

Article

A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy

Jibo Yue^{1,2,3,†} , Haikuan Feng^{1,4,*,†}, Guijun Yang^{1,3,4,*}  and Zhenhai Li^{1,3,4}

¹ Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture China, Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China; yuejb.ha@gmail.com (J.Y.); lizh@nercita.org.cn (Z.L.)

² International Institute for Earth System Science, Nanjing University, Nanjing 210023, China

³ National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

⁴ Beijing Engineering Research Center for Agriculture Internet of Things, Beijing 100097, China

* Correspondence: fenghaikuan123@163.com (H.F.); guijun.yang@163.com (G.Y.); Tel.: +86-10-5150-3215 (H.F.); +86-10-5150-3647 (G.Y.); Fax: +86-10-5150-3750 (H.F.)

† Both authors contributed equally to this work and should be considered co-first authors.

Received: 16 November 2017; Accepted: 2 January 2018; Published: 5 January 2018

Abstract: Above-ground biomass (AGB) provides a vital link between solar energy consumption and yield, so its correct estimation is crucial to accurately monitor crop growth and predict yield. In this work, we estimate AGB by using 54 vegetation indexes (e.g., Normalized Difference Vegetation Index, Soil-Adjusted Vegetation Index) and eight statistical regression techniques: artificial neural network (ANN), multivariable linear regression (MLR), decision-tree regression (DT), boosted binary regression tree (BBRT), partial least squares regression (PLSR), random forest regression (RF), support vector machine regression (SVM), and principal component regression (PCR), which are used to analyze hyperspectral data acquired by using a field spectrophotometer. The vegetation indexes (VIs) determined from the spectra were first used to train regression techniques for modeling and validation to select the best VI input, and then summed with white Gaussian noise to study how remote sensing errors affect the regression techniques. Next, the VIs were divided into groups of different sizes by using various sampling methods for modeling and validation to test the stability of the techniques. Finally, the AGB was estimated by using a leave-one-out cross validation with these powerful techniques. The results of the study demonstrate that, of the eight techniques investigated, PLSR and MLR perform best in terms of stability and are most suitable when high-accuracy and stable estimates are required from relatively few samples. In addition, RF is extremely robust against noise and is best suited to deal with repeated observations involving remote-sensing data (i.e., data affected by atmosphere, clouds, observation times, and/or sensor noise). Finally, the leave-one-out cross-validation method indicates that PLSR provides the highest accuracy ($R^2 = 0.89$, RMSE = 1.20 t/ha, MAE = 0.90 t/ha, NRMSE = 0.07, CV (RMSE) = 0.18); thus, PLSR is best suited for works requiring high-accuracy estimation models. The results indicate that all these techniques provide impressive accuracy. The comparison and analysis provided herein thus reveals the advantages and disadvantages of the ANN, MLR, DT, BBRT, PLSR, RF, SVM, and PCR techniques and can help researchers to build efficient AGB-estimation models.

Keywords: regression techniques; biomass; vegetation indexes; sampling methods; noise immunity; biomass estimation model; hyperspectral; multi-collinearity

1. Introduction

Accurate estimates of crop biophysical variables are crucial for monitoring vegetation growth and for analyzing important physiological parameters during the crop growth cycle [1,2]. One such variable, above-ground biomass (AGB), plays an important role in plant functioning because it reflects the status of crop growth and is related to solar-energy consumption, yield, and grain quality [3,4]. Therefore, AGB is considered as one of the most important crop biophysical parameters, and its accurate estimation can help improve crop monitoring and yield prediction [5]. Traditional AGB estimates are based on destructive measurements, which are not only time and labor consuming, but more importantly, are difficult to apply over large areas [6]. In recent years, Hyperspectral remote-sensing data acquired from the ground [7,8], unmanned aerial vehicles [9], airborne platforms [10–12], and satellite platforms [13] have been able to capture crop canopy spectra in narrow bands and thereby provide information on the biochemical composition of the canopy. Crop physiology research shows that spectral absorption by plant leaves is mainly due to the leaf pigments, especially chlorophyll content (Chl) [14]. The reflectance is low in both the blue and red regions of the spectrum, due to absorption by chlorophyll for photosynthesis; it has a peak at the green region which gives rise to the green color of vegetation [15]. In the near-infrared region, the reflectance is much higher than that in the visible band due to the cellular structure in the leaves [16]. Previous studies have shown that near-infrared- and red-band vegetation indexes (VIs) are effective for estimating AGB [8,9,11]. However, during the reproductive growth of crops, with the senescence of leaves, the effectiveness of photosynthesis is reduced [14,17]. With clear decreases in both photosynthesis and the near-infrared reflectance, the correlation between AGB and the red- and near-infrared-based VIs reduced. Therefore, hyperspectral remote sensing of AGB has received increasing attention as an efficient and precise method for nondestructive monitoring in agricultural research [18].

Physically based models and empirical regression techniques are two essential approaches for estimating vegetation characteristics from hyperspectral measurements [19]. Physically based models were founded on physical principles. The two main examples of this approach are radiative transfer (RT) models and geometric optical models [19]. Because vegetation canopy reflectance depends on a number of factors [20] (e.g., leaf-area index, Chl, water content, matter content, soil reflectance, and bidirectional reflectance distribution function), physically based models require canopy biophysical parameters, soil parameters, and some external parameters to simulate canopy reflectance, and these are often not readily available. In contrast, empirical regression techniques require a large number of ground measurements, and offer a direct relationship between spectral features and vegetation parameters. Previous research has used many powerful empirical regression techniques that make full use of the narrow hyperspectral bands, VIs, and even different types of sensor data [21]. These techniques essentially fall into two categories: (i) machine-learning techniques such as artificial neural network (ANN) [22], decision tree regression (DT) [23], boosted binary regression tree (BBRT) [24], random forest regression (RF) [25], support vector machine regression (SVM) [26], and (ii) conventional regression techniques such as multivariable linear regression (MLR) [26,27], partial least squares regression (PLSR) [7,8,22], and principal component regression (PCR) [7]. Many studies have obtained promising results by using these techniques [8–11,26]. However, hyperspectral data redundancy is a big problem because of the high spectral dimensions and large number of bands [28]. In addition, the correlation between the spectral and AGB vary with the crop growth period, which is related to the physiological state of the crop [17]. To address this problem, many researchers have tried to extract features from narrow hyperspectral bands first, and many methods to do this have been proposed; for example, correlation analysis, continuum removal [29], red-edge position [30], gray relational analysis [31], and out-of-bag analysis [21]. Spectral vegetation indexes (VIs) have been widely used for decades, and more than 60 VIs [32] have been proposed for estimating biophysical variables [33].

Conventional regression techniques are more suitable for data that have a clear linear or exponential relationship with a distinct estimation equation, whereas machine-learning techniques are

typically better able to cope with the strong nonlinearity between the biophysical and biochemical parameters and the reflection spectra [34]. However, many studies indicate that empirical regression techniques are rarely transferable to other sites with different vegetation, or to data acquired from other types of sensors or under different acquisition conditions. Despite this, empirical regression techniques still have some advantages, such as fewer input variables, less computation, and ease of application, which have resulted in their widespread use under many conditions.

Numerous studies have used hyperspectral remote-sensing data and empirical regression techniques to estimate AGB [26], and some analyses of the performance of these techniques have also been carried out, although they focus mostly on comparing the estimation accuracy. No comprehensive study is available as yet that evaluates these regression techniques for estimating AGB, and no studies have evaluated the different statistical techniques to better understand their respective advantages and disadvantages.

The main objective of the present study is to evaluate the performance (in particular, data selection, sampling methods, noise immunity) of eight regression techniques for estimating AGB. The following four tests were applied:

- (1) VIs were used to train regression models, which were validated to select best VI input (Section 4.1).
- (2) The noise immunities of these techniques were compared by simulating remote-sensing errors by adding white Gaussian noise (Section 4.2).
- (3) The stability of these techniques was examined by using samples of varying sizes and different sampling methods for modeling and validation (Section 4.3).
- (4) Leave-one-out cross validation was used to evaluate the accuracy of the AGB estimation of these techniques (Section 4.4).

We discuss the performance of eight AGB estimation techniques and the advantages and disadvantages of each technique (Section 5), then summarize the optimal conditions for using these techniques.

2. Materials

2.1. Study Area

The study area was situated in Changping District, which is located in the northwest part of Beijing City, China (see Figure 1). Experiments were conducted at the National Precision Agriculture Research Center of China (116°26'36"E, 40°10'44"N). Changping District has an average altitude of 36 m, its total area is about 1352 km², and it has a warm temperate semi-humid continental monsoon climate, with an average rainfall of 450 mm, an average low temperature of −10 to 7.5 °C and an average high temperature of 35 to 40 °C.

The aim of the agronomy experiment was to increase the difference in AGB by using two crop varieties, three water treatments, and four nitrogen treatments. The AGB was measured by using ground-based techniques. The experiments involved two winter wheat cultivars, J9843 and ZM175, which are the main winter wheat varieties grown in northern China. The irrigation treatment included rainfall only (W0, see Figure 1), rainfall plus normal irrigation (W1, 100 mm), and rainfall plus double the normal irrigation (W2, 200 mm). The nitrogen fertilizer treatment included no fertilizer (N0), one-half the normal fertilization (N1, 195 kg/ha), normal fertilization (N2, 390 kg/ha), and twice the normal fertilization (N3, 780 kg/ha).

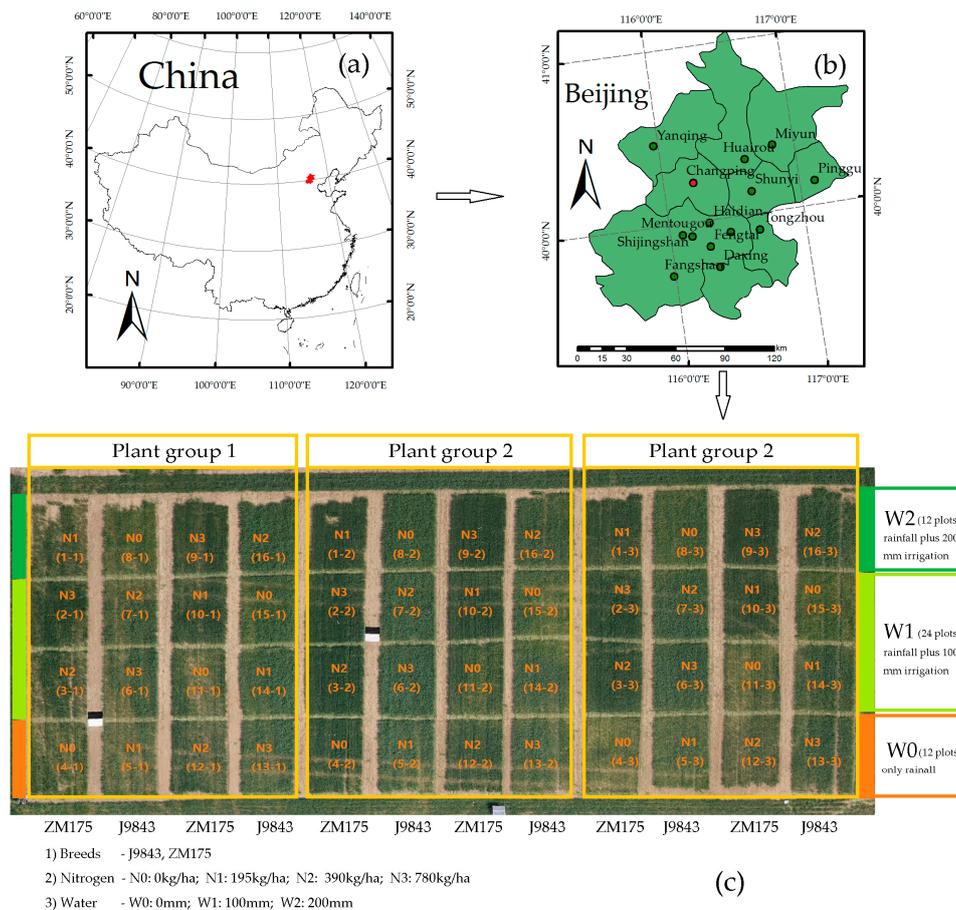


Figure 1. (a) Location of study area shown in red. (b) Map showing Changping District in Beijing City. (c) Design of treatments and an unmanned-aerial-vehicle image of the experimental field (acquired on 12 May 2015). Three plant groups are present with two winter wheat varieties (J9843 and ZM175), three water treatments (W0, W1, and W2), and four nitrogen treatments (N0, N1, N2, and N3).

