

Article

Exploring the Best Hyperspectral Features for LAI Estimation Using Partial Least Squares Regression

Xinchuan Li ¹, Youjing Zhang ^{1,*}, Yansong Bao ², Juhua Luo ³, Xiuliang Jin ⁴, Xingang Xu ⁴, Xiaoyu Song ⁴ and Guijun Yang ⁴

¹ School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China; E-Mail: rs_lxc@126.com

² Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science and Technology, Nanjing 210044, China; E-Mail: ysbao@hotmail.com

³ State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China; E-Mail: luojuhua@126.com

⁴ National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; E-Mails: jinxiuxiuliang@126.com (X.J.); xxgpaper@126.com (X.X.); songxy@nercita.org.cn (X.S.); guijun.yang@163.com (G.Y.)

* Author to whom correspondence should be addressed; E-Mail: zhangyj@hhu.edu.cn; Tel./Fax: +86-25-8378-7234.

Received: 6 March 2014; in revised form: 15 June 2014 / Accepted: 16 June 2014 /

Published: 1 July 2014

Abstract: The use of spectral features to estimate leaf area index (LAI) is generally considered a challenging task for hyperspectral data. In this study, the hyperspectral reflectance of winter wheat was selected to optimize the selection of spectral features and to evaluate their performance in modeling LAI at various growth stages during 2008 and 2009. We extracted hyperspectral features using different techniques, including reflectance spectra and first derivative spectra, absorption and reflectance position and vegetation indices. In order to find the best subset of features with the best predictive accuracy, partial least squares regression (PLSR) and variable importance in projection (VIP) were applied to estimated LAI values. The results indicated that the red edge–NIR spectral region (680 nm–1300 nm) was the most sensitive to LAI. Most features in this region exhibited a high correlation with LAI and had higher VIP values, especially the first derivative waveband at 750 nm ($r = 0.900$, $VIP = 1.144$). Adding a large number of features would not

significantly improve the accuracy of the PLSR model. The PLSR model based on the fourteen features with the highest VIP values predicted LAI with a mean bootstrapped R^2 value of 0.880 and a mean RMSE of 0.943 on the validation dataset and produced an estimated LAI result better than that, including the entire 54-feature dataset with a mean R^2 of 0.875 and a mean RMSE of 0.965. The results of this study thus suggest that the use of only a few of the best features by VIP values is sufficient for LAI estimation.

Keywords: hyperspectral remote sensing; leaf area index (LAI); spectral feature; partial least squares regression

1. Introduction

Defined as one-half of the total green leaf area (all-sided) per unit ground surface area [1], the leaf area index (LAI) is an important parameter and is commonly applied in environmental studies examining growth monitoring, yield estimation, evapotranspiration, radiation extinction, carbon cycling and climate [2–5]. Direct measurement of the LAI via the use of destructive field measurements is extremely labor intensive, tedious and limited to experimental plots. Remote sensing techniques have been recognized as a reliable method to provide a fast, non-destructive and relatively cheap way to measure LAI on different scales [6,7].

Hyperspectral remote sensing can produce hundreds or even thousands of narrow, contiguous spectral bands, which may provide crucial additional information, potentially representing a significant improvement over broad bands in quantifying biophysical and biochemical variables, such as LAI [8,9]. However, hyperspectral data are much more complex than multispectral data. Although they provide a vast amount of information, most adjacent wavebands are redundant and often highly correlated [10]. It is therefore essential to determine the best spectral features derived from hyperspectral data in order to accurately quantify LAI.

Various methods have been proposed, applied and improved in recent decades for the extraction of the spectral features of hyperspectral information. These selection techniques can be classified into three broad groups. (i) Waveband features: Compared with broad bands, narrow bands in specific portions of the spectrum are known to improve discrimination capabilities for various vegetation. Thenkabail *et al.* [10] determined 22 optimal hyperspectral wavebands with which to best characterize vegetation and agricultural crops over the spectral range of 400–2500 nm. Becker *et al.* [11] used second-derivative analysis to identify eight optimal spectral bands in the visible-NIR wavelength region that appeared to contain the majority of the coastal wetland information content of the full spectral resolution. Wang *et al.* [12] employed three methods, including correlation coefficient-based, vegetation index-based and the stepwise regression method to select 15 suitable wavebands for paddy rice LAI estimation. (ii) Spectral position features: Reflectance and absorption features that characterize hyperspectral data are also related to specific physical and chemical crop characteristics [13]. Pu *et al.* [14] found strong correlations between forest LAI and various red-edge parameters, including the red-edge position (REP) and the red-well position (RWP). Spectral features in the shortwave infrared (SWIR) regions (as well as those in the near-infrared) are also important in predicting LAI [9,15]. (iii) Vegetation indices: Spectral

vegetation indices are mathematical combinations of different spectral bands, mostly in the visible and near-infrared regions of the electromagnetic spectrum. Although the normalized difference vegetation index (NDVI) is by far the most well-known and widely used method of estimating LAI [16], it is sensitive to soil and saturates at a relatively low LAI level. In contrast, VIs, such as the soil adjusted vegetation index (SAVI; [17]), second modified SAVI (MSAVI2; [18]), renormalized difference vegetation index (RDVI; [19]) and triangular vegetation index (TVI; [20]), have now been devised to improve LAI estimation. It also has been demonstrated that VIs with red-edge bands are good predictors of widely variable green LAI, such as $CI_{Red-edge}$ [21,22].

The above-mentioned studies have made important progresses in detecting canopy information via the use of hyperspectral remote sensing. However, the research that systematically summarizes and analyzes the different spectral features of hyperspectral remote sensing data in terms of their performance in estimating LAI is rare, with the analysis of a single feature typically not sufficient to explore such rich information. Several studies have focused on statistical techniques, such as stepwise multiple linear regression (SMLR), which makes use of the information provided by several spectral features to estimate biochemical and biophysical vegetation properties [23,24]. In either case, multi-collinearity is a common problem inherent to hyperspectral datasets [25].

Partial least squares regression (PLSR) is a data compression technique that reduces a large number of collinear variables to a few non-correlated latent variables or factors [26–28]. A number of studies have shown that PLSR is a powerful tool able to extract significant signals and to create reliable models [8,29–32], and it has the potential to accurately predict LAI. Although these previous studies employed all available spectral wavelengths simultaneously for PLSR, others have revealed that the use of only a few features is sufficient to extract and discriminate essential information and characteristics [10]. As the use of full spectral subsets or the greatest available amount of spectral information would likely not improve retrieval performance, but simply increase computation time [24], it may therefore be more effective to obtain the most accurate biophysical vegetation data possible to build the model.

The objectives of this study were to: (1) systematically summarize the spectral features of hyperspectral canopy reflectance in terms of three aspects: feature wavebands, feature positions and vegetation indices; (2) evaluate every feature's potential for LAI estimation; and (3) identify optimal features (and their numbers) for LAI estimation via PLSR.

2. Materials and Methods

2.1. Study Area

The study was conducted in suburban Beijing, China, in an area characterized by a northern temperate monsoon climate. The experimental fields were located in Tongzhou District (39°36'–40°02'N, 116°32'–116°56'E) and Shunyi District (40°00'–40°18'N, 116°28'–116°58'E) on flat terrain, with the predominant soil texture being a fine clay loam. The types of wheat analyzed included Nongda 211 (erectophile), Zhongyou 206 (middle), Jingdong 8 (middle) and Jing 9428 (planophile).

2.2. In situ Data Collection

The experiment was carried out during different years and growth stages of winter wheat, including: 1 April 2008, and 2 April 2009 (tillering); 15 April 2008, and 13 April 2009 (jointing); 29 April 2008, and 30 April 2009 (heading); and 15 May 2008, and 18 May 2009 (anthesis). For each experimental point, winter wheat canopy spectral measurements were taken from 0.5 m × 0.5 m plots randomly selected in the central 30 m × 30 m field.

