

Chemical Composition and Calorific Value Prediction of Wheat Straw at Different Maturity Stages Using Near-Infrared Reflectance Spectroscopy

Wenjuan Niu, Guangqun Huang, Xian Liu, Longjian Chen, and Lujia Han^{*,†}

Biomass Resources and Utilization Laboratory, College of Engineering, China Agricultural University, Beijing 100083, China

Supporting Information

ABSTRACT: Rapid assessment of wheat straw at different maturity stages will help to reveal its growth mature and enable better process control that will optimize the sustainable value-added usage. This study explored the potential of near-infrared reflectance (NIR) spectroscopy to quantitatively and qualitatively analyze the multiple chemical composition and calorific value of wheat straw at different maturity stages. Partial least-squares (PLS) and genetic algorithm and partial least-squares (GA-PLS) models were used for NIR spectroscopy analysis. Results showed that PLS models and GA-PLS models could be both used for the estimation of chemical composition and calorific value of wheat straw at three maturity stages, and the GA-PLS method reduced the spectral variables for modeling and provided sensitive spectral variables correlated well with chemical composition and calorific value. NIR spectroscopy could successfully detect the contents of water-soluble carbohydrates (WSC), crude protein (CP), acid detergent fiber (ADF), neutral detergent fiber (NDF), ash, and nitrogen (N). It was also able to quantify dry matter (DM), cellulose (Cel), moisture (Moist), volatile matter (VM), higher heating value (HHV), and lower heating value (LHV). The NIR models for hemicellulose (Hem), lignin (Lig), carbon (C), sulfur (S), and hydrogen (H) had moderate accuracy, and could be used for qualitative analysis, whereas for fixed carbon (FC), it was suitable for screening purposes.

1. INTRODUCTION

Crop straw is an important resource in agricultural production systems. In 2011, its production in China was 842 million metric tons, including 127 million metric tons of wheat straw (National Bureau of Statistics, China).¹ Variations in the chemical composition and calorific value of wheat straw at different maturity stages play important roles in lodging resistance, energy properties, and biofuel utilization.^{1,2} Dry matter (DM), water-soluble carbohydrates (WSC), crude protein (CP), cellulose (Cel), hemicellulose (Hem), lignin (Lig), acid detergent fiber (ADF), and neutral detergent fiber (NDF) are related to the biomechanical properties, energy conversion efficiency, and nutrient characteristics of wheat straw.^{2,3} Nitrogen (N), carbon (C), sulfur (S), and hydrogen (H) are important components of organic matter and can be applied to estimate the energy value.³ The energy properties of wheat straw are related to moisture (Moist), ash (Ash), volatile matter (VM), fixed carbon (FC), higher heating value (HHV), and lower heating value (LHV).⁴ Analyses of the chemical composition and calorific value of wheat straw at different maturity stages can reveal the inherent nature of the material and explain related mechanisms in the growth process. This information is useful for breeding programs and also lays a foundation for efficiently evaluating and using wheat straw.

Conventional wet chemical analysis methods are usually accurate and credible but are also time-consuming, expensive, labor-intensive, and usually destructive.⁵ Thus, a rapid and cost-effective method is alternatively needed. Near infrared reflectance (NIR) spectroscopy is a kind of rapid, inexpensive, and nondestructive technology analysis method,^{6,7} and it has been widely used in agriculture, food, and industry.⁸ Its principle is based on the correlation between chemical composition and

absorption of electromagnetic radiation at different wavelength. The most common absorption bands that occur in the NIR regions are overtones and combinations bands of C—H, O—H, C=O, N—H, and aromatic C—H groups. NIR spectroscopy could be used to estimate the chemical composition and calorific value of crop straw with rich of C, H, O, and N.^{6,9}

NIR spectra typically exhibit broad ill-defined overlapping bands of chemical information, and chemometrics is usually needed to extract relevant signals and minimize irrelevant signals from the spectra information during analysis. Partial least-squares (PLS) regression is a bilinear calibration method that uses all available spectral variables, and it has been used to process NIR signal data.^{4,6} The genetic algorithm (GA) is a chemometrics method for selecting spectral variables based on Darwinian evolution and Mendelism, and natural selection is algorithmically mimicked in the strategy.¹⁰ GA employs a probabilistic, nonlocal search process to select variables from whole variable space and it can rapidly find good solutions for difficult high-dimensional problems. The selection of spectral variables by GA can refine the predictive efficiency of PLS models.¹⁰

In previous studies, NIR spectroscopy based on PLS method has been used to predict CP, Cel, Lig, Hem, ADF, NDF, C, H, N, S, Moist, VM, FC, Ash, and HHV in silage,¹¹ crop straw,^{12–14} cereal cultivars,¹⁵ and wood.¹⁶ The purpose of this study was to explore the feasibility and capacity of using NIR spectroscopy technology based on PLS and GA-PLS methods to determine

Received: June 29, 2014

Revised: November 12, 2014

Published: November 13, 2014



Table 1. Descriptive Statistics of Actual Values for Chemical Composition and Calorific Value of Wheat Straw at Different Maturity Stages^a

chemical composition and calorific value ^b	milky stage	dough stage	fully ripe stage	P ^c	all stages	CV ^d of all stages
DM (%), fb)	36.22 ± 3.28	71.14 ± 20.28	95.26 ± 1.96	<0.05	67.95 ± 25.98	38.23
WSC (%), db)	11.23 ± 4.17	3.11 ± 1.49	2.49 ± 1.45	<0.05	5.06 ± 4.72	93.28
CP (%), db)	7.87 ± 2.36	4.28 ± 1.23	3.26 ± 0.69	<0.05	5.00 ± 2.38	47.63
Cel (%), db)	28.91 ± 2.89	36.84 ± 3.97	39.30 ± 2.40	<0.05	35.29 ± 5.23	14.83
Hem (%), db)	19.37 ± 2.21	21.51 ± 2.53	23.47 ± 3.52	<0.05	21.42 ± 3.11	14.53
Lig (%), db)	19.37 ± 1.83	20.87 ± 2.83	19.98 ± 2.68	<0.05	20.22 ± 2.60	12.85
ADF (%), db)	35.61 ± 2.78	46.74 ± 3.75	49.96 ± 3.19	<0.05	44.61 ± 6.62	14.83
NDF (%), db)	54.98 ± 3.85	68.39 ± 4.57	73.43 ± 4.48	<0.05	66.03 ± 8.32	12.60
N (%), db)	1.39 ± 0.39	0.76 ± 0.22	0.57 ± 0.11	<0.05	0.88 ± 0.41	46.46
C (%), db)	42.88 ± 1.29	41.95 ± 1.21	43.00 ± 0.86	<0.05	42.47 ± 1.25	2.95
S (%), db)	0.46 ± 0.11	0.43 ± 0.12	0.32 ± 0.07	<0.05	0.41 ± 0.12	28.75
H (%), db)	5.96 ± 0.49	5.66 ± 0.34	5.39 ± 0.46	<0.05	5.67 ± 0.46	8.18
Moist (%), fb)	63.78 ± 3.28	28.86 ± 20.28	4.74 ± 1.96	<0.05	32.05 ± 25.98	81.06
Ash (%), db)	9.49 ± 1.65	10.89 ± 1.97	8.98 ± 2.06	<0.05	10.02 ± 2.08	20.72
VM (%), db)	69.60 ± 1.55	69.35 ± 1.94	70.94 ± 2.34	<0.05	69.84 ± 2.06	2.94
FC (%), db)	20.91 ± 1.71	19.76 ± 1.44	20.08 ± 1.37	<0.05	20.16 ± 1.57	7.77
HHV (MJ/Kg, db)	16.59 ± 0.91	16.43 ± 0.65	16.68 ± 0.47	<0.05	16.53 ± 0.70	4.24
LHV (MJ/Kg, db)	15.22 ± 0.58	15.21 ± 0.64	15.52 ± 0.44	<0.05	15.29 ± 0.59	4.02

^aIn mean ± standard deviation; n = 109. ^bfb, fresh basis; db, dry basis; DM, dry matter; WSC, water-soluble carbohydrates; CP, crude protein; Cel, cellulose; Hem, hemicellulose; Lig, lignin; ADF, acid detergent fiber; NDF, natural detergent fiber; Moist, moisture; VM, volatile matter; FC, fixed carbon; HHV, higher heating value; LHV, lower heating value. ^cSignificant differences between different maturity stages (p < 0.05). ^dCV, coefficient of variation value was the ratio of the mean to the standard deviation.