2.2. Measurement of Data

A 5000 m² square area was selected as the experimental field (see Figure 1) and divided into 48 plots each of size 6 m × 8 m. In each plot, winter wheat near the center of the given plot was selected for spectral, physiological, and biochemical measurements and analyses. AGB, Chl and canopy spectral measurements were made at four growth stages: the winter wheat jointing stage (13 and 14 April 2015), the flag leaf stage (26 and 27 April 2015), the flowering period (12 to 14 May 2015), and the filling period (25 to 27 May 2015). Four ground-based measurements allowed 192 sets of Chl, winter wheat biomass, and canopy hyperspectral data to be collected.

2.2.1. Measurements of Winter Wheat Canopy Reflectance

Canopy hyperspectral reflectance was acquired by using an ASD FieldSpec 3 spectrometer (FieldSpec 3 spectrometer, Analytical Spectral Devices, Boulder, Colorado, CO, USA) from 10:00~14:00 (Beijing time, UTC/GMT+08:00) in windless and cloudless conditions. We calibrated the field spectrometer based on the reflectance from a 40 cm × 40 cm BaSO₄ white board, and the vertical height from the canopy is 1.3 m. The winter wheat canopy reflectance was measured 10 times (the scanning time was 0.2 s) at the center of each plot, and the average reflectance was recorded. To reduce the influence of sky and field conditions on the spectral measurements, each plot was measured three times, and the mean value was used as the canopy reflectance for the given experimental plot. Figure 2a shows the average hyperspectral reflectance spectrum for the four growing stages.

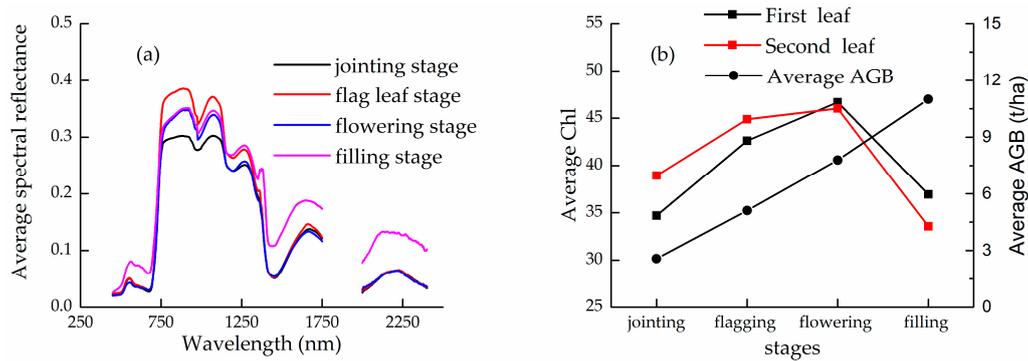


Figure 2. (a) Average hyperspectral reflectance spectrum for the four growing stages. (b) Average Chl and above-ground biomass (AGB) for the four growing stages.

2.2.2. Measurements of Winter Wheat Chlorophyll and Above-Ground Biomass

During measurements, the planting density of winter wheat (row spacing 15 cm) was investigated, and 20 stems were collected near the center of each plot. Chl was measured from the first and second uppermost leaves by using a Dualex 4 (Dualex Scientific Portable Sensor for Leaf Measurements, Force-a, Université Paris Sud, Orsay, France) and the average values were processed (see Figure 2b).

After ground measurements, the winter wheat organs were processed in the laboratory. They were first put into paper bags and dried at 80 °C to remove moisture, then, once the sample weight became constant (about 24 h), they were weighed by using a balance with an accuracy of 0.001 g. Finally, the biomass per unit area was calculated based on the measured planting density and sample dry weight. The winter wheat AGB was calculated by using

$$AGB = \frac{m \times n}{20 \times l} C \quad (1)$$

where m is the dry weight of the sample, n is the number of winter wheat ears per unit area, and l is the row spacing. The statistics of the AGB measurement for different growing periods is shown in Table 1.

Table 1. Statistics of AGB measurement in study area.

Period	Sample	Min (t/ha)	Max (t/ha)	Mean (t/ha)	Standard Deviation (t/ha)	Coefficient of Variation (%)
Jointing	48	1.201	4.526	2.569	0.685	26.664
Flag	48	2.194	8.266	5.114	1.468	28.706
Flowering	48	3.419	12.737	7.790	1.960	25.160
Grain filling	48	5.456	17.599	10.993	2.793	25.407

3. Methods

Data selection, sampling methods, noise immunity, and prediction performance were analyzed by using a series of VIs and ground-based measurements of the AGB. The flowchart in Figure 3 illustrates the process.

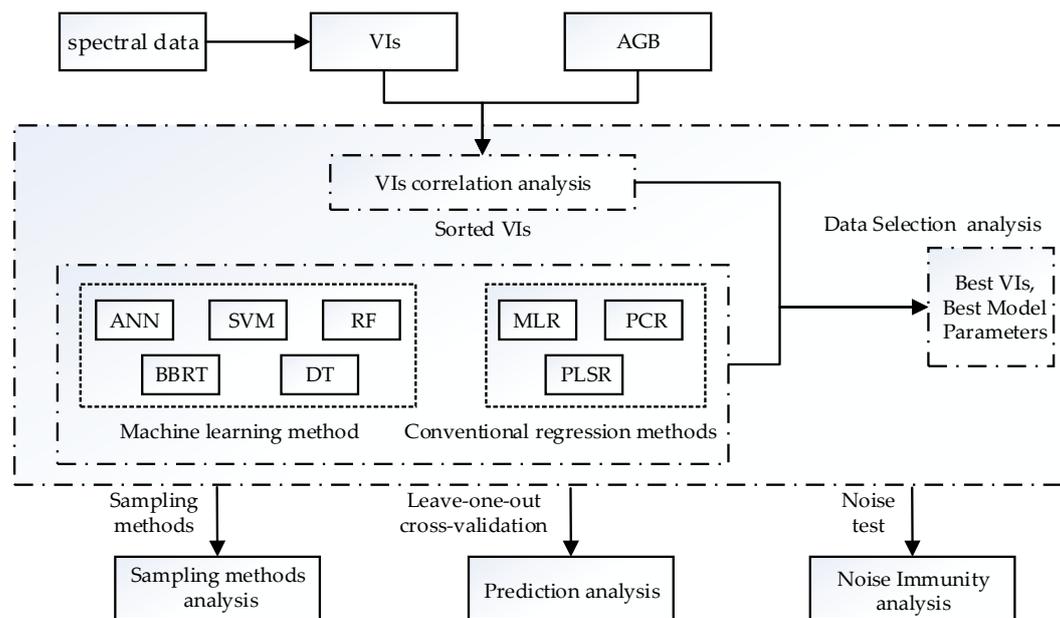


Figure 3. Flowchart showing experiment methodology. Data selection, sampling methods, noise immunity, and prediction accuracy were analyzed.

3.1. Regression Techniques

3.1.1. Machine Learning Techniques

(1) Artificial neural networks have represented a hot research topic in artificial intelligence since the 1980s; this method is very powerful in dealing with nonlinear relationships [35]. An ANN is based on a collection of connected units called artificial neurons, and each neuron can transmit a signal to other neurons. ANN is composed of a large number of neurons, with each neuron representing a particular output function. The connection between two neurons represents the weighted value of the signal through the connection. The network outputs different weighted values and the incentive function, according to the network connection mode. (2) Support vector machines were proposed by Cortes and Vapnik [36] in 1995 and offer many unique advantages for dealing with complex multidimensional data. A SVM constitutes a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis. A SVM can be used as a regression method, maintaining all the main features that characterize the algorithm (i.e., maximal margin). Support vector regression (SVR) uses the same principles as a SVM for classification, with only a few minor differences. Herein, we use LIBSVM (LIBSVM 3.1.2—A Library for Support Vector Machines, Version 3.12 [37]) for the tests. (3) A decision tree is a tree structure in which each internal node represents a test of an attribute, each branch represents a test output, and each leaf node represents a category [38]. A decision node has two or more branches, each representing values for the attribute tested. Breaking down a dataset into smaller and smaller subsets incrementally develops an associated decision tree. The final result is a tree with decision nodes and leaf nodes. A leaf node represents a decision for the numerical target. (4) A boosted binary regression tree is a powerful regression method proposed by Friedman [24] in 2001. Boosted binary regression trees combine binary regression trees by using a gradient-boosting technique [39]. (5) Random forest regression is a data analysis and statistical method that is widely used in machine-learning research. It was proposed by Breiman and Cutler [25] in 2001 and is an ensemble learning method for classification and regression that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or the mean prediction (regression) of the individual trees. RF has a higher accuracy, better tolerance to outliers and noise, and makes excellent use of the full spectral information.

3.1.2. Conventional Regression Techniques

(1) Multiple linear regression is a regression method in which two or more independent variables are used to analyze a dependent variable. The regression equation is used to calculate the parameters by using the least squares method in which the sum of the errors squared is minimized. (2) Partial least squares regression is a data analysis method proposed by Wold [40] in 1966. PLSR has also been widely used in studies of vegetation because it provides an efficient way to make full use of hyperspectral information. Previous studies [8–10] indicate that PLSR makes excellent use of the full spectral information and is a flexible method for monitoring agricultural crop parameters. (3) Principal component analysis (PCA) is a technique to simplify data sets based on a linear transformation of data into a new coordinate system. After that transformation, the largest variance in the data projection appears in the first coordinate (called the first principal component), the second largest variance appears in second coordinate (second principal component), and so on. PCA often reduces the dimensionality of data sets. This method can reduce the dimensionality of hyperspectral data, thus avoiding the problem of collinear variables that can occur in PLSR and MLR regression [7].

In the present work, we analyze the ANN, SVM, RF, BBRT, DT, MLR, PLSR, and PCR regression by using Matlab2014a (Matrix Laboratory 2014a, MathWorks, Inc., Natick, MA, USA) on a Microsoft Windows platform.

3.2. Selection of Vegetation Indexes

A VI is a combination of two or more characteristic spectra acquired by multispectral or hyperspectral remote sensing. It is a simple, effective, and empirical measure of the surface vegetation status. VIs are widely used to classify vegetation and environmental changes, determine crop and forage yield, monitor droughts, etc. After many years of research on narrow-band hyperspectral spectra, incomplete statistics show that dozens of VIs exist that can be used to estimate biophysical parameters [13,36–74].