An ASD FieldSpec Pro spectrometer (Analytical Spectral Devices, Boulder, CO, USA) was used to measure canopy reflected radiances in the field. This instrument records reflectance in the range between 350 and 1050 nm, with a sampling interval of 1.40 nm, and between 1050 and 2500 nm, with a sampling interval of 2 nm. Spectral data were interpolated to a spectral band width of 1 nm using the ASD software. The instrument also has a 25° fiber optic field of view. Spectral measurements operated at a nadir of 1.3 m above the winter wheat canopy. All spectral measurements were made during cloud-free periods ±2 h from solar noon, between 10:00 a.m. and 2:00 p.m. local time. Prior to each reflectance measurement, the radiance of a 40 cm × 40 cm BaSO₄ standard panel was recorded under the same illumination conditions to convert the spectral radiance measurements into reflectance. Vegetation reflectance measurements were obtained by averaging 20 scans at optimized integration times. Due to severe noise associated with water absorption, the spectral regions of 1350–1480 nm, 1781–1990 nm and 2400–2500 nm were also excluded from the analysis. A moving Savitzky–Golay filter [33] with a frame-size of 17 data points and a second polynomial was employed to smooth the spectra.

All of the plants within a 0.25 m² area of winter wheat in each experimental point were harvested immediately after the spectral measurement, with the samples placed in black plastic bags and transported to the laboratory for subsequent analysis. In the lab, plants were dissected into green leaves, dead leaves, stems and roots. The green leaves were used to measure the leaf area by a leaf area meter (Li-Cor 3100, LICOR, Inc., Lincoln, NE, USA). Table 1 shows summary statistics of green LAI measured. The LAI range for the calibration dataset in 2009 is from 0.40 to 7.49, with the average of 3.07 and a standard deviation of 1.86. Similarly, the statistical parameters for the validation dataset in 2008 are 0.21–8.85, 3.69 and 2.45, respectively.

Table 1. Descriptive statistics of LAI in two years' data.

Dataset	Year	Samples	Max	Min	Mean	Standard Deviation
Calibration dataset	2009	76	7.49	0.40	3.07	1.86
Validation dataset	2008	71	8.85	0.21	3.69	2.45

2.3. Spectral Features of Hyperspectral Information

2.3.1. Spectral Waveband Features

Both reflectance spectra and first derivative spectra were studied in order to select waveband features. The first derivative of reflectance was calculated from each reflectance spectrum using the following equation:

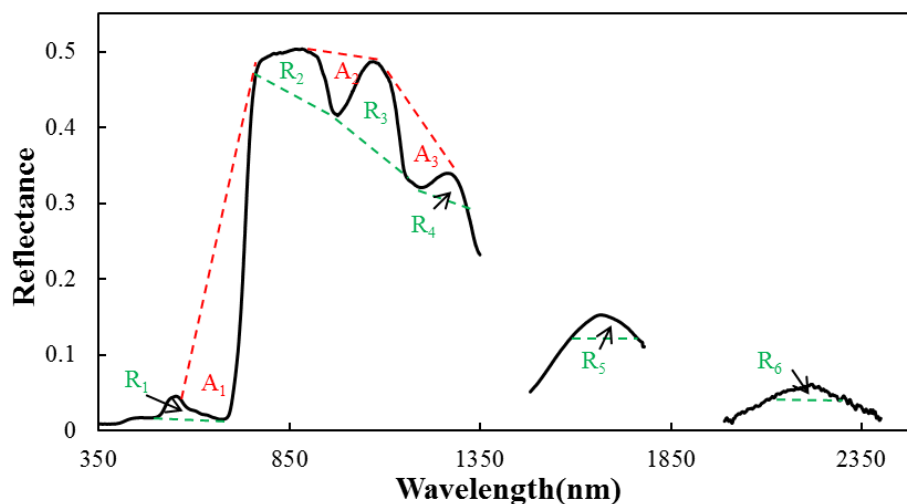
$$\text{FDS}_{\lambda(i)} = (R_{\lambda(j+1)} - R_{\lambda(j)}) / \Delta\lambda \quad (1)$$

where FDS is the first derivative of reflectance at wavelength midpoint i between wavebands j and $j + 1$, $R_{\lambda(j)}$ is the reflectance at waveband j , $R_{\lambda(j+1)}$ is the reflectance at waveband $j + 1$ and $\Delta\lambda$ the difference in wavelength between j and $j + 1$.

2.3.2. Spectral Position Features

Figure 1 shows three absorption (560–760 nm, 920–1080 nm and 1120–1280 nm) and six reflectance positions (500–670 nm, 780–970 nm, 980–1200 nm, 1200–1350 nm, 1480–1720 nm and 2000–2300 nm) for winter wheat, used for the extraction of absorption and reflectance feature parameters. Continuum removal was applied throughout the full spectrum in order to enhance the differences in absorption and reflectance [34]. The three parameters proposed were employed in the present study at every absorption and reflectance position: (1) depth; (2) area; and (3) normalized depth [35,36].

Figure 1. Three absorption (A) and six reflectance (R) characteristics of the wheat canopy spectrum.



For the three absorption positions, each absorption depth feature (A_Depth_i) was defined as:

$$A_Depth_i = 1 - R'_i(\lambda_{min}) = 1 - \frac{R_i(\lambda_{min})}{R_{ci}(\lambda_{min})} \quad (2)$$

where continuum-removed reflectance $R'_i(\lambda_{min})$ is obtained by dividing the minimum reflectance value $R_i(\lambda_{min})$ in the absorption position by the continuum line $R_{ci}(\lambda_{min})$ at the corresponding wavelength and i is the number of absorption positions ($i = 1, 2, 3$).

The absorption area feature (A_Area_i) was calculated as the area bounded by the reflectance spectrum and continuum line in each absorption region as follows:

$$A_Area_i = \int_{\lambda_{Si}}^{\lambda_{Ei}} (R_{ci}(\lambda) - R_i(\lambda)) d\lambda \quad (3)$$

where $R_{ci}(\lambda)$ and $R_i(\lambda)$ are, respectively, the reflectance of the continuum line and reflectance at the corresponding wavelength λ in the absorption region and λ_{Si} and λ_{Ei} are the start and end wavelengths, respectively, in each absorption region.

The normalized absorption depth (A_ND_i) was calculated by dividing the absorption depth feature by the absorption area feature, as follows:

$$A_ND_i = \frac{A_Depth_i}{A_Area_i} \quad (4)$$

For the six reflectance positions, each reflectance depth feature (R_Depth_i) was defined as:

$$R_Depth_i = 1 - R'_i(\lambda_{max}) = 1 - \frac{R_{ci}(\lambda_{max})}{R_i(\lambda_{max})} \quad (5)$$

where the continuum-removed reflectance $R_i(\lambda_{max})$ is obtained by dividing the maximum reflectance value $R_i(\lambda_{max})$ in the reflectance position by the inner continuum line $R_{ci}(\lambda_{max})$ at the corresponding wavelength and i is the number of reflectance positions ($i = 1, 2, 3, 4, 5, 6$).

The reflectance area feature (R_Area_i) was calculated as the area bounded by the reflectance spectrum and inner continuum line in the reflectance region as follows:

$$R_Area_i = \int_{\lambda_{Si}}^{\lambda_{Ei}} (R_i(\lambda) - R_{ci}(\lambda)) d\lambda \quad (6)$$

where $R_{ci}(\lambda)$ and $R_i(\lambda)$ are the reflectance of the inner continuum line and reflectance, respectively, at the corresponding wavelength λ in the reflectance region and λ_{Si} and λ_{Ei} are, respectively, the start and end wavelengths in each reflectance region.

The normalized reflectance depth (R_ND_i) was calculated by dividing the reflectance depth feature by the reflectance area feature, as follows:

$$R_ND_i = \frac{R_Depth_i}{R_Area_i} \quad (7)$$

After a literature review, other spectral position-based variables obtained from first derivative spectra [37] were also used in this study and are listed in Table 2.

Table 2. Spectral features derived from first derivative spectra.