Table 2. Pearson Correlation Coefficient between Pairs of Parameters Based On Actual Values of Chemical Composition and Calorific Value of Wheat Straw at Different Maturity Stages^a

chemical composition and calorific value ^b	DM (%), fb)	WSC (%), db)	CP (%), db)	Cel (%), db)	Hem (%), db)	Lig (%), db)	ADF (%), db)	NDF (%), db)	N (%), db)	C (%), db)
DM (%), fb)	1									
WSC (%), db)	-0.70 ^c	1								
CP (%), db)	-0.67 ^c	0.51 ^c	1							
Cel (%), db)	0.74 ^c	-0.75 ^c	-0.74 ^c	1						
Hem (%), db)	0.40 ^c	-0.46 ^c	-0.40 ^c	0.50 ^c	1					
Lig (%), db)	-0.03	-0.22 ^c	0.16	0.31 ^c	0.11	1				
ADF (%), db)	0.80 ^c	-0.80 ^c	-0.81 ^c	0.91 ^c	0.38 ^c	0.17	1			
NDF (%), db)	0.77 ^c	-0.81 ^c	-0.78 ^c	0.90 ^c	0.68 ^c	0.20 ^d	0.94 ^c	1		
N (%), db)	-0.67 ^c	0.56 ^c	0.97 ^c	-0.75 ^c	-0.45 ^c	0.20 ^d	-0.81 ^c	-0.81 ^c	1	
C (%), db)	-0.05	0.01	0.19 ^d	0.02	0.28 ^c	0.07	-0.07	0.07	0.15	1
S (%), db)	-0.40 ^c	0.25 ^c	0.43 ^c	-0.46 ^c	-0.24 ^d	-0.07	-0.46 ^c	-0.45 ^c	0.40 ^c	-0.05
H (%), db)	-0.45 ^c	0.27 ^c	0.46 ^c	-0.49 ^c	-0.07	0.00	-0.35 ^c	-0.34 ^c	0.65 ^c	0.44 ^c
Moist (%), fb)	0.96 ^c	0.73 ^c	0.67 ^c	-0.77 ^c	-0.43 ^c	0.07	-0.82 ^c	-0.81 ^c	0.67 ^c	0.05
Ash (%), db)	-0.10	-0.08	0.01	-0.13	-0.33 ^c	0.04	-0.05	-0.16	0.04	-0.80 ^c
VM (%), db)	0.29 ^c	-0.11	-0.27 ^c	0.38 ^c	0.42 ^c	-0.02	0.31 ^c	0.40 ^c	-0.25 ^d	0.50 ^c
FC (%), db)	-0.24 ^d	0.24 ^d	0.34 ^c	-0.33 ^c	-0.12	-0.04	-0.34 ^c	-0.31 ^c	0.28 ^c	0.42 ^c
HHV (MJ/kg, db)	0.04	-0.07	0.09	0.23 ^d	0.28 ^c	0.11	0.05	0.14	0.06	0.61 ^c
LHV (MJ/kg, db)	0.04	-0.07	0.09	0.23 ^d	0.29 ^c	0.14	0.05	0.14	0.06	0.62 ^c

^an = 109. ^bfb, fresh basis; db, dry basis; DM, dry matter; WSC, water-soluble carbohydrates; CP, crude protein; Cel, cellulose; Hem, hemicellulose; Lig, lignin; ADF, acid detergent fiber; NDF, natural detergent fiber; Moist, moisture; VM, volatile matter; FC, fixed carbon; HHV, higher heating value; LHV, lower heating value. ^cCorrelation is significant at the 0.01 level (2-tailed). ^dCorrelation is significant at the 0.05 level (2-tailed).

the multiple chemical composition and calorific value of wheat straw at different maturity stages.

2. MATERIALS AND METHODS

2.1. Samples Collection and Preparation. In total, 109 wheat straw samples of 34 cultivars were collected in Beijing, Tianjin, Shandong, Hebei, and Shanxi provinces of China during 2011 to 2013. The 109 wheat straw samples were grown in different soil and climatic conditions and were harvested at different times. To study the chemical composition and calorific

value of wheat straw at different maturity stages, the samples were collected at the milky stage, dough stage, and fully ripe stage.¹⁷ There were 29 milky-stage samples, 47 dough-stage samples, and 33 fully ripe-stage samples (see Table S1, Supporting Information). Each wheat straw sample was taken from a different position in the field and was thoroughly mixed to obtain a representative batch of approximately 2 kg. A little part of fresh wheat straw was used for the measurement of Moist and DM, the other part of wheat straw samples were oven-dried at 45 °C for 48 h and milled to pass through a

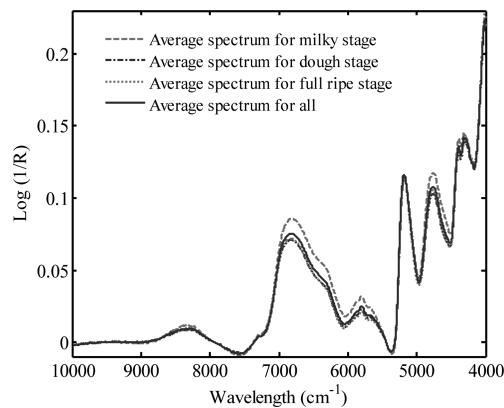


Figure 1. Raw spectra of wheat straw samples at different maturity stages.

0.90 mm mesh sieve. Then, the dried samples were used for NIR spectra scanning and other wet chemical analyses.

2.2. Reference Analysis of Chemical Composition and Calorific Value. Moist and DM were analyzed at 45 °C using fresh samples, and the other chemical composition and calorific value analyses of wheat straw were conducted using wheat straw samples dried at 45 °C according to ASTM E1757-01.¹⁸ The WSC content was determined by measuring absorbance at 620 nm using a UV-2550 spectrophotometer, according to

anthrone sulfuric acid method.¹⁹ CP was determined using the official method.²⁰ Cel, Hem, and Lig contents were measured according to the National Renewable Energy Laboratory (NREL) method.²¹ ADF and NDF contents were measured using a Fibertec 2010 extractor (FOSS, Höganäs, Sweden) according to the method of Van Soest.²² The C, H, N, and S contents were measured using an Elementar Vario EL II Analyzer (Vario Macro, Hanau, Germany). Ash content was measured at 550 °C by burning samples to a constant weight in a muffle furnace. VM was analyzed by burning samples for 7 min at 900 °C in a muffle furnace, and FC was calculated by subtracting Ash and VM from 100%. Parr 6300 Oxygen Bomb Calorimeter (Parr 6300, America) was used to determine HHV and LHV values. The chemical analyses were performed in triplicate aliquots, and the recorded result was the average.