Data redundancy and multi-collinearity can seriously affect regression performance. By using selected 54 VIs (Table 2) [32], the abilities of eight techniques to solve the multi-collinearity problem can be analyzed. Section 4.1 gives the best input VIs (we set seven levels: 5, 10, 15, 20, 30, 40 and 54) based on eight techniques by analyzing the modeling and validation as a function of the VIs used as the input.

Table 2. Summary of VIs used in this study.

VIs	Equation	Reference	VIs	Equation	Reference
ATSAVI	$\frac{a(R_{800} - aR_{670} - b)}{X(1 + a^2)}$, where $X = 0.08$, $a = 1.22$, and $b = 0.03$	[41]	MND ₆₈₀	$(R_{800} - R_{680}) / (R_{800} + R_{680} - 2R_{445})$	[42]
EVI	$2.5(R_{NIR} - R_{Red}) / (R_{NIR} + 6R_{Red} - 7.5R_{Blue} + 1)$	[43]	MND ₇₀₅	$(R_{750} - R_{705}) / (R_{750} + R_{705} - 2R_{445})$	[42]
EVI2	$2.5(R_{NIR} - R_{Red}) / (R_{NIR} + 2.4R_{Red} + 1)$	[44]	MSR ₇₀₅	$(R_{750} - R_{445}) / (R_{705} - R_{445})$	[42]
GI	R_{554} / R_{677}	[45]	NPCI	$(R_{680} - R_{430}) / (R_{680} + R_{430})$	[46]
LAIDI	R_{1250} / R_{1050}	[47]	NPQI	$(R_{415} - R_{435}) / (R_{415} + R_{435})$	[48]
MSAVI	$0.5[2R_{800} + 1 - ((2R_{800} + 1)^2 - 8(R_{800} - R_{670}))^{1/2}]$	[49]	PBI	R_{810} / R_{560}	[50]
MSR	$(R_{800} / R_{670} - 1) / (R_{800} / R_{670} + 1)^{1/2}$	[51]	PRI	$(R_{531} - R_{570}) / (R_{531} + R_{570})$	[52]
MTVI1	$1.2[1.2(R_{800} - R_{550}) - 2.5(R_{670} - R_{550})]$	[53]	PSSR	R_{800} / R_{500}	[54]
MTVI2	$\frac{\{1.5[1.2(R_{800} - R_{550}) - 2.5(R_{670} - R_{550})]\}}{\{(2R_{800} + 1)^2 - [6R_{800} - 5(R_{670})^{1/2}] - 0.5\}^{1/2}}$	[53]	RARS	R_{760} / R_{500}	[55]
NDVI	$(R_{NIR} - R_{Red}) / (R_{NIR} + R_{Red})$	[56]	RGR	R_{Red} / R_{Green}	[57]
OSAVI	$1.16(R_{800} - R_{670}) / (R_{800} + R_{670} + 0.16)$	[58]	SIPI	$(R_{800} - R_{445}) / (R_{800} - R_{680})$	[59]
PSND	$(R_{800} - R_{470}) / (R_{800} + R_{470})$	[54]	TVI	$0.5[120(R_{750} - R_{550}) - 200(R_{670} - R_{550})]$	[16]
PVI _{hyp}	$(R_{1148} - aR_{807} - b) / (1 + a^2)^{1/2}$, where $a = 1.17$ and $b = 3.37$	[12]	CAI	$0.5(R_{2020} + R_{2220}) - R_{2100}$	[60]

Table 2. Cont.

VIs	Equation	Reference	VIs	Equation	Reference
RDVI	$(R_{800} - R_{670}) / (R_{800} + R_{670})^{1/2}$	[61]	NDLI	$[\log(1/R_{1754}) - \log(1/R_{1680})] / [\log(1/R_{1754}) + \log(1/R_{1680})]$	[62]
SLAIDI	$S(R_{1050} - R_{1250}) / (R_{1050} + R_{1250})$, where $S = 5$	[47]	NDNI	$[\log(1/R_{1510}) - \log(1/R_{1680})] / [\log(1/R_{1510}) + \log(1/R_{1680})]$	[62]
SPVI	$0.4[3.7(R_{800} - R_{670}) - 1.2 R_{530} - R_{670}]$	[63]	DSWI	$(R_{802} + R_{547}) / (R_{1657} + R_{682})$	[13]
TCARI	$3[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})(R_{700}/R_{670})]$	[64]	LWVI1	$(R_{1094} - R_{893}) / (R_{1094} + R_{893})$	[13]
SR	R_{NIR} / R_{Red}	[65]	LWVI2	$(R_{1094} - R_{1205}) / (R_{1094} + R_{1205})$	[13]
VARI _{green}	$(R_{Green} - R_{Red}) / (R_{Green} + R_{Red})$	[66]	MSI	R_{1600} / R_{819}	[67]
WDRVI	$(0.1 R_{NIR} - R_{Red}) / (0.1 R_{NIR} + R_{Red})$	[68]	NDII	$(R_{819} - R_{1600}) / (R_{819} + R_{1600})$	[69]
ARI	$(R_{550})^{-1} - (R_{700})^{-1}$	[70]	NDWI	$(R_{860} - R_{1240}) / (R_{860} + R_{1240})$	[71]
BGI	R_{450} / R_{550}	[45]	RV _{I_{hyp}}	R_{1088} / R_{1148}	[12]
BRI	R_{450} / R_{690}	[45]	SIWSI	$(R_{860} - R_{1640}) / (R_{860} + R_{1640})$	[72]
LCI	$(R_{850} - R_{710}) / (R_{850} + R_{680})$	[15]	SRWI	R_{860} / R_{1240}	[73]
MCARI	$[(R_{701} - R_{671}) - 0.2(R_{701} - R_{549})] / (R_{701} / R_{671})$	[74]	WI	R_{900} / R_{970}	[75]
MCARI1	$1.2[2.5(R_{800} - R_{670}) - 1.3(R_{800} - R_{550})]$	[53]	PSRI	$(R_{680} - R_{500}) / R_{750}$	[76]
MCARI2	$[1.5[2.5(R_{800} - R_{670}) - 1.3(R_{800} - R_{550})]] / [1(2R_{800} + 1)^2 - [6R_{800} - 5(R_{670})^{1/2}] - 0.5]^{1/2}$	[53]	RVSI	$[(R_{712} + R_{752}) / 2] - R_{732}$	[77]

Note: R_{Green} , R_{Red} , and R_{NIR} represent bands at 470, 670, and 800 nm of hyperspectral reflectance, respectively. R_{470} and R_{800} represent bands at 470 and 800 nm of hyperspectral reflectance, etc.

3.3. Noise Simulation

Many error sources exist in remote-sensing imaging and sensor systems (see Figure 4), including radiation errors caused by the atmosphere, topography, or other geometric errors and systematic errors related to the charge-coupled device (CCD) sensor [78,79]. Although radiometric calibration and radiometric correction are applied to correct for sensor degradation and atmospheric effects, the noise cannot be completely removed. Noise such as shot noise due to the quantum properties of light and readout noise generated by the output amplifier remain, and both follow a Poisson distribution [78]. In addition, the dark-current noise and thermal noise are present and are proportional to the CCD temperature; these follow a Gaussian distribution [78,79].

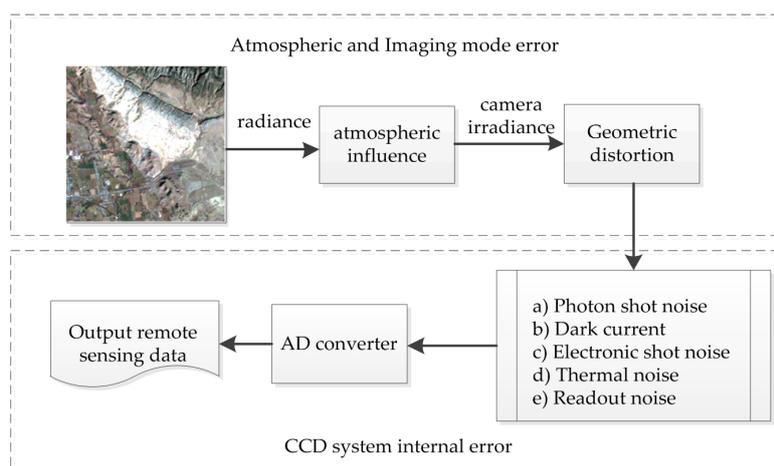


Figure 4. Remote-sensing imaging process and noise sources in a charge-coupled device (CCD) system.

To evaluate how sensor noise and other uncertainty sources affect these data-analysis techniques, we simulate the internal noise (dark current, random noise) of a CCD used for remote sensing. We add random noise to the validation VIs to analyze how remote-sensing noise affects the stability of these techniques with the help of white Gaussian noise [7]. These models to estimate the AGB use the original VIs and are validated by using VIs with noise added. In these tests, the signal-to-noise ratio

(SNR) is set to 5, 10, 15, 20, 30, and 50 (the noise increases with decreasing SNR). We also compare the results to those obtained with noiseless VIs.

3.4. Modeling Parameters and Sampling Methods

Before the test, the ability to deal with multiple VIs (best input VIs) and to obtain the modeling parameters must be analyzed. In this test, the optimal modeling parameters are determined according to the accuracy of validation by using an exhaustive method within limits (Table 3). In the next modeling test, these optimal parameters were selected for testing. Through thousands of rough modeling and verification processes (Note in Table 3), we obtained the best optimal model parameters under the optimal VIs input.

Table 3. Maximum, minimum, and step length (SL) of ANN, RF, and SVM parameters.

Method	ANN			RF			SVM				
	Hidden layer 1	Hidden layer 2	SL	ntree	SL	mtry	SL	c	SL	g	SL
Min value	1	1	1	0	20	1	1	−10	0.5	−10	0.5
Max value	20	20	1	2000	20	10	1	10	0.5	10	0.5

Note: Only parameters to be optimized appear in this table; other parameters were determined as per Refs. [34,80,81].
Number of modeling and verification: ANN, $20 \times 20 \times 7 = 2800$; RF: $100 \times 10 \times 7 = 7000$; SVM: $40 \times 40 \times 7 = 11,200$.

Leave one sampling (LOS) was used to evaluate the performance of each technique, and global random sampling (GRS) and growth-period sampling (GPS) were used to evaluate the performance and stability of each technique with different sampling methods [81,82]. Global random sampling represents random samples from all samples, with a total of three samplings taken and denoted GRS1/3 (64 samples for modeling, the remaining 128 samples for validation), GRS1/2 (96 samples for modeling, the remaining 96 samples for validation), and GRS2/3 (128 samples for modeling, the remaining 64 samples for validation). Growth-period sampling uses random samples from each period and ensures an equal number of samples per birth period. Therefore, all samples from the whole growth period were divided into four layers, with each layer including one growth period. Again a total of three samplings were taken and are likewise denoted GPS1/3, GPS1/2, and GPS2/3. Leave one sampling represents leave-one-out cross-validation [83] in which only one sample is selected for verification, with all other samples taken as training samples (191 samples for modeling, the remaining one sample used for validation).