Title	Definition and Description
Db	Maximum value of 1st derivative with blue edge (490–530 nm)
λ_b	Wavelength at Db
Dy	Maximum value of 1st derivative with yellow edge (550–582 nm)
λ_y	Wavelength at Dy
Dr	Maximum value of 1st derivative with red edge (680–780 nm)
λ_r	Wavelength at Dr
Rg	Maximum reflectance with green peak (510–560 nm)
λ_g	Wavelength at Rg
Ro	Lowest reflectance with red well (640–680 nm)
λ_o	Wavelength at Ro
SDb	Sum of 1st derivative values within blue edge
SDy	Sum of 1st derivative values within yellow edge
SDr	Sum of 1st derivative values within red well

2.3.3. Vegetation Index Features

Many different optical indices have been reported in the literature and have proven to be well correlated with vegetation parameters. A total of twenty-four vegetation indices were employed in the

present study (Table 3 [16–20,38–56]). The atmospherically-resistant vegetation index (ARVI), DVI, EVI, green normalized difference vegetation index (GNDVI), modified nonlinear vegetation index (MNLI), MSAVI2, modified simple ratio (MSR), NDVI, nonlinear vegetation index (NLI), optimization of SAVI (OSAVI), RDVI, ratio vegetation index (RVI), SAVI, three gradient difference vegetation index (TGDVI), TVI and modified triangular vegetation index (MTVI2) were calculated using simulated reflectance bands of the Moderate Resolution Imaging Spectrometer (MODIS) (blue: 459–479 nm; green: 545–565 nm; red: 620–670 nm; NIR: 841–876 nm). The reflectance spectra were also resampled to the spectral bands of the Medium Resolution Imaging Spectrometer (MERIS) (green: 555–565 nm; red: 660–670 nm; red edge: 703–712 nm; NIR: 750–760 nm) using the MERIS spectral response function for calculating red edge indices, such as red edge NDVI, $CI_{Red-edge}$ and MTCI. As for the SWIR region, water indices (WI, NDWI, normalized difference infrared index (NDII), disease water stress index (DSWI) and standardized LAI-determining index (sLAIDI*)) were also tested.

Table 3. Vegetation indices compiled from the literature.

Vegetation Index	Formulas	Ref.
Atmospherically-resistant vegetation index (ARVI)	$ARVI = \frac{\rho_{NIR} - RB}{\rho_{NIR} + RB}$ $RB = R - \gamma(B - R), \gamma = 1$	[38]
Difference vegetation index (DVI)	$DVI = \rho_{NIR} - \rho_R$	[39]
Enhanced vegetation index (EVI)	$EVI = 2.5 \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + 6\rho_R - 7.5\rho_B + 1}$	[40]
Green normalized difference vegetation index (GNDVI)	$GNDVI = \frac{\rho_{NIR} - \rho_G}{\rho_{NIR} + \rho_G}$	[41]
Modified nonlinear vegetation index (MNLI)	$MNLI = \frac{1.5(\rho_{NIR}^2 - \rho_R)}{\rho_{NIR}^2 + \rho_R + 0.5}$	[42]
The second modified SAVI (MSAVI2)	$MSAVI2 = \frac{2\rho_{NIR} + 1 - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_R)}}{2}$	[18]
Modified simple ratio (MSR)	$MSR = \frac{\rho_{NIR}/\rho_R - 1}{\sqrt{\rho_{NIR}/\rho_R} + 1}$	[43]
Normalized difference vegetation index (NDVI)	$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}$	[44]
Nonlinear vegetation index (NLI)	$NLI = \frac{\rho_{NIR}^2 - \rho_R}{\rho_{NIR}^2 + \rho_R}$	[45]
Optimization of soil-adjusted vegetation index (OSAVI)	$OSAVI = (1 + 0.16) \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R + 0.16}$	[46]
Renormalized difference vegetation index (RDVI)	$RDVI = \frac{\rho_{NIR} - \rho_R}{\sqrt{\rho_{NIR} + \rho_R}}$	[19]
Ratio vegetation index (RVI)	$RVI = \frac{\rho_{NIR}}{\rho_R}$	[47]
Soil-adjusted vegetation index (SAVI)	$SAVI = \frac{\rho_{NIR} - \rho_R}{(\rho_{NIR} + \rho_R + L)} (1 + L), L = 0.5$	[17]

Table 3. Cont.

Vegetation Index	Formulas	Ref.
Three gradient difference vegetation index (TGDVI)	$TGDVI = \frac{\rho_{NIR} - \rho_R}{\lambda_{NIR} - \lambda_R} - \frac{\rho_R - \rho_G}{\lambda_R - \lambda_G}$	[48]
Triangular vegetation index (TVI)	$TVI = 60(\rho_{NIR} - \rho_G) - 100(\rho_R - \rho_G)$	[20]
Modified triangular vegetation index (MTVI2)	$MTVI2 = \frac{1.5[(1.2(\rho_{NIR} - \rho_G) - 2.5(\rho_R - \rho_G))]}{\sqrt{(2\rho_{NIR} + 1)^2 - (6\rho_{NIR} - 5\sqrt{\rho_R})} - 0.5}$	[16]
Red edge NDVI	$NDVI_{Red-edge} = \frac{\rho_{NIR} - \rho_{Red-edge}}{\rho_{NIR} + \rho_{Red-edge}}$	[49]
Red-edge chlorophyll index	$CI_{Red-edge} = \frac{\rho_{NIR}}{\rho_{Red-edge}} - 1$	[50]
MERIS Terrestrial Chlorophyll Index	$MTCI = \frac{\rho_{NIR} - \rho_{Red-edge}}{\rho_{Red-edge} - \rho_{Red}}$	[51]
Water Index (WI)	$WI = \rho_{900} / \rho_{970}$	[52]
Normalized difference water index (NDWI)	$NDWI = \frac{\rho_{860} - \rho_{1240}}{\rho_{860} + \rho_{1240}}$	[53]
Normalized difference infrared index (NDII)	$NDII = \frac{\rho_{819} - \rho_{1600}}{\rho_{819} + \rho_{1600}}$	[54]
Disease water stress index (DSWI)	$DSWI = \frac{\rho_{803} + \rho_{549}}{\rho_{1659} + \rho_{681}}$	[55]
Standardized LAI-determining index (sLAIDI*)	$sLAIDI^* = s \cdot \left(\frac{\rho_{1050} - \rho_{1250}}{\rho_{1050} + \rho_{1250}} \right) \rho_{1555}, s = 1$	[56]

Notes: ρ_{NIR} , $\rho_{Red-edge}$, ρ_{Red} , ρ_{green} and ρ_{blue} are the reflectance in spectral bands of the near-infrared, red, green and blue light bands, respectively; ρ_i denotes reflectance at the band i wavelength (nanometers).

2.4. Partial Least Squares Regression

The partial least squares regression technique generalizes and combines features of principal component regression (PCR) and multiple linear regression. The method is recognized as a powerful modeling tool with which to model relations when the number of predictor variables is large and the collinearity among the variables is strong [57]. The aim of PLSR is to build a linear model as follows:

$$Y = X\beta + \varepsilon \quad (8)$$

where Y is a mean-centered vector of a dependent variable, X is a mean-centered matrix of independent variables, β is a matrix of regression coefficients, and ε is a matrix of residuals [28,58].

The optimal number of factors in PLSR analysis is determined by minimizing the prediction residual error sum of squares (PRESS) statistic. Here, the PRESS statistic is calculated via cross-validation (CV) prediction for each model [4,59]. The root mean squared error of cross validation (RMSCV) is also used to assess the predictive abilities of the PLS models.

In order to evaluate the relative importance of variables in the PLSR model, their variable importance in projection (VIP) scores were computed. VIP calculates the contribution of independent variables to

the contribution of the dependent variable, with the most influential predictors in the model selected according to the magnitude of their values [60]. VIP scores serve as an apt measure for identifying individual waveband importance and for providing valuable insight into the most effective spectral regions [60,61]. All calculations were carried out using MATLAB software (The Math Works, Inc., Natick, MA, USA).