2.3. NIR Spectra Collection. Spectra were measured in reflection using a Fourier Transform-NIR spectrometer system (Spectrum 400; PerkinElmer, USA) equipped with a high speed, low noise, indium gallium arsenide detector. For each sample, approximately 15 g of wheat straw was placed in a standard PerkinElmer sample cup, which had a quartz window of 10 cm in diameter and 3.5 cm in depth. The spectra were collected from 4000 to 10 000 cm⁻¹ at a resolution of 8 cm⁻¹. Each spectrum was the average of 32 scans. Each sample was analyzed three times, and the three spectra were averaged for further analysis. The spectrometer is sensitive to changes in

Table 3. Overtones and Combination Bands for the Functional Groups in Wavelengths of 10 000–4000 Cm⁻¹

wavelength (cm ⁻¹)	overtones and combination bands ²⁷	functional groups or structure ²⁷	possible related chemical composition ^a
8880, 8258, 8231	C—H stretching second overtone	·CH ₂ in hydrocarbon	C, H
7300	Combination band of C—H stretching and C—H bending	ArCH ₃ in hydrocarbon	C, H
7168, 7067, 6944	Combination band of C—H stretching and C—H deformation	·CH ₂ in hydrocarbon	C, H
7042	O—H stretching first overtone	ArOH	H
6852	N—H stretching first overtone	p-NH ₂ in aromatic amine	CP, N
6835, 6798, 6743, 6711	N—H stretching first overtone	·CONH ₂ in protein	CP, C, N
6494	O—H stretching first overtone	·O—H in polyol	H
6330	O—H stretching first overtone	alcohol or water	Moist, O, H
5920, 5900, 5865	C—H stretching first overtone	·CH ₃ in methyl	C, H
5797	C—H stretching first overtone	·CH ₂ in submethyl	C, H
5618	C—H stretching first overtone	·CH ₂ in cellulose	Cel, C, H
5495	combination band of O—H stretching and C—O stretching	O—H/C—O in cellulose	Cel, C, H
5208	combination band of C=O stretching and N—H stretching	·C=ONH in amide	CP, C, N, H
5186, 5181	combination band of O—H stretching and HOH deformation	·O—H in polysaccharide	WSC, Cel, Hem, Lig, ADF, NDF, H
5100	combination band of N—H asymmetric stretching and amide Π deformation	R—C=O—NH ₂ in primary amide	CP, C, N, H
5095	combination band of N—H stretching and N—H bending	N—H in aromatic amine	CP, N, H
4762	combination band of O—H stretching and C—O stretching	O—H/C—O in polysaccharide	WSC, Cel, Hem, Lig, ADF, NDF, C, H
4701, 4690, 4660	combination band of N—H stretching and C=O stretching	·CONH ₂ in amino acid	CP, C, N, H
4587	N—H deformation second overtone	·N—H in amino acid	CP, N, H
4460	combination band of N—H stretching and NH ₃ deformation	·N—H in amino acid	CP, N, H
4747	S—H stretching first overtone	·S—H in mercaptan	S
4420	combination band they are sum of several fundamental vibrations from different vibrations	·C—H in cellulose	Cel, C, H
4405, 4400	combination band of O—H stretching and C—O stretching	O—H/C—H in cellulose or glucose	WSC, Cel, C, H, O
4386, 4307, 4292, 4283	combination band of C—H stretching and CH ₂ deformation	·C—H and ·CH ₂ in polysaccharide	WSC, Cel, Hem, Lig, ADF, NDF, C, H
4063, 4019	combination band of C—H stretching and C—C stretching	C—H/C—C in cellulose	Cel

^a WSC, water-soluble carbohydrates; CP, crude protein; Cel, cellulose; Hem, hemicellulose; Lig, lignin; ADF, acid detergent fiber; NDF, natural detergent fiber; Moist, moisture.

Table 4. NIR Calibration, Cross-Validation, and Validation for Chemical Composition and Calorific Value of Wheat Straw at Different Maturity Stages Using PLS Method with Full Spectral Variables^a

chemical composition and calorific value ^b	pretreatment ^c	Factors	calibration and cross-validation ^d (n = 73)				validation ^e (n = 36)			
			R _c ²	RMSEC	R _{cv} ²	RMSECV	R _v ²	RMSEP	RPD	SD _a
DM (%), fb)	derivative (15, 2, 1) +normalize	6	0.91	8.15	0.82	10.96	0.87	10.32	2.53	1.03
WSC (%), db)	derivative (15, 2, 1) +normalize	6	0.91	1.28	0.90	1.30	0.90	1.12	5.13	0.24
CP (%), db)	detrend	8	0.98	0.40	0.96	0.44	0.96	0.47	6.17	0.12
Cel (%), db)	none	7	0.88	1.81	0.87	1.92	0.86	2.16	2.49	0.94
Hem (%), db)	none	8	0.76	1.53	0.70	1.66	0.72	1.56	2.00	0.92
Lig (%), db)	derivative (15, 2, 1) +baseline	4	0.72	1.38	0.65	1.46	0.58	1.60	1.62	0.72
ADF (%), db)	derivative (7, 2, 1) +MSC	4	0.93	1.75	0.93	1.76	0.93	1.68	3.84	0.76
NDF (%), db)	derivative (7, 2, 1) +MSC	5	0.95	1.91	0.93	2.19	0.94	2.21	3.76	0.80
N (%), db)	detrend	8	0.95	0.08	0.96	0.07	0.96	0.07	7.27	0.02
C (%), db)	MSC	6	0.73	0.56	0.66	0.63	0.72	0.65	2.16	0.33
S (%), db)	derivative (15, 2, 1) +MSC	3	0.66	0.05	0.57	0.06	0.54	0.06	1.93	0.03
H (%), db)	MSC	7	0.69	0.19	0.62	0.20	0.64	0.21	1.98	0.09
Moist (%), fb)	derivative (15, 2, 1) +normalize	6	0.91	8.15	0.82	10.96	0.87	10.32	2.53	1.03
Ash (%), db)	MSC	9	0.92	0.55	0.91	0.61	0.91	0.63	3.13	0.14
VM (%), db)	MSC+smoothing	8	0.81	0.94	0.80	0.96	0.84	0.91	2.26	0.37
FC (%), db)	MSC+normalize	6	0.45	1.01	0.40	1.14	0.46	1.02	1.82	0.77
HHV (MJ/kg), db)	derivative (15, 2, 1) +MSC	7	0.87	0.21	0.88	0.20	0.88	0.20	2.68	0.21
LHV (MJ/kg), db)	derivative (15, 2, 1) +MSC	7	0.88	0.22	0.89	0.22	0.88	0.19	2.79	0.22

^an = 109. ^bfb, fresh basis; db, dry basis; DM, dry matter; WSC, water-soluble carbohydrates; CP, crude protein; Cel, cellulose; Hem, hemicellulose; Lig, lignin; ADF, acid detergent fiber; NDF, natural detergent fiber; Moist, moisture; VM, volatile matter; FC, fixed carbon; HHV, higher heating value; LHV, lower heating value. ^cMSC, multiplicative scatter correction. ^dR_c², the coefficient of determination in the calibration set; RMSEC, the root-mean-square error of calibration; R_{cv}², the coefficient of determination in cross-validation for calibration; RMSECV, the root-mean-square error of cross-validation in calibration set. ^eR_v², the coefficient of determination in the validation set; RMSEP, the root-mean-square error of prediction; RPD, the ratio of standard error of performance to standard deviation for validation; SD_a, the average of standard deviation of the three parallel samples.

environmental conditions (temperature and humidity), so the room temperature was maintained at 25 °C and the humidity was maintained at a constant ambient level in the laboratory.^{7,23}

2.4. Data Processing and Analysis. SPSS 17.0 and Matlab R2012b software were used to analyze chemical measurements and spectral data. One-way analysis of variance (ANOVA), multifactorial ANOVA, and the Pearson correlation coefficient analyses were conducted using SPSS 17.0. PLS and GA packages were used in Matlab R2012b.