3.5. Precision Evaluation

We use the coefficient of determination R^2 , the root mean square error (RMSE), the mean absolute error (MAE), and normalized root mean square error (NRMSE) and coefficient of variance of the root mean square error [CV(RMSE)] to evaluate the accuracy of each technique. A larger R^2 value corresponds to a smaller RMSE, MAE, NRMSE, and CV(RMSE) and greater model accuracy. R^2 , RMSE, MAE, NRMSE, and CV(RMSE) are calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (3)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (4)$$

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}}{y_{\max} - y_{\min}} \quad (5)$$

$$\text{CV(RMSE)} = \frac{\sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}}{\bar{y}} \quad (6)$$

where x_i and y_i are the estimated and measured AGB values, respectively, y_{\max} and y_{\min} are the maximum and minimum measured values, respectively, \bar{x} and \bar{y} are the average estimated and measured values, respectively, and n is the sample number.

4. Results

The correlation coefficients r between the AGB and the VIs are shown in Table 4. The results show that all the measured VIs are correlated with biomass to varying degrees. Of the 54 VIs investigated, NPQI performs the best ($r = 0.757$). The correlation with AGB of the red-, green-, and blue-band-based spectral VIs [e.g., NPQI (0.757), BGI (0.555), TCARI (0.188), NPCI (0.474), BRI (0.519), and MCARI (0.237)] is greater than that of red- and near-infrared-band-based spectral VIs [e.g., MND₆₈₀ (0.193), MND₇₀₅ (0.086), NDVI (0.039), SR (0.029), EVI2 (0.182), OSAVI (0.088), and EVI (0.149)].

Table 4. Correlation coefficients r between AGB and VIs ($n = 192$).

Number	VI	r	Number	VI	r	Number	VI	r
1	NPQI	0.757 **	19	MCARI1	0.226 **	37	NDLI	0.106 ^{n.s.}
2	BGI	0.555 **	20	MTVI1	0.226 **	38	PSSR	0.102 ^{n.s.}
3	BRI	0.519 **	21	MND ₆₈₀	0.193 **	39	ATSAVI	0.092 ^{n.s.}
4	RVI _{hyp}	0.490 **	22	TCARI	0.188 **	40	OSAVI	0.088 ^{n.s.}
5	NPCI	0.474 **	23	EVI2	0.182 *	41	MND ₇₀₅	0.086 ^{n.s.}
6	CAI	0.442 **	24	PSND	0.180 *	42	RARS	0.079 ^{n.s.}
7	PVI _{hyp}	0.370 **	25	MSAVI	0.180 *	43	SIWSI	0.075 ^{n.s.}
8	LWVI2	0.352 **	26	TVI	0.177 *	44	NDWI	0.075 ^{n.s.}
9	LWVI1	0.337 **	27	GI	0.176 *	45	PBI	0.067 ^{n.s.}
10	SLAIDI	0.304 **	28	PRI	0.176 *	46	MSR ₇₀₅	0.061 ^{n.s.}
11	SRWI	0.300 **	29	VARI _{green}	0.165 *	47	NDVI	0.039 ^{n.s.}
12	LAI	0.397 **	30	SIPI	0.157 *	48	LCI	0.034 ^{n.s.}
13	RVSI	0.299 **	31	MCARI2	0.149 *	49	MSR	0.032 ^{n.s.}
14	WI	0.260 **	32	PSRI	0.149 *	50	WDRVI	0.031 ^{n.s.}
15	SPVI	0.251 **	33	RDVI	0.146 *	51	SR	0.029 ^{n.s.}
16	ARI	0.241 **	34	RGR	0.142 *	52	MSI	0.029 ^{n.s.}
17	MCARI	0.237 **	35	EVI	0.149 *	53	DSWI	0.027 ^{n.s.}
18	NDNI	0.228 **	36	MTVI2	0.121 *	54	NDII	0.012 ^{n.s.}

Note: Probability levels are indicated by ^{n.s.}, *, and ** for “not significant” (up to 0.119), 0.05 (greater than 0.141), and 0.01 (greater than 0.185), respectively. ** r (0.01, 192) = 0.185; * r (0.05, 192) = 0.141.

The correlation coefficients r among the 54 VIs are shown in Figure 5. For each VI, there are 53 colors that represent different correlation coefficient values. The results (Figure 5) show that complex correlations exist among these 54 VIs. NPQI (first), BGI (second), ARI (16th) and TCARI (22nd) are low correlated with other VIs. Among the top 22 VIs, a low correlation (zone a) was observed, but the remaining 32 VIs have a high correlation (zone b). The VIs analysis results of total explained variance and variance inflation factor (VIF) values are shown in Tables A1 and A2 (in Appendix A).

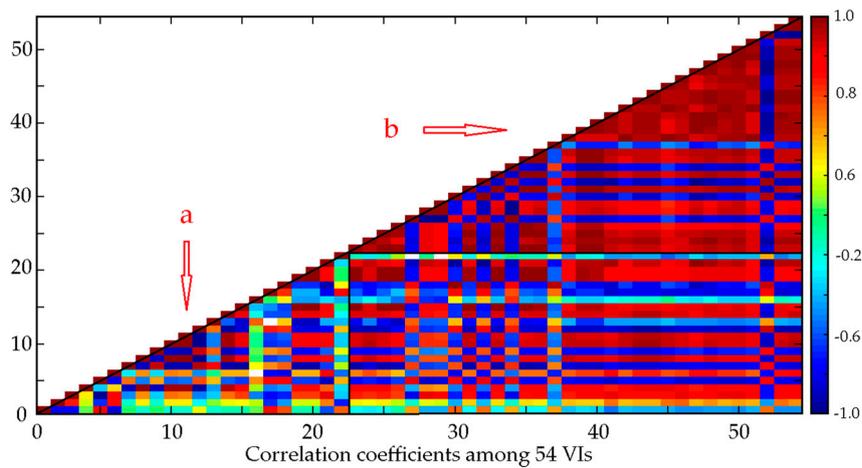


Figure 5. Correlation coefficients r among 54 VIs. NPQI (first), BGI (second), ARI (16th) and TCARI (22nd) are weakly correlated with other VIs. Zone a (orange and light blue), low correlation zone; Zone b (red and dark blue), high correlation zone. VIs are ordered according to corr. coeff to AGB.

4.1. Selection of Vegetation Indexes

The best AGB models and the associated validation accuracy of the eight techniques are shown in Figure 6. After a different number of VIs (Table 5) were incorporated into the modeling, the validation accuracy of ANN, BBRT, DT, and RF (Figure 6a–c,g) flattens out. BBRT and DT performed well when using the top five VIs as the input, although the validation accuracy decreases slightly after using lower-correlation VIs as the input for modeling (Figure 6b,c). The performance of ANN, PCR, and SVM becomes complex (Figure 6a,f,h) after using lower-correlation VIs as the input for modeling.

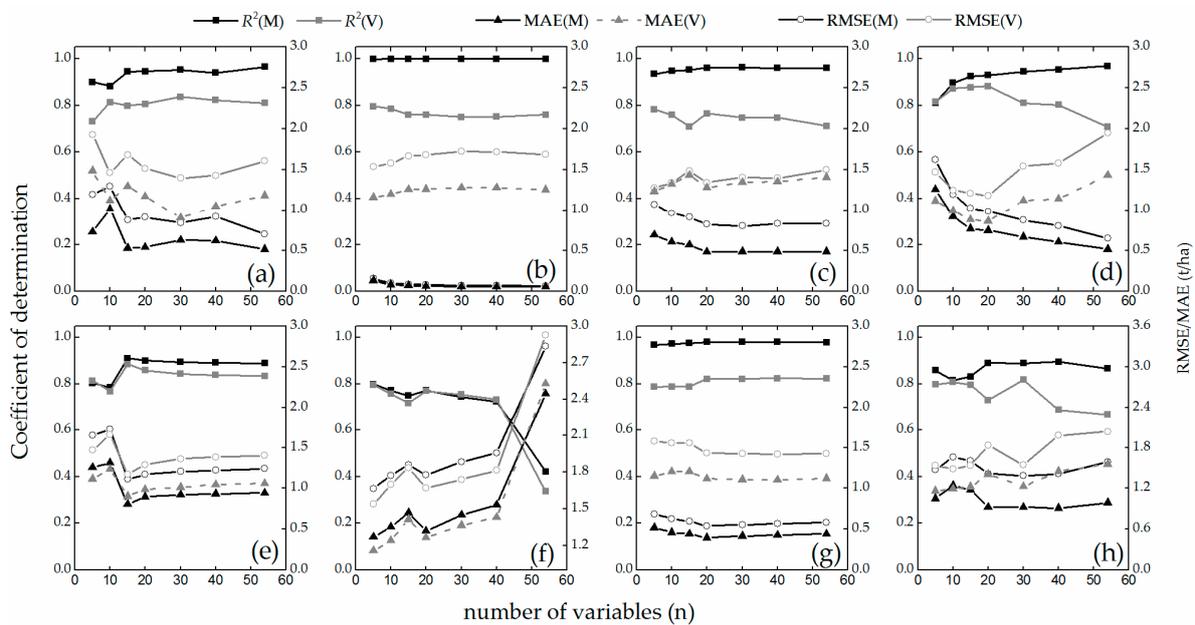


Figure 6. Best AGB modeling and validation accuracy for eight techniques with different input VIs: (a) ANN, (b) BBRT, (c) DT, (d) MLR, (e) PLSR, (f) PCR, (g) RF, (h) SVM. Note: R^2 (M) and R^2 (V) indicate R^2 for modeling and validation, respectively. The same notation is used for MAE (M) and MAE (V), RMSE (M), and RMSE (V).

Table 5. Optimal number of input VIs for the eight analytical techniques.

Technique	ANN	BBRT	DT	MLR	PLSR	PCR	RF	SVM
Number of input VIs	30	5	5	20	15	5	30	30

Note: the number of input VIs (n) represents the top n VIs in Table 4.

The results, given in Table 5, are the optimum numbers of input VIs determined by the change of validation accuracy (Figure 6). The numbers were confirmed when the best estimation accuracy was obtained. For example, with the input modeling of the top 30 VIs, the highest accuracy was acquired (Figure 6a). Accordingly, for RF, the parameters $n_{tree} = 520$ and $m_{try} = 8$, the ANN hidden layer is 10 and 2; $c = 10$ and $g = -2.5$ for LIBSVM. In PCR modeling, we use 85% as the threshold to determine the principal component, and when we use the top five VIs as the input, the cumulative variance of the top two principal components is 89.136%.

Considering that the correlation coefficients r among VIs were high (Figure 5), multi-collinearity may be a problem when using so many VIs for modeling. The VIs analysis results of total explained variance (Table A1) and VIF values (Table A2) also support this view. The results in Figure 6 show that different techniques have varying abilities in tackling multi-collinearity data. ANN, BBRT and RF show a good performance when dealing with collinear data (Figure 6a,b,g), which performed well in 30 VIs, 40 VIs and 54 VIs group modeling and validation. Greater modeling accuracy was obtained with MLR after using more VIs as the input for modeling (Figure 6d), but validation accuracy decreased, especially after more than 20 VIs were input (R^2 (V), MAE (V) and RMSE (V) in Figure 6d).

4.2. Test with White Gaussian Noise

A comparative analysis of different estimation accuracies resulting from the eight selected techniques with white Gaussian noise (different SNR values) is presented in Figure 7. For each technique, the three figures (Figure 7) represent R^2 , RMSE and MAE with different SNR.