2.5. Calibration and Validation

The 76 samples collected in 2009 were used as a calibration data set and the 71 samples collected in 2008 as an independent validation data set. Regression analyses were performed on the former, with the latter then employed to conduct an empirical validation of the regression models. The performance of the different PLSR models was compared using the coefficient of determination (R^2), root mean square error (RMSE) and normalized mean bias (NMB). Definitions of each metric are given below:

$$R^2 = \frac{\sum_{i=0}^n (P_i - \bar{P})^2 \times (Q_i - \bar{Q})^2}{\sum_{i=0}^n (P_i - \bar{P})^2 \times \sum_{i=0}^n (Q_i - \bar{Q})^2} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=0}^n (P_i - Q_i)^2} \quad (10)$$

$$NMB = \frac{\sum_{i=1}^n (P_i - Q_i)}{\sum_{i=1}^n Q_i} \times 100\% \quad (11)$$

where P is the estimated values, Q represents the measured values and n is the total number of samples.

The bootstrap process was performed on the independent validation data to assess the robustness of the regression models. The validation data set was bootstrapped with replacement for $n = 1000$ times. The mean and 95% confidence levels of R^2 values, as well as RMSE values for the validation data were calculated and recorded [62,63].

3. Results

3.1. Optimal Spectral Features in Three Datasets

3.1.1. Optimal Spectral Waveband Feature Dataset

Spectral reflectance data and their first derivative spectra were correlated with measured LAI values using both 2008 and 2009 data sets (Figure 2). Figure 2a reveals negative correlation coefficients in the blue (503 nm), red (661 nm) and shortwave infrared regions (1990–2400 nm) and a positive correlation coefficient in the near infrared region (740–1300 nm). The strongest correlations occur at 439 nm, 554 nm, 750 nm, 948 nm, 1030 nm, 1147 nm, 1236 nm, 1295 nm, 1602 nm and 1745 nm for the first derivative spectra (Figure 2b). The ten best wavebands in both reflectance spectra and first derivative spectra were then selected for establishing an optimal spectral waveband dataset (Table 4).

Figure 2. Correlation analysis of reflectance spectra (a) and first derivative spectra (b) against LAI for each wavelength.

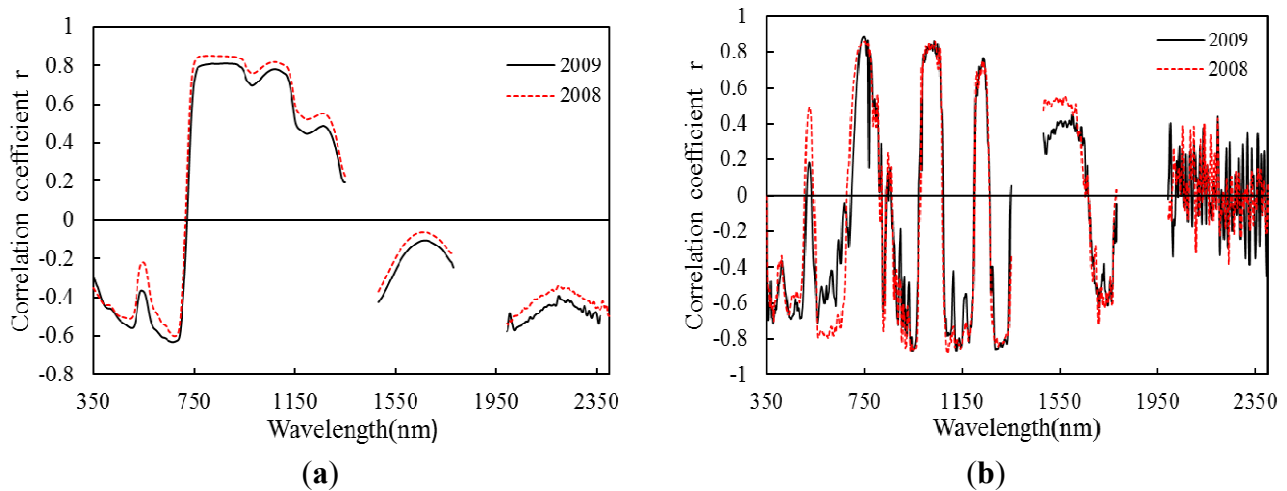


Table 4. Correlations between optimal spectral waveband features and LAI in different years.

Reflectance Spectra			First Derivative Spectra		
Number	Wavelength	2009-r/2008-r	Number	Wavelength	2009-r/2008-r
SP1	503 nm	−0.560 **/−0.558 **	FD1	439 nm	−0.670 **/−0.679 **
SP2	661 nm	−0.629 **/−0.649 **	FD2	554 nm	−0.688 **/−0.761 **
SP3	770 nm	0.793 **/0.873 **	FD3	750 nm	0.891 **/0.908 **
SP4	868 nm	0.810 **/0.877 **	FD4	948 nm	−0.850 **/−0.884 **
SP5	938 nm	0.789 **/0.860 **	FD5	1030 nm	0.836 **/0.824 **
SP6	1072 nm	0.796 **/0.844 **	FD6	1147 nm	−0.846 **/−0.873 **
SP7	1263 nm	0.499 **/0.557 **	FD7	1236 nm	0.732 **/0.678 **
SP8	1993 nm	−0.569 **/−0.595 **	FD8	1295 nm	−0.857 **/−0.863 **
SP9	2022 nm	−0.564 **/−0.558 **	FD9	1602 nm	0.505 **/0.459 **
SP10	2398 nm	−0.558 **/−0.547 **	FD10	1745 nm	−0.602 **/−0.517 **

2009-r and 2008-r: the correlation coefficients between features and LAI for the calibration data set (in 2009) and the validation data set (in 2008), respectively; ** correlation significant at the 0.01 level.

3.1.2. Optimal Spectral Position Feature Dataset

A total of 40 spectral position features were calculated using absorption position features, reflectance position features and other spectral position-based variables. Correlation analysis was carried out between these 40 spectral position features and LAI values (Table 5), with the results revealing A_Area_2 to exhibit the strongest correlation with LAI ($r = 0.854$). The top 10 spectral position features were then selected in order to establish an optimal spectral position feature dataset: R_Area_2 , R_Area_3 , R_Area_4 , A_Area_1 , A_ND_1 , A_Area_2 , A_Area_3 , A_ND_3 , Dr and SDr .

3.1.3. Vegetation Index Dataset

We studied the correlations between 24 vegetation indices (VIs) and LAI (Table 6), with the results indicating significant relationships existing between LAI and all spectral parameters. A vegetation index dataset was then established based on these vegetation indices.

Table 5. Correlations between spectral position features and LAI in different years.

Spectral Position	2009-r/ 2008-r	Spectral Position	2009-r /2008-r	Spectral Position	2009-r /2008-r	Spectral Position	2009-r /2008-r
R_Depth ₁	0.682 **/0.762 **	R_Area ₄	0.786 **/0.846 **	A_ND ₁	−0.765 **/−0.825 **	λb	0.009/−0.040
R_Area ₁	−0.004/0.390 **	R_ND ₄	−0.372 **/−0.143	A_Depth ₂	0.655 **/0.783 **	λy	0.002/0.182
R_ND ₁	0.467 **/0.276	R_Depth ₅	0.641 **/0.603 **	A_Area ₂	0.854 **/0.903 **	λr	0.602 **/0.658 **
R_Depth ₂	0.502 **/0.674 *	R_Area ₅	0.400 **/0.582 **	A_ND ₂	−0.663 **/−0.775 **	Rg	−0.382 **/−0.229 *
R_Area ₂	0.825 **/0.904 **	R_ND ₅	0.128/0.049	A_Depth ₃	0.599 **/0.723 **	λg	−0.629 **/−0.610 **
R_ND ₂	−0.466 **/−0.280	R_Depth ₆	0.612 **/0.596 **	A_Area ₃	0.830 **/0.890 **	Ro	−0.629 **/−0.600 **
R_Depth ₃	0.639 **/0.535 **	R_Area ₆	0.021/0.311 **	A_ND ₃	−0.716 **/−0.857 **	λo	0.130/0.102
R_Area ₃	0.849 **/0.900 **	R_ND ₆	0.389 **/0.286 *	Db	0.172/0.489 **	SDB	0.014/0.340 **
R_ND ₃	−0.643 **/−0.519 **	A_Depth ₁	0.703 **/0.760 **	Dy	−0.521 **/−0.314 **	SDy	−0.621 **/−0.784 **
R_Depth ₄	0.465 **/0.507 **	A_Area ₁	0.844 **/0.906 **	Dr	0.825 **/0.841 **	SDr	0.835 **/0.862 **

2009-r and 2008-r: the correlation coefficients between features and LAI for the calibration data set (in 2009) and the validation data set (in 2008), respectively; * correlation significant at the 0.05 level; ** correlation significant at the 0.01 level.