NIR spectroscopy models were performed by PLS based on the full spectral variables and selected variables by GA. All samples were arranged from the lowest to the highest by the actual concentration gradient of chemical composition and calorific value; every third sample was selected for validation and the remainder was kept for calibration. The calibration set consisted of 73 samples, and the validation set consisted of 36 samples (Table S2 of Supporting Information). In calibration, leave-one-out cross validation was used to select the optimal number of factors and avoid overfitting. In the process of cross validation, one variable was removed out of the spectral variables and the maximum number of factors was set as 20. Different preprocessing methods such as smoothing, baseline, derivative, detrend, normalize, multiplicative scatter correction (MSC), and autoscale were used to pretreat the data. Smoothing filtered out noise interference. Baseline eliminated baseline drift. Derivative (15, 2, 1) and derivative (7, 2, 1) could eliminate baseline drift or flat background interference; the first number indicates filter width, the second number indicates polynomial order, and the third number indicates the order of the derivative, respectively. Detrend removed the linear and quadratic curvature of each spectrum. Normalize and MSC

functions were used for normalization. Autoscale was used for scaling and centering.

In addition, GA analysis was used to select spectral variables. The spectra of the calibration set were used for variable selection by GA. The key parameters in the GA analyses were set as follows: 30 for the initial population size, 50% for the crossover rate, 1% for the probability of mutation, and 100 for the number of generations. GA analysis was run five times to eliminate the stochastic effect of the algorithm. Sensitive spectral variables of each run were selected when the coefficient of variation (CV) used to explain the variance as a function of the number of variables included reached a maximum. Finally, the total spectral variables selected after five runs were used for the PLS analysis, and the spectral variables selected by GA that have significant correlations with chemical composition and calorific value were defined as sensitive spectral variables.

2.5. Model Evaluation. The accuracy of PLS and GA-PLS models was evaluated by the statistical parameters both in calibration and validation set. Values for the coefficient of determination in the calibration set (R_c²), the root-mean-square error of calibration (RMSEC), the coefficient of determination in cross-validation (R_{cv}²), and the root-mean-square error of cross-validation (RMSECV) were derived for calibration. Values for the coefficient of determination in the validation set (R_v²), the root-mean-square error of prediction (RMSEP), and the ratio of standard error of performance to standard deviation (RPD) were derived for validation.²⁴ All of the above indicators were based on the average of triplicate analyses of each sample. The optimal number of PLS factors was determined by the highest values of R_c², R_{cv}² and the lowest values of RMSEC, RMSECV. The standard deviation (SD) of the three parallel samples for each sample in the validation set was also

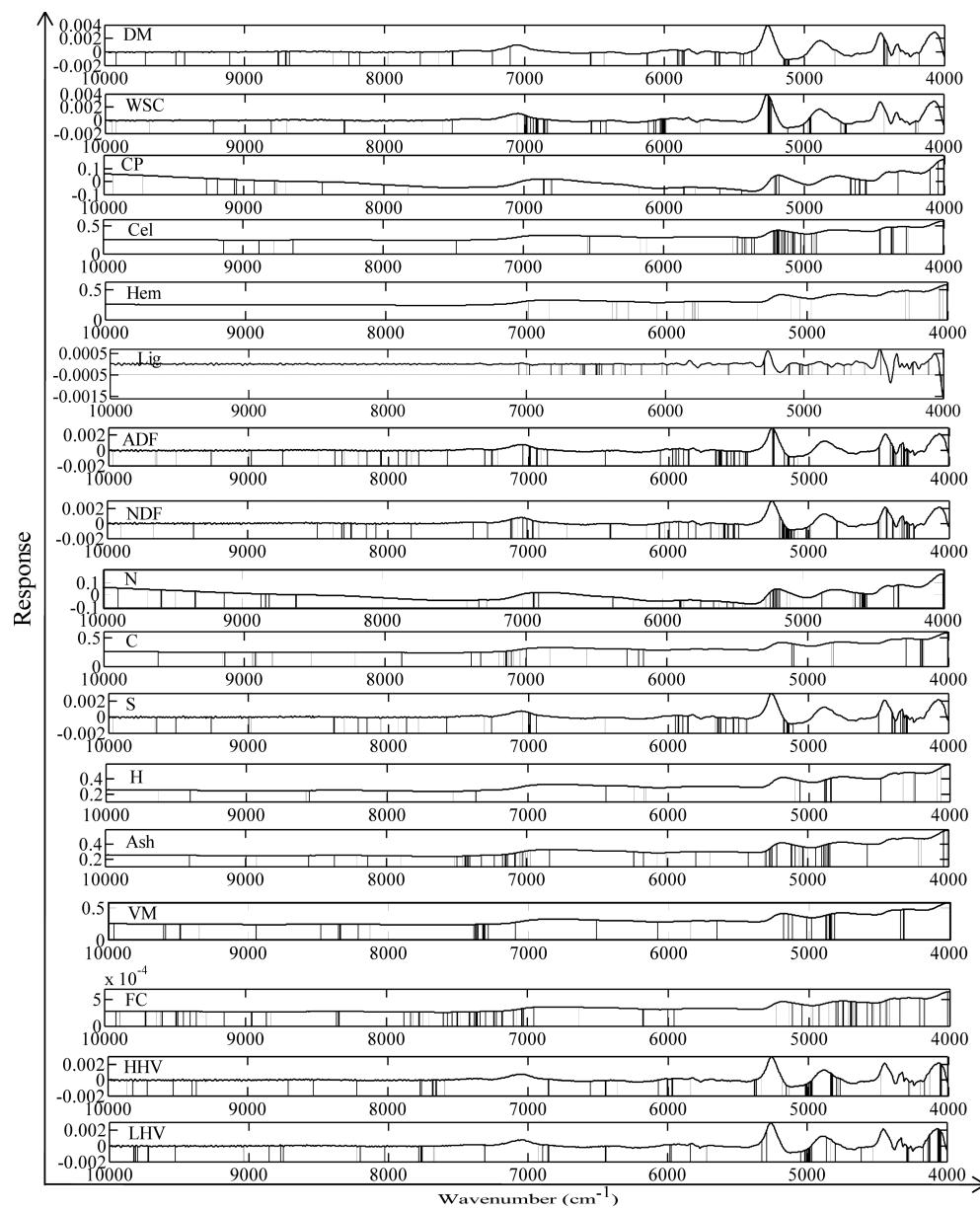


Figure 2. Sensitive spectral variables selected by GA analysis and different pretreated NIR spectra for the prediction of chemical composition and calorific value.

calculated to test the precision of the model. The SD_a value was the average of SD of 36 samples in the validation set.