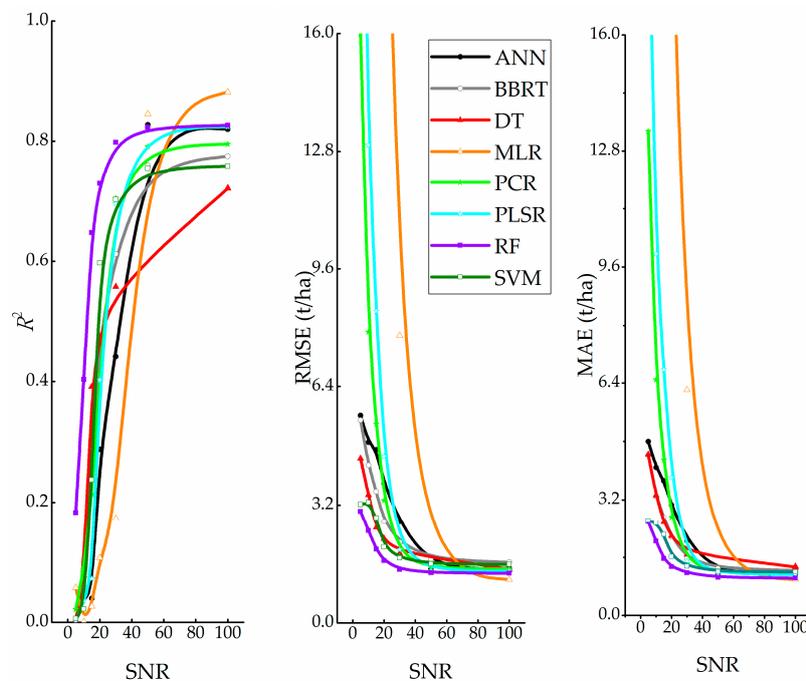


Figure 7. Three measures of accuracy as a function of the signal-to-noise ratio. Note: SNR 100 represents no noise. All techniques used the first 30 sets of VIs (Table 4) as the input, with SNR = 5, 10, 15, 20, 30, 50, and no noise.

Figure 7 shows the noise immunity for the eight analytical techniques. The results indicate RF > SVM > DT > BBRT > ANN > PCR > PLSR > MLR. RF performs best in this test with a validation R^2 near 0.2 and SNR = 5. For MLR, poor noise immunity was observed; its validation accuracy declined from SNR = 30 (Figure 7). Meanwhile, the validation of MLR (R^2 (<0.2), RMSE (about 8 t/ha) and MAE (about 6 t/ha)) is the worst in eight techniques. As the noise increases (Figure 7, SNR <30), MLR, PCR and PLSR are extremely sensitive to it, whereas ANN, SVM, DT, and BBRT are more robust against noise; however, the latter techniques also show a poor performance with increased levels of noise (Figure 7, SNR = 5).

4.3. Stability Test

Figure 8 shows the results for modeling and validation using GRS1/3, GRS1/2, and GRS2/3 with the parameters given in Table 5. The calculated absolute difference of AGB – GRS modeling and the validation accuracy (we use ∇R^2 , $\nabla RMSE$ and ∇MAE to show the difference) are given in Table 6.

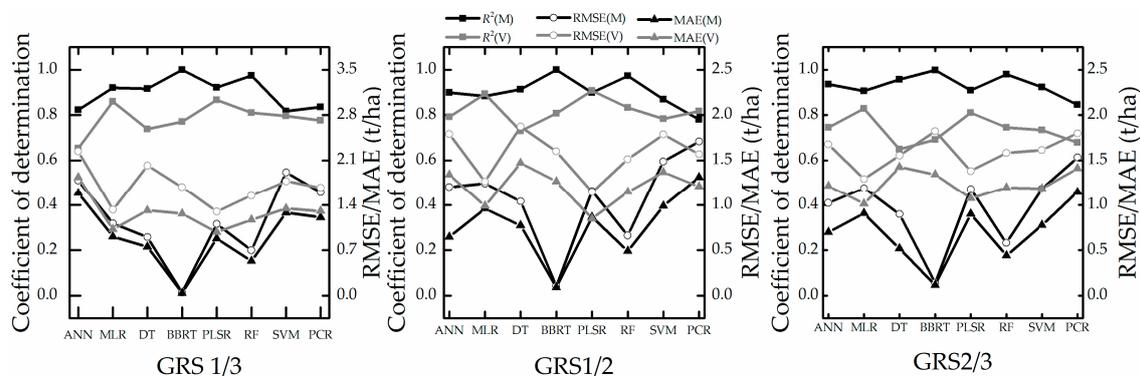


Figure 8. Modeling and validation results for global random sampling (GRS).

Table 6. Absolute difference of AGB – GRS modeling and accuracy of validation.

Technique	∇R^2			$\nabla RMSE$ (t/ha)			∇MAE (t/ha)		
	1/3	1/2	2/3	1/3	1/2	2/3	1/3	1/2	2/3
ANN	0.17 *	0.11	0.19 *	0.45	0.59	0.65	0.25	0.69	0.52
MLR	0.06	0.01	0.08	0.21	0.02	0.10	0.11	0.03	0.10
DT	0.18*	0.18*	0.31 *	1.12 *	0.83 *	0.66	0.56	0.70	0.91 *
BBRT	0.23 *	0.19*	0.31 *	1.64 *	1.49 *	1.68 *	1.23 *	1.17 *	1.23 *
PLSR	0.05	0.01	0.10	0.19	0.01	0.20	0.10	0.02	0.17
RF	0.16*	0.14	0.24 *	0.85 *	0.85*	1.00 *	0.64	0.66	0.76
SVM	0.02	0.09	0.19 *	0.14	0.30	0.43	0.06	0.38	0.41
PCR	0.06	0.04	0.17 *	0.07	0.14	0.26	0.10	0.11	0.26

Note: $\nabla R^2 > 0.15$, $\nabla RMSE$ and $\nabla MAE > 0.800$ t/ha are marked with *. ∇R^2 , $\nabla RMSE$, and ∇MAE show the absolute difference between modeling and validation R^2 , RMSE, MAE, respectively. The smaller the difference, the more stable and reliable the technique.

The difference in the modeling and validation results with eight techniques for global random sampling varies (Figure 8). For MLR and PLSR, a stable performance was observed with three sampling methods (Table 6). SVM and PCR have a poorer performance than MLR and PLSR in this test (Figure 8 and Table 6). For ANN, RF, DT and BBRT, the difference in modeling and validation accuracy is huge (Figure 8 and Table 6). As the sampling method changes (1/3, 1/2, 2/3), these four techniques still show a poor performance in all investigated techniques. In addition, the validation accuracy of almost all techniques is lower than modeling accuracy (Figure 8).

The modeling and validation results for GPS1/3, GPS1/2, and GPS2/3 are shown in Figure 9. The absolute difference of AGB – GRS between modeling and validation accuracy (we use ∇R^2 , $\nabla RMSE$ and ∇MAE to show the difference) appears in Table 7.

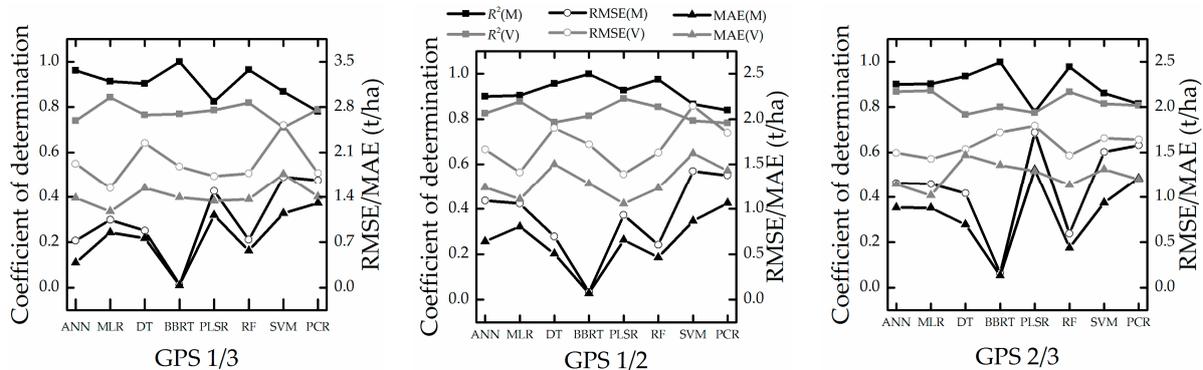


Figure 9. Modeling and validation results for growth period sampling (GPS).

Table 7. Absolute difference AGB – GPS between modeling and validation accuracy.

Technique	∇R^2			$\nabla RMSE$ (t/ha)			∇MAE (t/ha)		
	1/3	1/2	2/3	1/3	1/2	2/3	1/3	1/2	2/3
ANN	0.22 *	0.07	0.03	1.20 *	0.57	0.34	1.00 *	0.60	0.26
MLR	0.07	0.03	0.03	0.50	0.35	0.28	0.32	0.30	0.14
DT	0.14	0.17 *	0.17 *	1.37 *	1.20 *	0.50	0.78	1.00 *	0.77
BBRT	0.23 *	0.19 *	0.20 *	1.83 *	1.64 *	1.56 *	1.36 *	1.21 *	1.23 *
PLSR	0.04	0.04	0.00	0.23	0.46	0.07	0.22	0.40	0.02
RF	0.14	0.12	0.11	1.04 *	1.03 *	0.87 *	0.80 *	0.77	0.70
SVM	0.16 *	0.07	0.05	0.81 *	0.72	0.15	0.62	0.76	0.37
PCR	0.01	0.06	0.01	0.11	0.47	0.06	0.10	0.36	0.01

Note: $\nabla R^2 > 0.15$, $\nabla RMSE$ and $\nabla MAE > 0.800$ t/ha are marked with *. ∇R^2 , $\nabla RMSE$, and ∇MAE show the absolute difference between modeling and validation R^2 , RMSE, MAE, respectively. The smaller the difference, the more stable and reliable the technique.

The results showed that all techniques perform differently in different modeling and validation sample sizes (Tables 6 and 7, from 1/3 to 1/2 and 2/3). We observed a stable performance in MLR, PLSR and PCR (Figure 9 and Table 7) in the GPS sampling method. As modeling numbers increased (1/3, 1/2, 2/3), MLR and PLSR performed more stably [such as MLR: (∇R^2 : 0.07, 0.03, 0.03; $\nabla RMSE$: 0.50, 0.35, 0.28; ∇MAE : 0.32, 0.30, 0.14) in Table 7]; ANN, PLSR, RF and SVM with GPS sampling showed a clear growth in stability (Tables 6 and 7). The above results indicate that the GPS sampling method is more effective than GRS for obtaining stable estimation models.

4.4. Estimation Accuracy with Leave One Sampling

The results for validation with LOS appear in Figure 10. PLSR provides the highest accuracy for AGB–LOS modeling. The LOS validation results shown in Figure 10 suggest that all of these techniques have impressive accuracy: the R^2 values are at least 0.79 [PCR with RMSE = 1.63 t/ha, MAE = 1.24 t/ha, NRMSE = 0.10, CV(RMSE) = 0.25], and PLSR [$R^2 = 0.89$, RMSE = 1.20 t/ha, MAE = 0.90 t/ha, NRMSE = 0.07, CV(RMSE) = 0.18] has the highest accuracy. The leave-one-out cross validation indicates that the prediction performance of these techniques can be ranked as (Figure 10i) PLSR > MLR > RF > SVM > BBRT > ANN > DT > PCR.