Table 6. Correlation between vegetation indices and LAI in different years.

VIs	2009-r/2008-r	VIs	2009-r/2008-r	VIs	2009-r/2008-r
NDVI	0.724 **/0.768 **	MSR	0.840 **/0.901 **	NDVI _{Red-edge}	0.712 **/0.763 **
RVI	0.857 **/0.906 **	NLI	0.772 **/0.827 **	CI _{Red-edge}	0.856 **/0.898 **
SAVI	0.839 **/0.900 **	MNLI	0.852 **/0.910 **	MTCI	0.652 **/0.769 **
EVI	0.842 **/0.903 **	GNDVI	0.746 **/0.745 **	WI	0.712 **/0.804 **
ARVI	0.671 **/0.730 **	DVI	0.839 **/0.905 **	NDWI	0.710 **/0.784 **
TVI	0.835 **/0.905 **	OASVI	0.815 **/0.871 **	DSWI	0.768 **/0.761 **
TGDVI	0.833 **/0.905 **	RDVI	0.842 **/0.901 **	NDII	0.710 **/0.735 **
MASVI2	0.849 **/0.901 **	MTVI2	0.849 **/0.894 **	sLAIDI*	0.830 **/0.903 **

2009-r and 2008-r: the correlation coefficients between features and LAI for the calibration data set (in 2009) and the validation data set (in 2008), respectively; ** correlation significant at the 0.01 level.

3.2. Best Features and Models

The 20 spectral waveband features, 10 spectral position features and 24 vegetation indices were then combined into a single dataset in order to assess its collective strength in predicting LAI. Every VIP value of the 54 spectral features was calculated and ranked in descending order (*i.e.*, from highest to lowest; Figure 3), ranging from 1.144 to 0.746. A high VIP value indicated that the spectral feature was of major importance in estimating LAI and also had the greater coefficient of determination in two years' data. The best spectral feature was found to be FD3 (the first derivative at a wavelength of 750 nm). Most intermediate variables have similar R^2 to FD3, but their VIP values were smaller. Waveband features in the visible and SWIR regions had the lowest VIP values.

In order to identify the predictive accuracies when different numbers of features were analyzed via PLSR, we tested a forward variable selection procedure on the VIP-selected features. This process involved ranking the VIP features based on their VIP scores and then iteratively adding the best ranked indices in a new PLS model. Firstly, the three best VIP spectral features were employed to predict LAI via PLSR. Then, the four best VIP spectral features were analyzed, and so on, until all of the features had

been used to predict LAI. For each stage of the forward selection procedure, the coefficients of determination of the first ten PLS factors and the optimal components were recorded (Figure 4).

Figure 3. Variable importance in projection (VIP) and R^2 for each spectral feature.

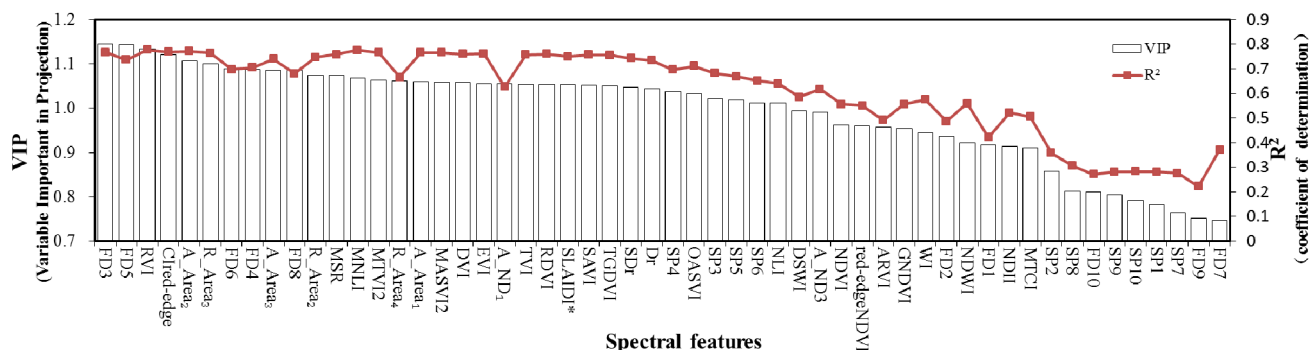
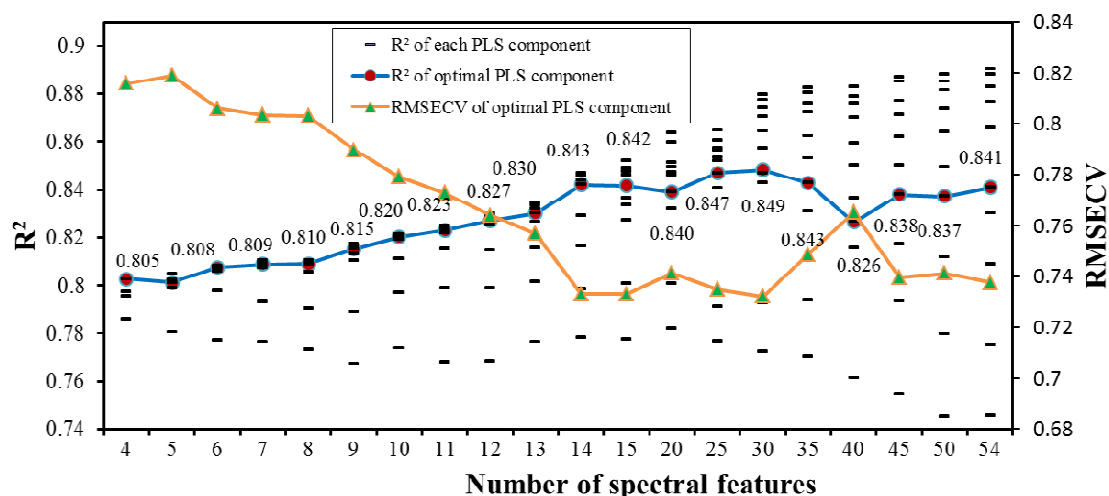


Figure 4. Variation in the coefficient of determination for different numbers of spectral features analyzed via partial least squares regression (PLSR). Note: The black horizontal dashes represent the coefficients of determination of the first ten PLS factors for different numbers of spectral features. Each optimal PLS factor is marked as a red circle, and the corresponding RMSECV is marked as green triangles.



PLSR searched the sensitive information from a different number of spectral features based on VIP values in descending order. The addition of more spectral features resulted in a decrease in the coefficient of determination of the first PLS factor and an increase in that of the tenth PLS factor. The optimum number of factors for use in the PLSR models was estimated via cross-validation, with the optimum number of PLS factors for the different datasets being four or five. The final R^2 values ranged from 0.805 to 0.849, with the top 30 features exhibiting the highest R^2 values ($R^2 = 0.849$). However, the increase in R^2 after the addition of more features was unremarkable; only a minor decrease in the number of VIP features was required in order to improve LAI estimation and to influence the final results. As analysis of Figure 4 reveals, the dataset comprising 14 spectral features was sufficient to reach an R^2 value ($R^2 = 0.842$) in the plateau region of the graph, beyond which only very small increases in R^2 could be observed. The RMSECV varied between 0.731 and 0.812, with the smallest value in the top 14 features. Therefore, these 14 spectral variables were finally used to construct a PLSR model that had both fewer spectral features and more accurate estimates.