A good model was chosen based on higher R_c^2 , R_{cv}^2 , and R_p^2 values and lower RMSEC, RMSECV, RMSEP, and SD_a values. The performance of the NIR model was evaluated by R_v^2 and RPD as described by Malley.²⁴ The values of R_v^2 and RPD were related to the predictive strength of the model, as follows: $R_v^2 > 0.95$ and $RPD > 4.00$, excellent prediction; $0.90 < R_v^2 < 0.95$ and $3.00 < RPD < 4.00$, successful prediction; $0.80 < R_v^2 < 0.90$, $2.25 < RPD < 3.00$, moderately successful prediction; $0.70 < R_v^2 < 0.80$, $1.75 < RPD < 2.25$, moderately useful prediction; $R_v^2 < 0.70$, $RPD < 1.75$, limited for screening purpose.²⁵

3. RESULTS AND DISCUSSION

3.1. Chemical Composition and Calorific Value. The results of one-way ANOVA and multifactorial ANOVA analyses are shown in Table S3 of Supporting Information. The one-way ANOVA showed that there were significant differences ($P < 0.05$)

in chemical composition and calorific value of wheat straw among the different stages of maturity. However, there was no significant difference ($P > 0.05$) in chemical composition and calorific value among the different provinces. The differences in DM, Hem, Lig, S, H, Moist, FC, HHV, and LHV among different cultivars were not significant ($P > 0.05$). The multifactorial ANOVA showed that the differences in chemical composition and calorific value among different stages of maturity and different cultivars were not significant ($P > 0.05$). Therefore, maturity stage had the strongest effect on the chemical composition and calorific value of wheat straw.

Table 1 showed the descriptive statistics of chemical composition values and calorific values of wheat straw at different maturity stages. Milky stage samples had low DM, Cel, Hem, Lig, ADF, and NDF contents, and higher WSC, CP, N, and H contents. Milky-stage and dough-stage samples had high S contents and low VM, HHV, and LHV contents. Milky-stage and fully ripe-stage samples had high C and low Ash contents.

Table 5. NIR Calibration, Cross-Validation, and Validation for Chemical Composition and Calorific Value of Wheat Straw at Different Maturity Stages Using GA-PLS Method with Selected Spectral Variables^a

chemical composition and calorific value ^b	pretreatment ^c	spectral variables	Factors	calibration and cross-validation ^d (n = 73)				validation ^e (n = 36)			
				R _c ²	RMSEC	R _{cv} ²	RMSECV	R _v ²	RMSEP	RPD	SD _a
DM (% fb)	derivative (15, 2, 1) +normalize	142	2	0.86	9.86	0.83	10.12	0.89	10.13	2.58	0.98
WSC (% db)	derivative (15, 2,1) +normalize	148	2	0.96	1.01	0.95	1.05	0.93	0.96	5.99	0.23
CP (% db)	detrend	84	7	0.98	0.39	0.97	0.42	0.97	0.45	6.44	0.12
Cel (% db)	none	114	4	0.90	1.76	0.90	1.78	0.88	1.85	2.92	0.83
Hem (% db)	none	47	6	0.74	1.54	0.72	1.58	0.72	1.54	2.03	0.87
Lig (% db)	derivative (15, 2, 1) +baseline	108	3	0.81	1.13	0.71	1.40	0.65	1.47	1.76	0.65
ADF (% db)	derivative (7, 2, 1) +MSC	197	2	0.95	1.54	0.95	1.56	0.94	1.59	4.06	0.73
NDF (% db)	derivative (7, 2, 1) +MSC	192	4	0.97	1.50	0.96	1.59	0.95	2.09	3.98	0.79
N (%) db)	detrend	132	6	0.98	0.06	0.97	0.06	0.98	0.06	8.48	0.02
C (%) db)	MSC	73	5	0.75	0.55	0.70	0.60	0.74	0.62	2.26	0.26
S (%) db)	derivative (15, 2, 1) +MSC	142	4	0.78	0.05	0.68	0.06	0.71	0.06	2.32	0.02
H (%) db)	MSC	46	7	0.73	0.18	0.75	0.16	0.79	0.17	2.45	0.07
Moist (% fb)	derivative (15, 2, 1) +Normalize	142	2	0.86	9.86	0.83	10.12	0.89	10.13	2.58	0.98
Ash (% db)	MSC	148	7	0.97	0.36	0.95	0.41	0.93	0.58	3.40	0.15
VM (% db)	MSC + smoothing	100	8	0.83	0.92	0.80	0.95	0.84	0.90	2.29	0.35
FC (% db)	MSC+normalize	192	4	0.52	1.03	0.45	1.10	0.52	1.01	1.84	0.68
HHV (MJ/kg, db)	derivative (15, 2, 1) +MSC	159	7	0.91	0.18	0.90	0.19	0.91	0.18	3.03	0.20
LHV (MJ/kg, db)	derivative (15, 2, 1) +MSC	151	7	0.90	0.18	0.89	0.20	0.89	0.18	3.06	0.21

^an = 109. ^bfb, fresh basis; db, dry basis; DM, dry matter; WSC, water-soluble carbohydrates; CP, crude protein; Cel, cellulose; Hem, hemicellulose; Lig, lignin; ADF, acid detergent fiber; NDF, natural detergent fiber; Moist, moisture; VM, volatile matter; FC, fixed carbon; HHV, higher heating value; LHV, lower heating value. ^cMSC, multiplicative scatter correction. ^dR_c², the coefficient of determination in the calibration set; RMSEC, the root-mean-square error of calibration; R_{cv}², the coefficient of determination in cross-validation for calibration; RMSECV, the root-mean-square error of cross-validation in calibration set. ^eR_v², the coefficient of determination in the validation set; RMSEP, the root-mean-square error of prediction; RPD, the ratio of standard error of performance to standard deviation for validation; SD_a, the average of standard deviation of the three parallel samples.

Milky-stage samples had a high FC content. Across all samples, the coefficients of variation (CV) values were as follows: 93.28% for WSC, 47.63% for CP, 46.46% for N, 20.72% for Ash, and 28.75% for S. The CV values ranged from 12.60% to 14.83% for Cel, Hem, Lig, ADF, and NDF. The CV values for the other components were less than 10%. The DM content of samples ranged from 32.09% to 97.96% and the Moist content ranged from 2.04% to 67.91%. These ranges are wider than those reported in previous studies on straw.¹⁴ The ranges of CP and N were 1.99–14.92% and 0.38–2.36%, respectively. These ranges are also wider than those reported in other studies on straw.¹³

The Pearson correlation coefficients between pairs of parameters, based on actual values of chemical composition and calorific value of wheat straw, are shown in Table 2. The strong correlations may result from similar trends in the changes in the various parameters as the straw matured, which have been explained in our previous research.²⁶ The correlations between HHV, LHV, and the chemical composition of wheat straw were mostly nonsignificant, mainly because HHV and LHV remained stable as the straw matured, whereas other components changed.