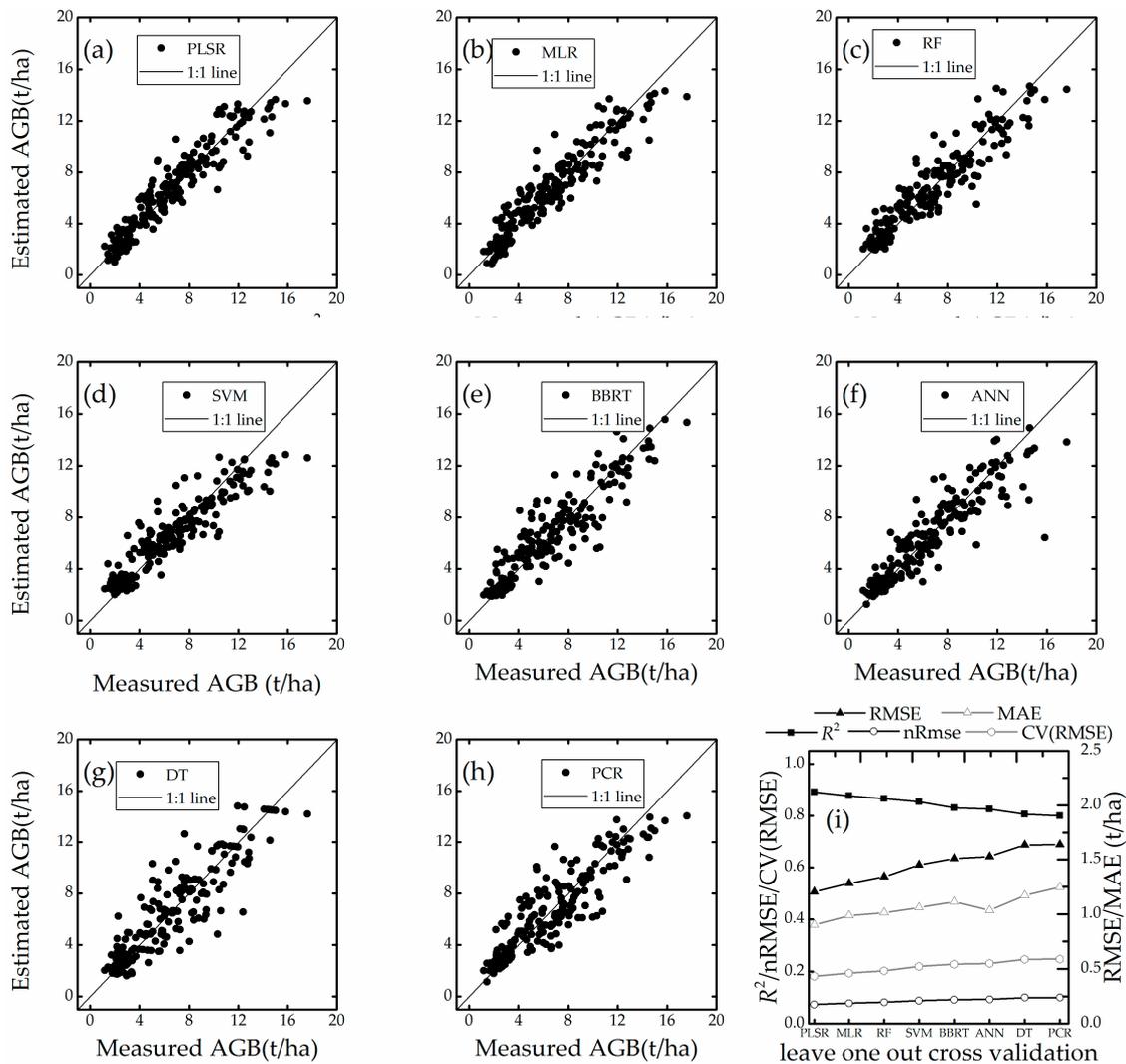


Figure 10. Measured and estimated AGB using leave one sampling. (a): PLSR; (b): MLR; (c): RF; (d): SVM; (e): BBRT; (f): ANN; (g): DT; (h): PCR; (i): R^2 , RMSE, MAE, NRMSE and CV(RMSE) of all techniques.

5. Analysis and Discussion

5.1. Analysis and Selection of Vegetation Indexes

Our correlation analysis shows that VIs are correlated with AGB to varying degrees. The results of the correlation analysis (Table 4) demonstrate that only the correlation of the top 20 VIs exceeds 0.2. Previous studies have shown that near-infrared- and red-band VIs are effective for estimating AGB [8,9,11]. The correlation between AGB and red- and near-infrared-band VIs is low in this study (Table 4). This may be because, during the reproductive stage, photosynthesis and the near-infrared reflectance [84] both clearly decrease (Figure 2), reducing the correlation between AGB and the red- and near-infrared-based VIs. This result is consistent with the results of a previous study [17]. By contrast, the correlation between the entire growth stage of winter wheat AGB estimates and the red-, green-, and blue-band spectral indexes is more promising (Table 4). Our study also demonstrated that these vegetation indices are effective in estimating AGB (Figure 6, top five VIs input). Thus, red-, green-, and blue-band spectral indexes are useful as they can be used to estimate AGB during vegetative growth and reproductive growth stages.

A serious multi-collinearity problem arises in the investigation of 54 VIs (Figure 5, Tables A1 and A2); our results (Figure 6) show that the verification accuracy of eight techniques differs when fed multi-collinearity data (Figure 5) as the input. Previous studies have shown that machine learning techniques (ANN [81], BBRT [24] and RF [21,85]) can make full use of the narrow hyperspectral bands (strongly collinear data) and VIs. In the current study, ANN, BBRT and RF are almost unaffected by using multi-collinearity data (Figure 6a,b,g), which may indicate that these techniques are robust against noise, which may relate to the principles of these techniques. Garg et al. [86] indicated that the machine learning technique is suitable for tackling the multi-collinearity problem; our results showed that machine learning techniques have better abilities in tackling the multi-collinearity problem than that of conventional regression techniques (Figure 6). In addition, the results show that MLR has a poor performance when using multi-collinearity data to estimate AGB, which confirms the results of a previous study [87]; PLSR performs best in three conventional regression techniques for tackling the multi-collinearity problem, which confirms the results of a previous study [87]. Thus, PLSR is a useful tool that can be used to estimate several response variables simultaneously, while accounting for multi-collinearity variables [88].

5.2. Analysis of Noise Immunity

Our results show that machine learning techniques are more immune to powerful noise than conventional regression techniques (Figure 7); RF performs best in this noise test. This may be because the RF method randomly changes the input variable and validates the importance of the input data, thus generating a large number of decision trees and reducing the impact of noise; this result corresponds to the results of previous studies [25,89,90]. Our results (Figure 7) show that MLR is more sensitive to noise than PLSR, which is consistent with the findings of Zhao et al. [91]. The results of Atzberger et al. [7] indicate that the noise immunity of PCR, PLSR, and SMLR is ranked as PCR > PLSR > MLR, which is exactly the same ranking as obtained in the present work (Figure 7). The present noise immunity results (Figure 7) are important because repeated observations by remote-sensing techniques occur at different times; so techniques with poor noise immunity may lead to low accuracy because of data errors [26,78,79] (e.g., due to atmosphere, clouds, observation times, sensor noise). Our noise immunity results may explain why different regression studies of vegetation parameters based on remote sensing obtain significantly different results.

5.3. Analysis of Stability and Prediction Performance

PLSR and MLR both perform better in stability tests than machine learning techniques (Tables 6 and 7, from 1/3 to 1/2 and 2/3 sampling). Farifteh et al. [92] indicated that PLSR performs more stably in soil salinity estimation than ANN (PLSR: R^2 : 0.6~0.98, RMSE% = 11.6~48%; ANN: R^2 = 0.46~0.97, RMSE% = 12.5~57%). Thus, PLSR and MLR may be suitable for works in which fewer samples are available for modeling. BBRT and DT perform poorly in stability tests, and the AGB estimation model seems to be over fit because R^2 is close to unity in all tests (Figure 6b,c). However, the validation accuracy of BBRT remains high, whereas that of DT is poor. Fewer studies evaluating DT for AGB estimation are available, which may be because DT does not deliver high accuracy for AGB estimation by remote sensing. Yuan et al. [81] indicated that the accuracy of the simple random sampling method is lower than stratified sampling, and our results are in agreement with that study; our results also indicate that all GPS models are more stable than GRS with 2/3 sampling (Table 6 GRS 2/3, and Table 7 GPS 2/3). This may be because the inappropriate sample selection method affects modeling and validation accuracy, which may indicate that GPS sampling is more suitable for these techniques. A previous study showed that stratified sampling helps to generate a good calibration set [82]; this may explain why GPS performed better than GRS in this study.

PLSR has the highest accuracy in leave-one-out cross validation (Figure 10a), while PCR has the lowest precision (Figure 10h). This comparison between PLSR and PCR is consistent with the results of Atzberger et al. [7], who estimated the aboveground-canopy Chl content (PLSR: R^2 = 0.85, RMSE = 51;

PCR: $R^2 = 0.57$, RMSE = 82). In addition, the comparison herein between PLSR, ANN, and PCR is consistent with that of Mirzaie et al. [22] in their estimate of the water content of vegetation (PLSR: $R^2 = 0.93$, RRMSE = 0.23; ANN: $R^2 = 0.83$, RRMSE = 0.41; PCR: $R^2 = 0.78$, RRMSE = 0.41). Thus, PLSR is a useful tool that can be used to estimate AGB with high accuracy.

6. Conclusions

We have provided herein a series of machine learning and conventional regression techniques to estimate hyperspectral winter wheat AGB and select input data for the sampling methods. We have also analyzed the noise immunity and prediction accuracy. The results allow the following conclusions to be drawn:

- (1) Machine learning is the correct technique for tackling the multi-collinearity problem. ANN, BBRT and RF are almost unaffected by the multi-collinearity problem (Figure 6a,b,g), while MLR and PCR could not solve it.
- (2) Machine learning techniques are much more immune to noise than conventional regression techniques. In terms of noise immunity, the techniques are ranked as follows (Figure 7): RF > SVM > DT > BBRT > ANN > PCR > PLSR > MLR. Thus, RF may be suitable for work that requires repeated observations via remote sensing.
- (3) The growth-period random sampling method performed better in stability tests. PLSR and MLR perform well in all stability tests (Figures 8 and 9 and Tables 6 and 7); these techniques and the sampling method may be suitable for work in which only a few samples are available for high-accuracy and stability estimation modeling.
- (4) This study demonstrated the potential application of VIs, machine learning and conventional regression techniques in estimating winter wheat biomass. The experimental results indicated that PLSR, MLR, and RF may be suitable for work that requires high-accuracy estimation models.

Acknowledgments: This study was supported by the National Key Research and Development Program (2016YFD0300603-5, 2016YFD0300602) the Natural Science Foundation of China (41601346, 41601369, 61661136003, 41771370, 41471285, 41471351), U.K. Science and Technology Facilities Council through the PAFiC project: Precision Agriculture for Family-farms in China (Ref.: ST/N006801/1), and the Special Funds for Technology innovation capacity building sponsored by the Beijing Academy of Agriculture and Forestry Sciences (KJ CX20170423).

Author Contributions: Haikuan Feng, Zhenhai Li and Guijun Yang designed the experiments, Zhenhai Li, Jibo Yue and Haikuan Feng collected the AGB, Chl and ASD hyperspectral data. Jibo Yue and Haikuan Feng analyzed the data and wrote the manuscript, Guijun Yang and Zhenhai Li provided comments and suggestions for the manuscript and checked the writing.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Principal component analysis of the VIs was conducted by SPSS software (Statistical Product and Service Solutions, IBM, Armonk, New York, NY, USA). Total explained variance is shown in Table A1. For each group of VIs, the different values represent the total explained variance with different components.