3.3. Calibration and Validation

PLSR analyses were conducted using the optimal waveband feature dataset, the optimal position feature dataset, the vegetation index dataset, the all-feature dataset and the 14 best features dataset. Table 7 shows the results produced by these models for both calibration and validation samples. Among the three different optimal spectral feature datasets, the PLSR model based on the 24-variable vegetation index dataset produced the highest estimation accuracy. The PLSR model based on the 10-variable optimal position feature dataset exhibited only intermediate accuracy, with that based on the 20-variable optimal waveband feature dataset being associated with the lowest level of accuracy. However, these three models with different features were able to extract spectral information and produce similar LAI estimation results. The PLSR model based on the combination of all features produced estimation accuracies for both the calibration sample ($R^2 = 0.845$, RMSE = 0.733) and the validation sample ($R^2 = 0.878$, RMSE = 0.940). The best feature dataset constructed via VIP forward selection, which reduced the number of spectral features from 54 to 14, produced a similar estimation accuracy with that of the all-feature dataset in the calibration dataset ($R^2 = 0.842$, RMSE = 0.731). The 14 best feature dataset using the developed regression equation from the calibration data set yielded higher R^2 to the all-feature dataset in the independent data set ($R^2 = 0.881$, RMSE = 0.937). The prediction results of estimation models tended to be underestimated (NMBs less 0). The NMBs of the best feature model tended to be lower than that of the all dataset model for both calibration and validation analyses.

Table 7. Performance of the PLSR models in predicting LAI based on the use of different datasets. NMB, normalized mean bias.

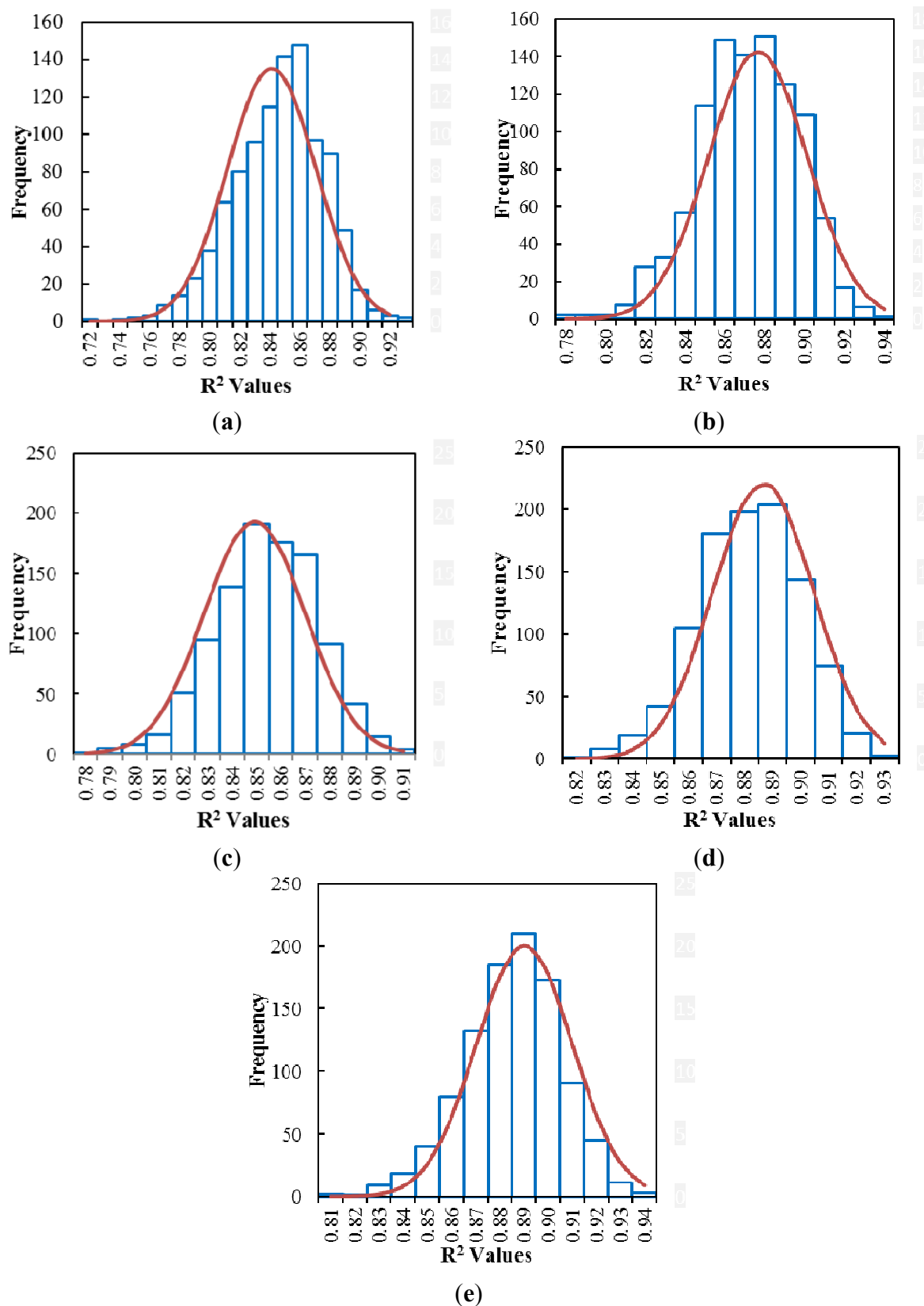
PLSR Models	No. of Variables	No. of Factors	Calibration ($n = 76$)			Validation ($n = 71$)		
			R^2	RMSE	NMB	R^2	RMSE	NMB
Optimal waveband feature model	20	3	0.803	0.843	−6.59%	0.844	1.209	−13.48%
Optimal position feature model	10	4	0.804	0.826	−3.34%	0.861	1.161	−12.36%
Vegetation index model	21	3	0.813	0.794	−3.15%	0.847	0.972	−10.1%
All-feature model	54	5	0.841	0.735	−2.77%	0.878	0.940	−7.52%
Best feature model	14	5	0.842	0.731	−1.20%	0.881	0.937	−5.23%

To assess the capability of the results, the bootstrap process was adopted. Figure 5 shows the normal distribution of the R^2 values calculated from the predicted and measured LAI in the independent data set by the bootstrapping methodology. Table 8 details the mean bootstrapped regression results in the form of the mean R^2 values, as well as the mean RMSE using PLS models. The best feature model performed the best with a mean R^2 of 0.880 and a mean RMSE of 0.943.

Table 8. Bootstrap methods for PLS models in the independent data set.

PLS Models	Mean R^2	Mean RMSE	Standard Error	95% Confidence Limit
Optimal waveband feature model	0.844	1.206	0.029	0.002
Optimal position feature model	0.866	1.155	0.025	0.002
Vegetation index model	0.849	0.974	0.021	0.001
All-feature model	0.875	0.965	0.018	0.001
Best feature model	0.880	0.943	0.020	0.001

Figure 5. Histograms showing the frequency of R^2 values between the measured and predicted LAI for the independent dataset. (a) Optimal waveband feature model, (b) Optimal position feature model, (c) Vegetation Index model, (d) All feature model, (e) Best feature model.



4. Discussion

To obtain adequate information from hyperspectral data, many features were identified based on spectral wavebands, spectral positions and vegetation indices. Additionally, the correlations between these features and wheat LAI values were studied. The analysis of the spectral waveband features revealed important features correlating with LAI across a broad range. However, compared to the performance of data produced via first derivative analysis, as well as absorption and reflectance position features and vegetation indices, the spectral features exhibited lower correlation coefficients, due to the influence of external factors, such as underlying soil brightness, leaf angle distribution and leaf optical properties [15,24].

The red-edge region is characterized by a sharp rise in the reflectance of green vegetation between the local minimum reflectance band in the red spectral region and the maximum reflectance band in the NIR spectral region. This region is considered to contain more information regarding biomass quantity and LAI than other parts of the electromagnetic spectrum [64,65]. In this study, many spectral features identified in the red-edge region had high precision and were more accurate (*i.e.*, FD3, CI_{Red-edge} and A_Area₁). FD3 was the first derivative at a wavelength of 750 nm and had both the strongest correlation with LAI ($R^2 = 0.800$) and the highest VIP value (VIP = 1.144) of all features. This confirmed previous findings by Wang *et al.* [12] with 723 nm and Thenkabail *et al.* [10] with 735 nm. CI_{Red-edge} led to an R^2 of 0.766 and a VIP of 1.121, which was more sensitive to LAI variability than the NDVI. Viña [21] also showed that the CI_{Red-edge} exhibited low sensitivity to soil background effects, and it constitutes a simple, yet robust tool for the remote and synoptic estimation of green LAI.