3.2. NIR Spectral Analysis. The raw NIR spectra of wheat straw samples at different maturity stages are shown in Figure 1. The overall forms of the spectra were similar among wheat straw samples with different maturity stages. There was little absorption from 10 000 to 9000 cm⁻¹, and most of the absorption peaks were in the 9000–4000 cm⁻¹ region. The spectra of wheat straw at different maturity stages showed a common

phenomenon. Absorption in the 8600–8000, 7300–5500, and 5100–4250 cm⁻¹ regions was higher in milky-stage samples than in samples at the other two maturity stages, and these regions have great variations. The strongest absorption in the 5200–5130 cm⁻¹ and 4200–4000 cm⁻¹ regions was in the fully mature-stage samples. The NIR spectra comprise the overtones and combination bands of the chemical bonds that correlate with the chemical composition and calorific value of the wheat straw samples. The overtones and combination bands for the functional groups in wavelengths of 10 000–4000 cm⁻¹ are shown in Table 3. The spectra of absorption bands in the 8600–8000 cm⁻¹ regions are mostly related to C—H stretching second overtone in hydrocarbons,²⁷ which are possibly related to C and H of wheat straw. Those of bands in the 7300–5500 cm⁻¹ and 5100–4250 cm⁻¹ regions are related to N—H, C—H, O—H, C—O, and C=O overtones and combination bands in hydrocarbons, polysaccharides, amino acid, aromatic amines, and celluloses,²⁷ which are possibly related to WSC, CP, Cel, Hem, Lig, ADF, NDF, C, H, N, and S of wheat straw (Table 3). The spectra in the 5200–5130 and 4200–4000 cm⁻¹ regions correspond to O—H, C—O, C—H, and C—C overtones and combination bands in celluloses and other polysaccharides,²⁷ which are possibly related to WSC, Cel, Hem, Lig, ADF, NDF, C, and H of wheat straw. The bond strength reflected in the spectra of wheat straw samples at three maturity stages was consistent with the changes in chemical composition and calorific value detected in the chemical analyses (Figure 1 and Table 1).

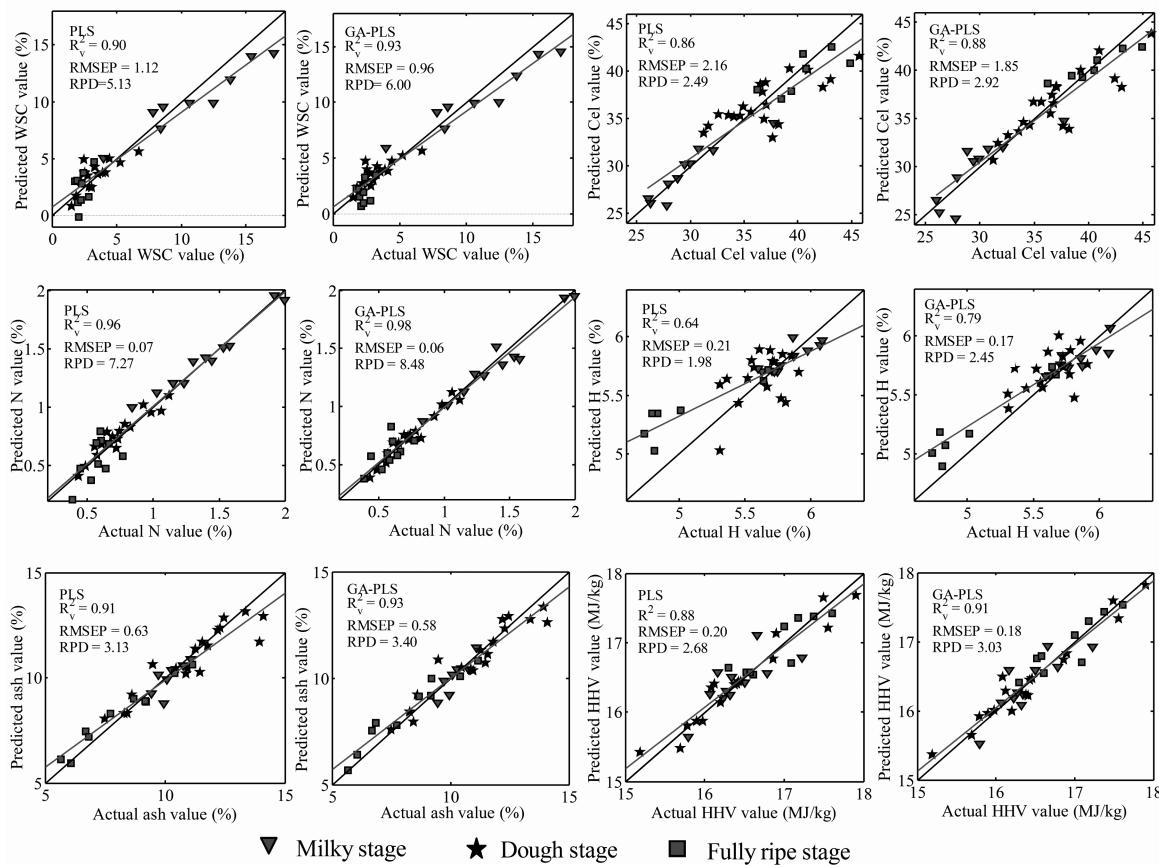


Figure 3. Comparison of scatter plots of predicted versus actual values of chemical composition and calorific value of wheat straw in validation using PLS to GA-PLS method ($n = 36$).

3.3. Comparison of GA-PLS and PLS Models. The PLS models used full spectral variables to predict chemical composition and calorific value of wheat straw at different maturity stages (Table 4). The GA-PLS models were constructed using sensitive spectral variables selected by GA analysis (Figure 2 and Table 5). The number of factors were lower in the GA-PLS models than in the PLS models. Fewer than 200 sensitive spectral variables were selected for the GA-PLS method, while 3,001 spectral variables were selected for the PLS method. The GA-PLS method also provided sensitive spectral variables correlated well with chemical composition and calorific value just shown in Figure 2. GA analyses of DM and Moist, CP and N, ADF and NDF, C and VM, and HHV and LHV selected many of the same spectral variables.

As shown in Tables 4 and 5, compared with those obtained using the PLS method, the R_c^2 , and R_v^2 values obtained using the GA-PLS method were slightly higher or the same. However, compared with PLS models using full spectral regions, the RMSECV, RMSEP, and SD_a values were slightly lower or the same in the GA-PLS models when sensitive spectral variables were selected. To validate the models, some scatter plots of actual values of chemical composition and calorific value of wheat straw versus values predicted using the PLS method and the GA-PLS method were also constructed (Figure 3). These plots reflected not only the accuracy of PLS and GA-PLS models but also the distributions of chemical composition and calorific value of wheat straw at the three different maturity stages. In summary, GA-PLS models can be used for the prediction of the chemical composition and calorific values of wheat

straw at different stages just as PLS models and can also provide sensitive spectral variables correlated well with chemical composition and calorific value.

3.4. Evaluation of NIR Models Using GA-PLS Method.

Good GA-PLS models are indicated by lower RMSEC, RMSECV, RMSEP, and SD_a values, and higher R_c^2 , R_{cv}^2 , R_v^2 , and RPD values. As shown in Table 5, the RMSEC, RMSECV, and RMSEP values for WSC, CP, ADF, NDF, N, and Ash were all lower than 2.09, and the SD_a values were all lower than 0.79, which indicated good accuracy and precision of the models. The R_c^2 , R_{cv}^2 , and R_v^2 values for CP, WSC, ADF, and N were higher than 0.90, and the RPD values were higher than 4.00, indicating excellent NIR models for CP, WSC, ADF, and N. The R_c^2 , R_{cv}^2 , and R_v^2 values for NDF and Ash were higher than 0.90 and the RPD values were between 3.00 and 4.00. The GA-PLS models for NDF and Ash were fully acceptable with good prediction (Table 5). As wheat straw matured from the milky stage to the fully ripe stage, the contents of WSC, CP, and N decreased rapidly while those of ADF and NDF increased rapidly (Figure 3,4). As shown in Table 5, the GA-PLS models for WSC, CP, ADF, NDF, N, and Ash in this study have lower RMSEP values and higher R_v^2 and RPD values than those reported in previous studies on corn stalk silage, straw silage, rice straw, and wheat straw.^{11,13,28–30}

WSC consists of small sugar molecules such as glucose, sucrose, maltose, and soluble hemicelluloses. The C—H, C—O, and O—H bonds of the molecules have strong absorption peaks in the 7100–6800, 6200–5850, 5250–4900, 4750–4600, and 4400–4300 cm⁻¹ regions of the spectra (Figure 2 and Table 3).