Table A1. Total explained variance of each group of VIs (%).

Component	54VIs	40VIs	30VIs	20VIs	15VIs	10VIs	5VIs
1	75.891	70.822	65.584	62.495	65.402	60.494	68.135
2	85.003	82.513	80.511	81.348	87.255	89.711	89.136
3	92.211	91.435	91.553	92.157	94.391	94.269	97.232
4	95.228	94.586	94.757	94.774	96.541	97.199	99.927
5	97.350	97.109	97.179	97.235	98.056	98.596	100.000
6	98.149	98.025	98.226	98.246	99.039	99.514	-
7	98.706	98.683	98.841	98.817	99.483	99.887	-
8	99.217	99.137	99.242	99.306	99.779	99.956	-
9	99.425	99.391	99.551	99.662	99.881	99.985	-
10	99.614	99.633	99.696	99.824	99.940	100.000	-
11	99.726	99.746	99.801	99.889	99.965	-	-
12	99.817	99.829	99.870	99.927	99.986	-	-
13	99.872	99.885	99.903	99.958	99.994	-	-
14	99.902	99.914	99.931	99.974	100.000	-	-
15	99.927	99.933	99.951	99.988	100.000	-	-
16	99.942	99.950	99.967	99.994	-	-	-
17	99.956	99.963	99.979	99.998	-	-	-
18	99.966	99.975	99.985	100.000	-	-	-
19	99.974	99.984	99.989	100.000	-	-	-
20	99.980	99.989	99.994	100.000	-	-	-
21	99.985	99.992	99.996	-	-	-	-
22	99.988	99.994	99.998	-	-	-	-
23	99.991	99.996	99.998	-	-	-	-
24	99.993	99.997	99.999	-	-	-	-
25	99.994	99.998	100.000	-	-	-	-
26	99.996	99.999	100.000	-	-	-	-
27	99.996	99.999	100.000	-	-	-	-
28	99.997	99.999	100.000	-	-	-	-
29	99.998	100.000	100.000	-	-	-	-
30	99.998	100.000	100.000	-	-	-	-
31	99.999	100.000	-	-	-	-	-
32	99.999	100.000	-	-	-	-	-
33	99.999	100.000	-	-	-	-	-
34	99.999	100.000	-	-	-	-	-
35	100.000	100.000	-	-	-	-	-
36	100.000	100.000	-	-	-	-	-
37	100.000	100.000	-	-	-	-	-
38	100.000	100.000	-	-	-	-	-
39	100.000	100.000	-	-	-	-	-
40	100.000	100.000	-	-	-	-	-
41	100.000	-	-	-	-	-	-
42	100.000	-	-	-	-	-	-
43	100.000	-	-	-	-	-	-
44	100.000	-	-	-	-	-	-
45	100.000	-	-	-	-	-	-
46	100.000	-	-	-	-	-	-
47	100.000	-	-	-	-	-	-
48	100.000	-	-	-	-	-	-
49	100.000	-	-	-	-	-	-
50	100.000	-	-	-	-	-	-
51	100.000	-	-	-	-	-	-
52	100.000	-	-	-	-	-	-
53	100.000	-	-	-	-	-	-
54	100.000	-	-	-	-	-	-

Note: 54, 40, 30, 20, 15 and 5 represent each group of data; Components 1~54 represent the first to 54th component, respectively. Symbol “-” stand for “None”.

VIF (Table A2) provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity. VI 1~VI 54 represent different VIs in Table 4 which fed into MLR modeling and validation.

Table A2. Variance inflation factor (VIF) of VIs.

VI	54VIs	VI	40VIs	VI	30VIs	VI	20VIs	VI	15VIs	VI	10VIs	VI	5VIs
1	121.3	1	84.0	1	46.6	1	35.0	1	20.7	1	13.7	1	7.1
2	793.8	2	400.9	2	323.2	2	28.3	2	7.2	2	6.0	2	3.3
3	1118.0	3	615.0	3	511.4	3	317.0	3	177.2	3	143.8	3	121.2
4	492.3	4	399.1	4	346.3	4	203.3	4	173.3	4	94.1	4	3.5
5	2673.3	5	2015.7	5	1492.1	5	471.6	5	258.9	5	215.0	5	150.6
6	16.2	6	14.0	6	13.9	6	11.7	6	9.1	6	6.5	-	-
7	1122.8	7	1526.1	7	1190.1	7	395.0	7	326.8	7	19.9	-	-
8	2441.3	8	1548.6	8	1290.8	8	1023.6	8	779.8	8	287.1	-	-
9	172.7	9	114.8	9	79.4	9	41.8	9	32.8	9	7.7	-	-
11	2138.9	11	1271.3	11	828.0	11	497.9	11	438.6	10	408.9	-	-
12	1637.2	12	1188.4	12	1133.6	12	805.6	12	616.8	-	-	-	-
13	358.1	13	1227.3	13	956.0	13	86.9	13	29.4	-	-	-	-
14	175.4	14	156.0	14	148.3	14	111.3	14	72.1	-	-	-	-
15	3285.7	15	3332.0	15	3899.8	15	1486.7	15	146.8	-	-	-	-
16	191.9	16	101.7	16	85.5	16	39.2	-	-	-	-	-	-
17	1130.6	17	1128.6	17	918.1	17	23.0	-	-	-	-	-	-
18	156.2	18	118.5	18	79.0	18	18.7	-	-	-	-	-	-
21	6258.6	21	3884.6	21	886.0	20	1280.1	-	-	-	-	-	-
22	663.8	22	572.1	22	470.4	-	-	-	-	-	-	-	-
24	996.9	24	706.2	24	207.5	-	-	-	-	-	-	-	-
25	3304.2	26	8495.2	26	2382.6	-	-	-	-	-	-	-	-
28	5845.4	27	2114.8	27	7627.7	-	-	-	-	-	-	-	-
30	1939.9	28	3010.1	28	1554.9	-	-	-	-	-	-	-	-
32	8499.5	30	1287.1	30	2371.2	-	-	-	-	-	-	-	-
34	3063.2	32	6092.3	32	177.4	-	-	-	-	-	-	-	-
37	59.0	37	42.0	-	-	-	-	-	-	-	-	-	-
38	2705.9	38	148.6	-	-	-	-	-	-	-	-	-	-
41	4203.8	40	4259.2	-	-	-	-	-	-	-	-	-	-
45	4723.4	-	-	-	-	-	-	-	-	-	-	-	-
46	2086.8	-	-	-	-	-	-	-	-	-	-	-	-
50	4010.9	-	-	-	-	-	-	-	-	-	-	-	-
51	1151.1	-	-	-	-	-	-	-	-	-	-	-	-
52	4666.7	-	-	-	-	-	-	-	-	-	-	-	-
54	3874.4	-	-	-	-	-	-	-	-	-	-	-	-

Note: 54VIs, 40VIs, 30VIs, 20VIs, 15VIs and 5VIs represent different VIF values of each VI in four groups of data. Symbol “-” stand for “None”.

References

1. Wang, J.; Zhao, C.; Huang, W. *Fundamental and Application of Quantitative Remote Sensing in Agriculture*; Science China Press: Beijing, China, 2008.
2. Jin, X.; Li, Z.; Yang, G.; Yang, H.; Feng, H.; Xu, X.; Wang, J.; Li, X.; Luo, J. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 24–37. [[CrossRef](#)]
3. Huang, J.; Sedano, F.; Huang, Y.; Ma, H.; Li, X.; Liang, S.; Tian, L.; Zhang, X.; Fan, J.; Wu, W. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agric. For. Meteorol.* **2016**, *216*, 188–202. [[CrossRef](#)]
4. Hensgen, E.; Böhle, L.; Wachendorf, M. The effect of harvest, mulching and low-dose fertilization of liquid digestate on above ground biomass yield and diversity of lower mountain semi-natural grasslands. *Agric. Ecosyst. Environ.* **2016**, *216*, 283–292. [[CrossRef](#)]

5. Jin, X.; Kumar, L.; Li, Z.; Xu, X.; Yang, G.; Wang, J. Estimation of winter wheat biomass and yield by combining the aquacrop model and field hyperspectral data. *Remote Sens.* **2016**, *8*. [[CrossRef](#)]
6. Boschetti, M.; Bocchi, S.; Brivio, P.A. Assessment of pasture production in the Italian Alps using spectrometric and remote sensing information. *Agric. Ecosyst. Environ.* **2007**, *118*, 267–272. [[CrossRef](#)]
7. 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. [[CrossRef](#)]
8. Fu, Y.; Yang, G.; Wang, J.; Song, X.; Feng, H. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. *Comput. Electron. Agric.* **2014**, *100*, 51–59. [[CrossRef](#)]
9. Yue, J.; Yang, G.; Li, C.; Li, Z.; Wang, Y.; Feng, H.; Xu, B. Estimation of Winter Wheat Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Snapshot Hyperspectral Sensor and Crop Height Improved Models. *Remote Sens.* **2017**, *9*, 708. [[CrossRef](#)]
10. Atzberger, C.; Darvishzadeh, R.; Immitzer, M.; Schlerf, M.; Skidmore, A.; le Maire, G. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *43*, 19–31. [[CrossRef](#)]
11. Cho, M.A.; Skidmore, A.; Corsi, F.; van Wieren, S.E.; 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. Geoinf.* **2007**, *9*, 414–424. [[CrossRef](#)]
12. Schlerf, M.; Atzberger, C.; Hill, J. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens. Environ.* **2005**, *95*, 177–194. [[CrossRef](#)]
13. Galvão, L.S.; Formaggio, A.R.; Tisot, D.A. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. *Remote Sens. Environ.* **2005**, *94*, 523–534. [[CrossRef](#)]
14. Wang, W. *The Associations of Photosynthesis and Grain Filling during Grain Filling Period in Flag Leaves of Wheat Species*; Ocean University of China: Qingdao, China, 2007.
15. Datt, B. A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves. *J. Plant Physiol.* **1999**, *154*, 30–36. [[CrossRef](#)]
16. 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. [[CrossRef](#)]
17. Sun, H.; Li, M.Z.; Zhao, Y.; Zhang, Y.E.; Wang, X.M.; Li, X.H. The Spectral Characteristics and Chlorophyll Content at Winter Wheat Growth Stages. *Spectrosc. Spectr. Anal.* **2010**, *30*, 192–196. [[CrossRef](#)]
18. Thenkabail, P.S. Biophysical and yield information for precision farming from near-real-time and historical Landsat TM images. *Int. J. Remote Sens.* **2003**, *24*, 2879–2904. [[CrossRef](#)]
19. Rivera, J.; Verrelst, J.; Delegido, J.; Veroustraete, F.; Moreno, J. On the Semi-Automatic Retrieval of Biophysical Parameters Based on Spectral Index Optimization. *Remote Sens.* **2014**, *6*, 4927–4951. [[CrossRef](#)]
20. Atzberger, C. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sens. Environ.* **2004**, *93*, 53–67. [[CrossRef](#)]
21. Yue, J.; Yang, G.; Feng, H. Comparative of remote sensing estimation models of winter wheat biomass based on random forest algorithm. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 175–182. [[CrossRef](#)]
22. 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. [[CrossRef](#)]
23. Du, H.; Sun, X.; Han, N.; Mao, F. RS estimation of inventory parameters and carbon storage of Moso bamboo forest based on synergistic use of object-based image analysis and decision tree. *Chin. J. Appl. Ecol.* **2017**, *28*. [[CrossRef](#)]
24. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
25. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
26. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sens.* **2015**, *7*, 16398–16421. [[CrossRef](#)]