In the NIR region (800–1300 nm), reflectance spectra and first derivative spectra (SP4, SP5, FD4 and FD5), absorption and reflectance features (R_Area₃, A_Area₃, A_Area₂, R_Area₂ and R_Area₄) and most vegetation indices (RVI, EVI, DVI and MTVI2) performed more effectively. The reflectance in this spectral region is mainly influenced by the arrangement of cells within the mesophyll layer of leaves, as well as by canopy structure, especially the number of vertical leaf layers. NIR water absorption regions are also sensitive to leaf moisture content [12].

The advantage of vegetation indices is that they can be used to obtain relevant information rapidly and easily, and the underlying mechanisms are well-understood [66]. For the 24 VIs examined in this study, NDVI had a lower accuracy than most of the studied vegetation indices. The obtained results are in agreement with those found elsewhere in the literature [5,16,42]. Most modified vegetation indices were better than their respective originals, including: MNLI *vs.* NLI, MSAVI2 *vs.* SAVI and RDVI *vs.* DVI, which is consistent with previous studies [18,19,42]. Some vegetation indices based on three discrete bands also produced strong correlation, including the following: EVI, TVI, TGDVI, MTVI2 and sLAIDI*, which take advantage of sensitive spectral regions to reduce external factor and are highly sensitive to LAI [16,56].

The PLSR approach is considered to be the most useful explorative tool with which to unravel the relationship between canopy spectral reflectance and grass characteristics at the canopy scale. It is able to effectively address strong collinearity and noise in dependent variables [24]. Although the parameter number of the optimal spectral waveband dataset was double that of the optimal spectral position dataset, the prediction accuracy of the 10-variable optimal position feature model performed better than that of the 20-variable optimal waveband feature model. Indeed, the all-feature (54 variables) dataset

yielded a lower level of accuracy than the top 30 variables dataset. The results therefore indicate that the continuous addition of variables may not always improve LAI estimation accuracy. Indeed, the inclusion of an increasing number of less important spectral features in PLSR models can negatively influence prediction accuracy [22,26].

The calculated VIP scores provided an insight into the usefulness of each variable in the PLS model. The spectral dataset containing the top 14 variables was able to achieve a high level of estimation accuracy with the use of fewer spectral features; the subsequent inclusion of additional features resulted in only a minor improvement in model accuracy. These 14 features included the red-edge region (FD3), the NIR region (FD5, FD6, FD4, FD8, A_Area₂, A_Area₃, R_Area₂ and R_Area₃) and the best vegetation indices (RVI, CI_{Red-edge}, MSR, MNLI and MTVI2); these were also the best features in the three datasets employed for LAI estimation discussed above. The presented results demonstrate the potential of PLSR and VIP techniques in identifying important variables for the estimation of LAI. It is important to select appropriate features and to determine the optimal variable number(s). Selecting only the very best features selected by VIP values may therefore be sufficient in terms of exploring the rich information available for LAI estimation, with the use of whole feature and/or full spectrum datasets being unnecessary. This finding is in agreement with that of [25], who identified the most significant indices (chosen via VIP) producing the best PLS model prediction of *T. peregrinus* damage.

5. Conclusions

In order to select suitable spectral features for LAI estimation, different features based on spectral wavebands, spectral positions and vegetation indices were evaluated, with all exhibiting the same changing tendency in two years of hyperspectral data. The best features in three different spectral feature groups exhibited a similar correlation with LAI. Derivative analysis, a combination of vegetation index, as well as absorption and reflectance position features generally proved to be better predictors of LAI variability. Spectral features in the red-edge and NIR regions were the most sensitive for predicting LAI. The first derivative at a wavelength of 750 nm exhibited the highest correlation with LAI for all features.

PLSR and VIP analyses were conducted on spectral feature data to estimate LAI and to identify the subset of features with the best predictive accuracy. Our findings suggest that LAI estimation accuracy could be improved by employing the most sensitive spectral features in conjunction with PLSR models. The application of these methods made it possible to extract sufficient signals covering the full spectral range of information, reducing the dimensionality of the hyperspectral data and improving the steady estimation accuracy of winter wheat LAI. The 14 features with the highest VIP values provided a higher level of accuracy in predicting LAI than the entire 54-feature dataset. The validation of the new model indicated that the best feature model performed the best with the mean R^2 of 0.880 and the mean RMSE of 0.943.

Compared to other multivariate statistical models, such as principal component regression (PCR) and stepwise multiple linear regression (SMLR), PLSR outperformed other techniques in estimating canopy chlorophyll content, vegetation water content, nitrogen content, LAI, and so on [24,67,68]. However, some methods, such as support vector machines (SVM) and artificial neural networks (ANNs), are also useful for nonlinear models and vegetation canopy property estimations. To evaluate applications of these features and models proposed in this study, other vegetation types and the radiative transfer model approach will be conducted.

Acknowledgments

This work was supported in part by the National Basic Research Program of China (Grant No. 2010CB951103) and the Scientific Research Innovation Project for Graduate Students of Jiangsu province, China (Grant No. KYLX_0494). We acknowledge the support given by the National Engineering Research Center for Information Technology in Agriculture, Beijing. We are very grateful to Weiguo Li and Hong Chang for data collection. Finally, we acknowledge the anonymous reviewers who provided useful comments regarding this manuscript.

Authors Contributions

Xinchuan Li analyzed data and wrote the manuscript; Youjing Zhang, Yansong Bao, Juhua Luo and Xiuliang Jin gave comments on the manuscript and checked the writing; Xingang Xu, Xiaoyu Song and Guijun Yang provided data and data acquisition capacity.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Chen, J.M.; Black, T.A. Defining leaf area index for non-flat leaves. *Plant Cell Environ.* **1992**, *15*, 421–429.
2. Casa, R.; Varella, H.; Buis, S.; Gu E Rif, M.; de Solan, B.; Baret, F. Forcing a wheat crop model with LAI data to access agronomic variables: Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *Eur. J. Agron.* **2012**, *37*, 1–10.
3. Chen, J.M.; Cihlar, J. Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sens. Environ.* **1996**, *55*, 153–162.
4. Pu, R. Comparing canonical correlation analysis with partial least squares regression in estimating forest leaf area index with multitemporal Landsat TM imagery. *GISci. Remote Sens.* **2012**, *49*, 92–116.
5. Walthall, C.; Dulaney, W.; Anderson, M.; Norman, J.; Fang, H.; Liang, S. A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. *Remote Sens. Environ.* **2004**, *92*, 465–474.
6. Eklundh, L.; Harrie, L.; Kuusk, A. Investigating relationships between Landsat ETM+ sensor data and leaf area index in a boreal conifer forest. *Remote Sens. Environ.* **2001**, *78*, 239–251.
7. Bréda, N.J. Ground-based measurements of leaf area index: A review of methods, instruments and current controversies. *J. Exp. Bot.* **2003**, *54*, 2403–2417.
8. Hansen, P.M.; Schjoerring, J.K. 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*, 542–553.
9. Lee, K.; Cohen, W.B.; Kennedy, R.E.; Maier-sperger, T.K.; Gower, S.T. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sens. Environ.* **2004**, *91*, 508–520.

10. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; van der Meer, B. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.* **2004**, *91*, 354–376.
11. Becker, B.L.; Lusch, D.P.; Qi, J. Identifying optimal spectral bands from *in situ* measurements of Great Lakes coastal wetlands using second-derivative analysis. *Remote Sens. Environ.* **2005**, *97*, 238–248.
12. Wang, F.; Huang, J.; Zhou, Q.; Wang, X. Optimal waveband identification for estimation of leaf area index of paddy rice. *J. Zhejiang Univ. Sci. B* **2008**, *9*, 953–963.
13. Strachan, I.B.; Pattey, E.; Boisvert, J.B. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sens. Environ.* **2002**, *80*, 213–224.
14. Pu, R.; Gong, P.; Biging, G.S.; Larrieu, M.R. Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 916–921.
15. Darvishzadeh, R.; Atzberger, C.; Skidmore, A.K.; Abkar, A.A. Leaf Area Index derivation from hyperspectral vegetation indices and the red edge position. *Int. J. Remote Sens.* **2009**, *30*, 6199–6218.
16. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. 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*, 337–352.
17. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309.
18. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126.
19. Roujean, J.; Breon, F. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384.
20. Broge, N.H.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* **2001**, *76*, 156–172.
21. Viña, A.; Gitelson, A.A.; Nguy-Robertson, A.L.; Peng, Y. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sens. Environ.* **2011**, *115*, 3468–3478.
22. Clevers, J.; de Jong, S.M.; Epema, G.F.; van der Meer, F.D.; Bakker, W.H.; Skidmore, A.K.; Scholte, K.H. Derivation of the red edge index using the MERIS standard band setting. *Int. J. Remote Sens.* **2002**, *23*, 3169–3184.
23. Ryu, C.; Suguri, M.; Umeda, M. Multivariate analysis of nitrogen content for rice at the heading stage using reflectance of airborne hyperspectral remote sensing. *Field Crop. Res.* **2011**, *122*, 214–224.
24. 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. Remote Sens.* **2008**, *63*, 409–426.
25. Oumar, Z.; Mutanga, O.; Ismail, R. Predicting *Thaumastocoris peregrinus* damage using narrow band normalized indices and hyperspectral indices using field spectra resampled to the Hyperion sensor. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 113–121.

26. Wold, H. Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis*; Krishnaiah, P.R., Ed.; Academic Press: Waltham, MA, USA, 1966; pp. 391–420.
27. Chin, W.W. The partial least squares approach to structural equation modeling. *Modern Methods Bus. Res.* **1998**, *295*, 295–336.
28. Haenlein, M.; Kaplan, A.M. A beginner's guide to partial least squares analysis. *Underst. Stat.* **2004**, *3*, 283–297.
29. Huang, Z.; Turner, B.J.; Dury, S.J.; Wallis, I.R.; Foley, W.J. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens. Environ.* **2004**, *93*, 18–29.
30. Nguyen, H.T.; Lee, B. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. *Eur. J. Agron.* **2006**, *24*, 349–356.
31. Vyas, D.; Christian, B.; Krishnayya, N.S.R. Canopy level estimations of chlorophyll and LAI for two tropical species (teak and bamboo) from Hyperion (EO1) data. *Int. J. Remote Sens.* **2012**, *34*, 1676–1690.
32. Li, F.; Mistele, B.; Hu, Y.; Chen, X.; Schmidhalter, U. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. *Eur. J. Agron.* **2014**, *52*, 198–209.
33. Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
34. Clark, R.N.; Roush, T.L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *J. Geophys. Res.: Sol. Earth* **1984**, *89*, 6329–6340.
35. Mutanga, O.; Skidmore, A.K. Hyperspectral band depth analysis for a better estimation of grass biomass (*Cenchrus ciliaris*) measured under controlled laboratory conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 87–96.
36. Kokaly, R.F.; Clark, R.N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* **1999**, *67*, 267–287.
37. Gong, P.; Pu, R.; Heald, R.C. Analysis of *in situ* hyperspectral data for nutrient estimation of giant sequoia. *Int. J. Remote Sens.* **2002**, *23*, 1827–1850.
38. Kaufman, Y.J.; Tanre, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 261–270.
39. Richardson, A.J.; Weigand, C.L. Distinguishing vegetation from soil background information. *Photogramm. Eng. Remote Sens.* **1977**, *43*, 1541–1552.
40. Hui, Q.L.; Huete, A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465.
41. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298.
42. Gong, P.; Pu, R.; Biging, G.S.; Larrieu, M.R. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1355–1362.

43. Chen, J.M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242.
44. Rouse, J.W.; Hass, R.H.; Shell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of Third ERTS Symposium, Washington, DC, USA, 10–14 December 1973.
45. Goel, N.S.; Qin, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and Fpar: A computer simulation. *Remote Sens. Rev.* **1994**, *10*, 309–347.
46. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107.
47. Jordan, C.F. Derivation of leaf area index from quality of light on the forest floor. *Ecology* **1969**, *50*, 663–666.
48. Tang, S.; Zhu, Q.; Wang, J.; Zhou, Y.; Zhao, F. Theoretical bases and application of three gradient difference vegetation index (In Chinese). *Sci. China Ser. D Earth Sci.* **2003**, *33*, 1094–1102.
49. Gitelson, A.; Merzlyak, M.N. Spectral reflectance changes associated with autumn senescence of aesculus Hippocastanum L. and acer Platanoides L. leaves. Spectral features and relation to chlorophyll estimation. *J. Plant Physiol.* **1994**, *143*, 286–292.
50. Gitelson, A.A. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* **2003**, *30*, doi:10.1029/2002GL016450.
51. Dash, J.; Curran, P.J. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413.
52. Penuelas, J.; Filella, I.; Serrano L.; Savé, R. Cell wall elasticity and Water Index (R970 nm/R900 nm) in wheat under different nitrogen availabilities. *Int. J. Remote Sens.* **1996**, *17*, 373–382.
53. Gao, B.; Goetz, A.F.H. Retrieval of equivalent water thickness and information related to biochemical components of vegetation canopies from AVIRIS data. *Remote Sens. Environ.* **1995**, *52*, 155–162.
54. Hardisky, M.A.; Klemas, V.; Smart, R.M. The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of spartina alterniflora canopies. *Photogramm. Eng. Remote. Sens.* **1983**, *49*, 77–83.
55. Apan, A.; Held, A.; Phinn, S.; Markley, J. Detecting sugarcane “range rust” disease using EO-1 Hyperion hyperspectral imagery. *Int. J. Remote Sens.* **2004**, *25*, 489–498.
56. Delalieux, S.; Somers, B.; Hereijgers, S.; Verstraeten, W.W.; Keulemans, W.; Coppin, P. A near-infrared narrow-waveband ratio to determine Leaf Area Index in orchards. *Remote Sens. Environ.* **2008**, *112*, 3762–3772.
57. Abdi, H. Partial Least Square Regression (PLS Regression). In *Encyclopedia of Social Science Research Methods*; SAGE: Thousand Oaks, CA, USA, 2003; pp. 792–795.
58. Helland, I.S. Partial least squares regression and statistical models. *Scand. J. Stat.* **1990**, *17*, 97–114.
59. Sheng, C.; Xia, H.; Harris, C.J.; Sharkey, P.M. Sparse modeling using orthogonal forward regression with PRESS statistic and regularization. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2004**, *34*, 898–911.

60. Chong, I.; Jun, C. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab.* **2005**, *78*, 103–112.
61. Lazraq, A.; Cléroux, R.; Gauchi, J. Selecting both latent and explanatory variables in the PLS1 regression model. *Chemom. Intell. Lab.* **2003**, *66*, 117–126.
62. Mutanga, O.; Skidmore, A.K.; Prins, H.H.T. Predicting *in situ* pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sens. Environ.* **2004**, *89*, 393–408.
63. Oumar, Z.; Mutanga, O. Predicting plant water content in Eucalyptus grandis forest stands in KwaZulu-Natal, South Africa using field spectra resampled to the Sumbandila Satellite Sensor. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 158–164.
64. Mutanga, O.; Skidmore, A.K. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* **2004**, *25*, 3999–4014.
65. Herrmann, I.; Pimstein, A.; Karnieli, A.; Cohen, Y.; Alchanatis, V.; Bonfil, D.J. LAI assessment of wheat and potato crops by VENUS and Sentinel-2 bands. *Remote Sens. Environ.* **2011**, *115*, 2141–2151.
66. Delegido, J.; Verrelst, J.; Meza, C.M.; Rivera, J.P.; Alonso, L.; Moreno, J. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *Eur. J. Agron.* **2013**, *46*, 42–52.
67. 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*, 165–173.
68. Mirzaie, M.; Darvishzadeh, R.; Shakiba, A.; Matkan, A.A.; 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. Geoinf.* **2014**, *26*, 1–11.