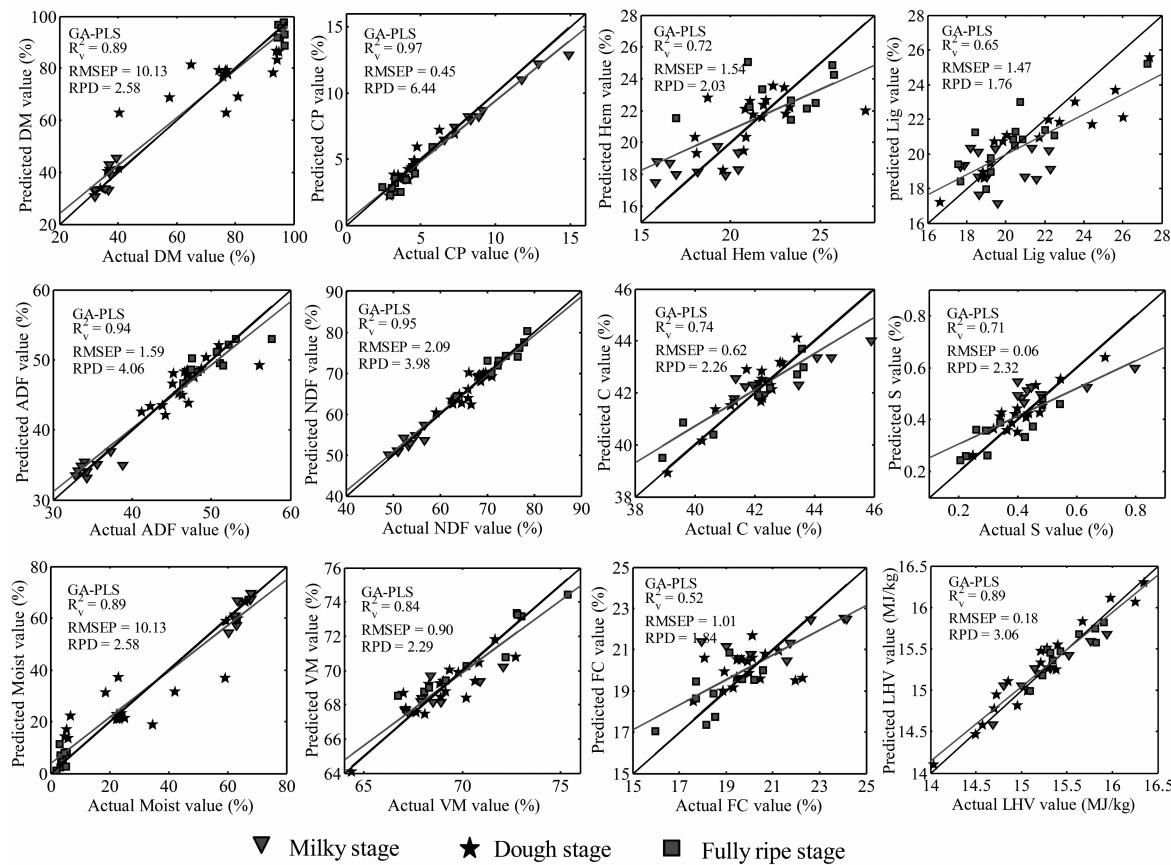


Figure 4. Scatter plots of predicted versus actual values of chemical composition and calorific value of wheat straw in validation using GA-PLS method ($n = 36$).

CP is an important constituent of organic matter, while N is an important element to constitute CP. The N—H bonds have strong absorption peak in the 6930–6700, 5500–5080, and 4750–4450 cm^{-1} regions of the spectra (Figure 2 and Table 3). The ADF and NDF models have good prediction because of the strong response of C—H, O—H, and C—O in the NIR spectra.³¹ NIR spectra also reflect absorption related to inorganic compounds. Ash consists of inorganic compounds including mineral oxides, sulfates, phosphates, and silicates. Those inorganic compounds may cause spectral changes to water or other molecules.

The R_c^2 , R_{cv}^2 , and R_v^2 values for DM, Cel, Moist, VM, HHV, and LHV were higher than 0.80. The RPD values were higher than 2.25. The RMSEP values for Cel, VM, HHV, and LHV were lower than 1.85, and the SD_a values were all lower than 0.83 (Table 5). The RMSEP values for DM and Moist were 10.32, but the SD_a values for DM and Moist were 0.98. These results indicate that the GA-PLS models for DM, Cel, Moist, VM, HHV, and LHV could be used in quantitative analysis with good precision. The Cel and Moist of wheat straw decreased, whereas DM increased as the straw matured (Figures 3 and 4). However, there is little accumulation of organic matter as wheat straw matures. VM is the sum of liquid and gas decomposed when organic matter is burned at 900 °C without air, whereas HHV is the heat released by complete combustion of organic matter. Consequently, VM, HHV, and LHV did not show significant differences among wheat straw at different stages of maturity (Figures 3 and 4).

In this study, the spectra were collected using dry wheat straw samples. It may be possible to predict the DM and Moist

of fresh wheat straw samples based on the strong correlations between actual values of Moist in fresh samples and certain chemical components in dry samples (Table 2). The absorption spectra bands in the 5620–4800, 4420–4237, 4063, and 4019 cm^{-1} regions correspond to Cel (Figure 2 and Table 3). VM, HHV, and LHV are combustion parameters of organic matter. The absorption bands of spectra for organic matter components such as CP, WSC, Cel, Hem, and Lig were strongly correlated well with VM, HHV, and LHV (Table 2).

The R_c^2 , R_{cv}^2 , and R_v^2 values for Hem, C, S, and H were between 0.70 and 0.80, the R_{cv}^2 value for Lig was between 0.70 and 0.80, the RMSEP values were below 1.58, the SD_a values were all less than 0.87, and the RPD values were between 1.75 and 2.25 (Table 5). Based on these values, the GA-PLS models for Hem, Lig, C, S, and H can be used for qualitative analysis. The R_v^2 value for FC was less than 0.70 and the RPD was 1.84 (Table 5). Thus, NIR spectroscopy can only be used for screening FC in wheat straw at different maturity stages. The NIR models for Hem, Lig, S, H, and FC of wheat straw were not very good, possibly because wheat straw does not photosynthesize at maturity, and the absolute contents of C, S, H, and FC did not change significantly (Figures 3 and 4). Alternatively, the poor prediction of S may also be because of the relatively weak absorption peak of S—H bonds of NIR spectra.