27. Cheng, J.H.; Dai, Q.; Sun, D.W.; Zeng, X.A.; Liu, D.; Pu, H. Bin Applications of non-destructive spectroscopic techniques for fish quality and safety evaluation and inspection. *Trends Food Sci. Technol.* **2013**, *34*, 18–31. [[CrossRef](#)]
28. Pan, L.; Li, H.C.; Deng, Y.J.; Zhang, F.; Chen, X.D.; Du, Q. Hyperspectral dimensionality reduction by tensor sparse and low-rank graph-based discriminant analysis. *Remote Sens.* **2017**, *9*. [[CrossRef](#)]
29. Gomez, C.; Lagacherie, P.; Coulouma, G. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma* **2008**, *148*, 141–148. [[CrossRef](#)]
30. Blackburn, G.A.; Pitman, J.I. Biophysical controls on the directional spectral reflectance properties of bracken (*Pteridium aquilinum*) canopies: Results of a field experiment. *Int. J. Remote Sens.* **1999**, *20*, 2265–2282. [[CrossRef](#)]
31. 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*, 1385–1392. [[CrossRef](#)]
32. Pu, R.; Gong, P. Hyperspectral Remote Sensing of Vegetation Bioparameters. In *Advances in Environmental Remote Sensing*; CRC Press: Boca Raton, FL, USA, 2011; Volume 4, pp. 101–142, ISBN 9781420091816.
33. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* **2000**, *71*, 158–182. [[CrossRef](#)]
34. Liang, L.; Di, L.; Zhang, L.; Deng, M.; Qin, Z.; Zhao, S.; Lin, H. Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sens. Environ.* **2015**, *165*, 123–134. [[CrossRef](#)]
35. Gevrey, M.; Dimopoulos, I.; Lek, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.* **2003**, *160*, 249–264. [[CrossRef](#)]
36. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
37. Chang, C.-C.; Lin, C.-J. Libsvm. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [[CrossRef](#)]
38. Utgoff, P.E.; Corporation, C.; Street, W. Decision Tree Induction Based on Efficient Tree Restructuring. *Mach. Learn.* **1997**, *29*, 5–44. [[CrossRef](#)]
39. Fakiola, M.; Mishra, A.; Rai, M.; Singh, S.P.; O’Leary, R.A.; Ball, S.; Francis, R.W.; Firth, M.J.; Radford, B.T.; Miller, E.N.; et al. Classification and regression tree and spatial analyses reveal geographic heterogeneity in genome wide linkage study of indian visceral leishmaniasis. *PLoS ONE* **2010**, *5*, e15807. [[CrossRef](#)] [[PubMed](#)]
40. Wold, H. Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis*; Academic Press: New York, NY, USA, 1966; pp. 1391–1420, ISBN 0471411256.
41. Baret, F.; Guyot, G. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sens. Environ.* **1991**, *35*, 161–173. [[CrossRef](#)]
42. Sims, D.A.; Gamon, J.A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* **2002**, *81*, 337–354. [[CrossRef](#)]
43. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
44. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [[CrossRef](#)]
45. Zarco-Tejada, P.J.; Berjón, A.; López-Lozano, R.; Miller, J.R.; Martín, P.; Cachorro, V.; González, M.R.; De Frutos, A. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* **2005**, *99*, 271–287. [[CrossRef](#)]
46. Peñuelas, J.; Gamon, J.A.; Fredeen, A.L.; Merino, J.; Field, C.B. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sens. Environ.* **1994**, *48*, 135–146. [[CrossRef](#)]
47. 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. [[CrossRef](#)]
48. Barnes, J.D.; Balaguer, L.; Manrique, E.; Elvira, S.; Davison, A.W. A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. *Environ. Exp. Bot.* **1992**, *32*, 85–100. [[CrossRef](#)]

49. 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. [CrossRef]
50. Rama Rao, N.; Garg, P.K.; Ghosh, S.K.; Dadhwal, V.K. Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery. *J. Agric. Sci.* **2008**, *146*, 65–75. [CrossRef]
51. Chen, J.M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242. [CrossRef]
52. Gamon, J.A.; Peñuelas, J.; Field, C.B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* **1992**, *41*, 35–44. [CrossRef]
53. 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. [CrossRef]
54. Blackburn, G.A. Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sens. Environ.* **1998**, *66*, 273–285. [CrossRef]
55. Chappelle, E.W.; Kim, M.S.; McMurtrey, J.E. Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves. *Remote Sens. Environ.* **1992**, *39*, 239–247. [CrossRef]
56. Rouse, J.W.; Hass, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the great plains with ERTS. In Proceedings of the Third Earth Resources Technology Satellite Symposium, Washington, DC, USA, 10–14 December 1973; pp. 309–317.
57. Gamon, J.A.; Surfus, J.S. Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* **1999**, *143*, 105–117. [CrossRef]
58. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [CrossRef]
59. Penuelas, J.; Baret, F.; Filella, I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica* **1995**, *31*, 221–230.
60. Nagler, P.L.; Daughtry, C.S.T.; Goward, S.N. Plant litter and soil reflectance. *Remote Sens. Environ.* **2000**, *71*, 207–215. [CrossRef]
61. Roujean, J.L.; Breon, F.M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [CrossRef]
62. Serrano, L.; Peñuelas, J.; Ustin, S.L. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sens. Environ.* **2002**, *81*, 355–364. [CrossRef]
63. Vincini, M.; Frazzi, E.; D'Alessio, P. Angular Dependence of Maize and Sugar Beet VIs from Directional CHRIS/Proba Data. Available online: https://www.researchgate.net/profile/Ermes_Frazzi/publication/228413259_Angular_dependence_of_maize_and_sugar_beet_VIs_from_directional_CHRISProba_data/links/0046352d50c18b3fe6000000.pdf (accessed on 3 January 2018).
64. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* **2002**, *81*, 416–426. [CrossRef]
65. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Source Ecol.* **1969**, *50*, 663–666. [CrossRef]
66. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [CrossRef]
67. Hunt, E.R.; Rock, B.N. Detection of changes in leaf water content using Near- and Middle-Infrared reflectances. *Remote Sens. Environ.* **1989**, *30*, 43–54. [CrossRef]
68. Stamatiadis, S.; Taskos, D.; Tsadilas, C.; Christofides, C.; Tsadila, E.; Schepers, J.S. Relation of ground-sensor canopy reflectance to biomass production and grape color in two merlot vineyards. *Am. J. Enol. Vitic.* **2006**, *57*, 415–422. [CrossRef]
69. Hardisky, M.A.; Klemas, V.; Smart, R.M. The influence of Soil Salinity, Growth Form, and Leaf Moisture on the Spectral Radiance of *Spartina alterniflora* Canopies. *Photogramm. Eng. Remote Sens.* **1983**, *49*, 77–83.
70. Gitelson, A.A.; Merzlyak, M.N.; Chivkunova, O.B. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochem. Photobiol.* **2001**, *74*, 38–45. [CrossRef]

71. Datt, B.; McVicar, T.R.; Van Niel, T.G.; Jupp, D.L.B.; Pearlman, J.S. Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1246–1259. [[CrossRef](#)]
72. Fensholt, R.; Sandholt, I. Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sens. Environ.* **2003**, *87*, 111–121. [[CrossRef](#)]
73. Zarco-Tejada, P.J.; Rueda, C.A.; Ustin, S.L. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. *Remote Sens. Environ.* **2003**, *85*, 109–124. [[CrossRef](#)]
74. Daughtry, C.S.T.; Walthall, C.L.; Kim, M.S.; De Colstoun, E.B.; McMurtrey, J.E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. [[CrossRef](#)]
75. Penuelas, J.; Pinol, J.; Ogaya, R.; Filella, I. Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). *Int. J. Remote Sens.* **1997**, *18*, 2869–2875. [[CrossRef](#)]
76. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y.U. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [[CrossRef](#)]
77. Merton, R.; Huntington, J. Early Simulation Results of the Aries-1 Satellite Sensor for Multi-Temporal Vegetation Research Derived from Aviris. In Proceedings of the Eighth Annual JPL Airborne Earth Science Workshop, Pasadena, CA, USA, 9–11 February 1999; pp. 1–10.
78. Matsushita, Y.; Lin, S. Radiometric calibration from noise distributions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007. [[CrossRef](#)]
79. Tsin, Y.; Ramesh, V.; Kanade, T. Statistical calibration of CCD imaging process. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; pp. 480–487. [[CrossRef](#)]
80. You, H.; Ma, Z.; Tang, Y.; Wang, Y.; Yan, J.; Ni, M.; Cen, K.; Huang, Q. Comparison of ANN (MLP), ANFIS, SVM, and RF models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators. *Waste Manag.* **2016**. [[CrossRef](#)] [[PubMed](#)]
81. Yuan, H.; Yang, G.; Li, C.; Wang, Y.; Liu, J.; Yu, H.; Feng, H.; Xu, B.; Zhao, X.; Yang, X. Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of RF, ANN, and SVM regression models. *Remote Sens.* **2017**, *9*. [[CrossRef](#)]
82. Cochran, W.G. *Sampling Technique*; China Statistical Publishing House: Beijing, China, 1985.
83. Volpe, V.; Manzoni, S.; Marani, M.; Katul, G. *Leave-One-Out Cross-Validation*; Springer: Berlin, Germany, 2011.
84. Humbeck, K.; Quast, S.; Krupinska, K. Functional and molecular changes in the photosynthetic apparatus during senescence of flag leaves from field-grown barley plants. *Plant Cell Environ.* **1996**, *19*, 337–344. [[CrossRef](#)]
85. Martínez-Muñoz, G.; Suárez, A. Out-of-bag estimation of the optimal sample size in bagging. *Pattern Recognit.* **2010**, *43*, 143–152. [[CrossRef](#)]
86. Garg, A.; Kang, T.; Tai, K. Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *Int. J. Model. Identif. Control* **2013**, *18*, 295–312. [[CrossRef](#)]
87. Abdullah, S.; Ismail, M.; Fong, S.Y.; Ahmed, A.M.A.N. Evaluation for long term PM₁₀ concentration forecasting using multi linear regression (MLR) and principal component regression (PCR) models. *Environ. Asia* **2016**, *9*, 101–110.
88. Jin, X.; Ma, J.; Wen, Z.; Song, K. Estimation of maize residue cover using Landsat-8 OLI image spectral information and textural features. *Remote Sens.* **2015**, *7*, 14559–14575. [[CrossRef](#)]
89. Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340–1347. [[CrossRef](#)] [[PubMed](#)]
90. Deng, H.; Runger, G.; Tuv, E. Bias of importance measures for multi-valued attributes and solutions. In Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN 2011), Espoo, Finland, 14–17 June 2011; pp. 293–300.
91. Zhao, W.; Hopke, P.K.; Qin, X.; Prather, K.A. Predicting bulk ambient aerosol compositions from ATOFMS data with ART-2a and multivariate analysis. *Anal. Chim. Acta* **2005**, *549*, 179–187. [[CrossRef](#)]
92. Farifteh, J.; Van der Meer, F.; Atzberger, C.; Carranza, E.J.M. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sens. Environ.* **2007**, *110*, 59–78. [[CrossRef](#)]

