresh biomass estimation in heterogeneous grassland using hyperspectral measurements and multivariate statistical analysis. In AGU Fall Meeting Abstracts, Vol. 1, No. 7, 2014.
[67] Clevers JGPW, Van der Heijden GW, Verzakov S, Schaepman ME. Estimating grassland biomass using SVM band shaving of hyperspectral data. Photogramm Eng Remote Sens 2007;73 (10):1141–8.
[68] Atzberger C, Guérif M, Baret F, Werner W. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. Comput Electron Agric 2010;73 (2):165–73.
[69] Jin X, Xu X, Song X, Li Z, Wang J, Guo W. Estimation of leaf water content in winter wheat using grey relational analysis–partial least squares modeling with hyperspectral data. Agron J 2013;105(5):1385–92.
[70] Kooistra L, Leuven RS, Wehrens R, Nienhuis PH, Buydens LM. A comparison of methods to relate grass reflectance to soil metal contamination. Int J Remote Sens 2003;24 (24):4995–5010.
[71] Kooistra L, Salas EA, Clevers JG, Wehrens R, Leuven RS, Nienhuis PH, et al. Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains. Environ Pollut 2004;127(2):281–90.
[72] Li X, Zhang Y, Bao Y, Luo J, Jin X, Xu X, et al. Exploring the best hyperspectral features for LAI estimation using partial least squares regression. Remote Sens 2014;6(7):6221–41.
[73] Panigrahi N, Das BS. Canopy spectral reflectance as a predictor of soil water potential in rice. Water Resour Res. 2018;54(4):2544–60.