In summary, NIR spectroscopy is a rapid and feasible technique to analyze the chemical composition and calorific value of wheat straw at three maturity stages, and it especially performs good function in predicting WSC, CP, ADF, NDF, Ash, N, DM, Cel, Moist, VM, HHV, and LHV. There were significant

differences in accuracy among the NIR models used to predict chemical composition and calorific values (Table 5). DM, WSC, CP, Cel, Hem, ADF, NDF, and Moist were successfully quantified using NIR models. All of these parameters were correlated well with N (Table 2), whereas N can form active N—H and C—N bonds. Ash, VM, FC, HHV, and LHV were successfully or moderately successfully evaluated using NIR models, based on the significant correlations with Cel, Hem, Lig, and C, which contain C—H, C—C, and C=O bonds. The element of C can form organic matrices, and WSC, Cel, Hem, Lig, ADF, and NDF are the main components of organic matter containing C. N is also an element to form organic matter, and CP is organic matter containing N. VM, FC, HHV, and Ash have a close relationship with organic matter. The satisfactory prediction results may be because organic components have strong spectral absorption in the NIR spectra. The accuracy of NIR models may be affected by the contents and concentration ranges of chemical components and calorific value, the number and strength of absorption bands in the spectra, and the correlations with organic matter.

4. CONCLUSION

NIR spectroscopy was a rapid and feasible technique to analyze the multiple chemical composition and calorific value of wheat straw at three maturity stages. GA-PLS models could be used for the prediction of the chemical composition and calorific value of wheat straw at different stages just as PLS models, and GA analysis also provided sensitive spectral variables correlated well with chemical composition and calorific value. The GA-PLS models obtained for WSC, CP, ADF, NDF, Ash, and N had good prediction. The GA-PLS models for DM, Cel, Moist, VM, HHV, and LHV could be used in quantitative analysis. The GA-PLS models for Hem, Lig, C, S, and H showed moderate accuracy and could be applied in qualitative analysis. The GA-PLS model for FC showed poor accuracy and could only be used for screening. It appeared that NIR data processed using GA-PLS models could be used to fulfill different functions in predicting chemical composition and calorific value of wheat straw at different maturity stages, according to the actual situation.

■ ASSOCIATED CONTENT

Supporting Information

Information statistics for wheat straw samples at different maturity stages (Table S1); descriptive statistics for chemical composition and calorific value in calibrations and validations of wheat straw at three maturity stages (Table S2); results of one-way ANOVA and multifactorial ANOVA analysis of chemical composition and calorific value of wheat straw (Table S3). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hanlj@cau.edu.cn. Tel.: +86 10 6273 6313. Fax: +86 10 6273 6313.

Present Address

[†]China Agricultural University (East campus), Box 191, Beijing 100083, China.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Agro-Scientific Research Project in the Public Interest (No. 201003063), Natural Science Foundation of China (No. 31271611), and New Century Excellent Talents in University (No. NCET-12-0524).

■ REFERENCES

- (1) Zhou, X.; Wang, F.; Hu, H.; Yang, L.; Guo, P.; Xiao, B. *Biomass Bioenergy* **2011**, *35*, 1–11.
- (2) Talebnia, F.; Karakashev, D.; Angelidaki, I. *Bioresour. Technol.* **2010**, *101* (13), 4744–4753.
- (3) Vassilev, S. V.; Baxter, D.; Andersen, L. K.; Vassileva, C. G. *Fuel* **2013**, *105*, 40–76.
- (4) Bruun, S.; Jensen, J. W.; Magid, J.; Lindedam, J.; Engelsen, S. B. *Ind. Crop. Prod.* **2010**, *31*, 321–326.
- (5) Pasquini, C. *J. Brazil Chem. Soc.* **2003**, *14*, 198–219.
- (6) Lindedam, J.; Bruun, S.; DeMartini, J.; Jorgensen, H.; Felby, C.; Yang, B.; Wyman, C. E.; Magid, J. *J. Biobased Mater. Biol.* **2010**, *4*, 378–383.
- (7) Liu, L.; Ye, X. P.; Wormald, A. R.; Sokhansanj, S. *Carbohydr. Polym.* **2010**, *81*, 820–829.
- (8) Chen, L. J.; Xing, L.; Han, L. J.; Yang, Z. L. *Biosys. Eng.* **2009**, *103*, 313–320.
- (9) Lomborg, C. J.; Thomsen, M. H.; Jensen, E. S.; Esbensen, K. H. *Bioresour. Technol.* **2010**, *101*, 1199–1205.
- (10) Kaneko, H.; Funatsu, K. *AIChE J.* **2012**, *58*, 1829–1840.
- (11) Liu, X.; Han, L. J. *J. Anim. Feed Sci.* **2006**, *15*, 329–336.
- (12) Huang, C. J.; Han, L. J.; Liu, X.; Ma, L. *Energy Sources, A* **2011**, *33*, 114–120.
- (13) Huang, C. J.; Han, L. J.; Yang, Z. L.; Liu, X. *Waste Manage.* **2009**, *29*, 1793–1797.
- (14) Huang, C. J.; Han, L. J.; Yang, Z. L.; Liu, X. *Energy Convers. Manage.* **2008**, *49*, 3433–3438.
- (15) Stubbs, T. L.; Kennedy, A. C.; Fortuna, A. M. *J. Agric. Food Chem.* **2010**, *58*, 398–403.
- (16) Sheng, K. C.; Shen, Y. Y.; Yang, H. Q.; Wang, W. J.; Luo, W. Q. *Spectrosc. Spectr. Anal.* **2012**, *32* (10), 2805–2809.
- (17) Jacobs, J. L.; Hill, J.; Jenkin, T. *Anim. Prod. Sci.* **2009**, *49*, 595–607.
- (18) ASTM International. *Annual Book of ASTM Standards*; ASTM International: West Conshohocken, PA, 2007; ASTM E1757–01.
- (19) Grandy, A. S.; Erich, M. S.; Porter, G. A. *Soil Biol. Biochem.* **2000**, *32*, 725–727.
- (20) Thiex, N. J.; Manson, H.; Anderson, S.; Persson, J. A. *J. AOAC Int.* **2002**, *85*, 309–317.
- (21) Sluiter, J. B.; Ruiz, R. O.; Scarlata, C. J.; Sluiter, A. D.; Templeton, D. W. *J. Agric. Food Chem.* **2010**, *58*, 9043–9053.
- (22) M?ller, J. *J. AOAC Int.* **2009**, *92*, 74–90.
- (23) Jiang, H.; Liu, G. H.; Xiao, X. H.; Mei, C. L.; Ding, Y. H.; Yu, S. *Microchem. J.* **2012**, *102*, 68–74.
- (24) Malley, D. F.; McClure, C.; Martin, P. D.; Buckley, K.; McCaughey, W. P. *Commun. Soil Sci. Plan* **2005**, *36*, 455–475.
- (25) Chang, C. W.; Laird, D. A. *Soil Sci.* **2002**, *167*, 110–116.
- (26) Niu, W. J.; Liu, X.; Huang, G. Q.; Chen, L. J.; Han, L. J. *Energy Fuels* **2013**, *27*, 5940–5947.
- (27) Workman, J.; Weyer, L. *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*; CRC Press Inc.: Boca Raton, FL, 2012; pp 10–309.
- (28) Li, Y. M.; Yang, Z. P. *Asian J. Agric. Res.* **2009**, *1*, 32–34.
- (29) Liu, X.; Han, L. J. *Near Infrared Spec.* **2006**, *14*, 333–339.
- (30) Tamaki, Y.; Mazza, G. *J. Agric. Food Chem.* **2011**, *59*, 6346–6352.
- (31) Wittkop, B.; Snowdon, R. J.; Friedt, W. *J. Agric. Food Chem.* **2012**, *60*, 2248–2256.

■ NOTE ADDED AFTER ASAP PUBLICATION

Table 1 published November 20, 2014 with incorrect abbreviation presentation of uppercase and lowercase letters. The correct version published November 21, 2014.