

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-infrared Spectroscopy and Spectral Multi-Preprocessing Techniques

Bijendra Shrestha ¹, Jetsada Posom ^{2,3*}, Bim Prasad Shrestha ^{4,5} and Panmanas Sirisomboon ¹

¹ Department of Agricultural Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand; 63601254@kmitl.ac.th (B.S.); panmanas.si@kmitl.ac.th (P.S.)

² Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand; jetspo@kku.ac.th

³ Center for Alternative Energy Research and Development, Khon Kaen University, Thailand; jetspo@kku.ac.th (J.P.)

⁴ Department of Mechanical Engineering, Kathmandu University, Dhulikhel, PO Box 6250, Nepal; shrestha@ku.edu.np (B.P.S.)

⁵ Department of Bioengineering, University of Washington, Seattle, United State of America; shresth1@uw.edu (B.P.S)

* Correspondence: jetspo@kku.ac.th

Abstract: In this study, the partial least squares regression (PLSR) models were developed using no preprocessing, traditional preprocessing, multi-preprocessing 5 range, multi-preprocessing 3 range, genetic algorithm (GA), and successive projection algorithm (SPA) to assess the higher heating value (HHV) and ultimate analysis of grounded biomass for energy usage employing near-infrared (NIR) spectroscopy. A novel approach was utilized based on the assumption that using multiple pretreatment methods across different sections in the entire NIR wavenumber range would enhance the performance of the model. The performance of the model obtained from 200 biomass samples for HHV and 120 samples for ultimate analysis was compared, and the best model was selected based on the coefficient of determination of validation set, root mean square error of prediction, and the ratio of prediction to deviation values. Based on model performance results, the proposed HHV model from GA-PLSR, and the N and O models from the multi-preprocessing PLSR 5 range method could be used for most applications, including research, whereas the C and H models from GA-PLSR performance is fair and applicable only for rough screening. The overall findings highlight that the multi-preprocessing 5-range method, which was attempted as a novel approach in this study to develop the PLSR model, demonstrated better accuracy for HHV, C, N, and O, improving by 4.1839%, 8.1842%, 3.7587%, and 35.9404%, respectively. Therefore, it can be considered a reliable and non-destructive alternative method for rapidly assessing biomass properties for energy usage and can also be used effectively in biomass trading. However, due to the smaller number of samples used in the model development, more samples are needed to update the model for a robust application.

Keywords: Higher heating value; Ultimate analysis; Spectral multi-preprocessing method; Near-infrared spectroscopy; Partial least squares regression.

Highlights:

- NIRS was developed using PLSR to access HHV and ultimate analysis parameters of ground biomass for energy usage.
- Ground biomass scanning in transreflectance mode could improve model accuracy compared to the diffuse reflectance mode in NIRS.
- The novel PLSR multi-preprocessing 5 range method improved the model accuracy of HHV, C, N, and O models by 4.1839%, 8.1842%, 3.7587%, and 35.9404%, respectively.
- PLSR combined with GA improved model accuracy for HHV, C, and H by 8.5069%, 15.2159 %, and 0.0092 %, respectively.
- The use of the multi-preprocessing PLSR model could be beneficial for biomass trading.

Abbreviations

% = percentage	PLSR = partial least squares regression
C = carbon	R^2 = coefficient of determination
CHNS = CHNS Elemental analyzer	R^2_C = coefficient of determination of calibration set
GA = genetic algorithm	R^2_P = coefficient of determination of validation set
H = hydrogen	RMSEC = root mean square error of calibration set
HHV = higher heating value	RMSEP = root mean square error of prediction set
LVs = latent variable number	RPD = ratio of prediction to deviation
Max = maximum	S = sulfur
Min = minimum	SD = standard deviation
MP = multi-preprocessing	SEC = standard error of calibration set
MSC = multiplicative scatter correction	SEL = standard error of laboratory
N = nitrogen	SEP = standard error of validation set
N_T = total number of sample	SNV = standard normal variate
N_c = number of sample in calibration set	SPA = successive projection algorithm
NIRS = near infrared spectroscopy	SW = selected wavenumber
N_p = number of sample in validation set	wt.% = weight percentage

1. Introduction

Biomass is an important carbon-neutral, renewable bio-resource that is widely available throughout the world. It mainly consists of three polymers: cellulose, hemicellulose, and lignin, whose composition varies based on the type of biomass [1]. Hardwood and herbaceous biomass contain approximately 43-47% and 33-38% cellulose, 25-35% and 26-32% hemicellulose, and 16-24% and 17-19% lignin, respectively [2]. This composition of biomass can be converted into useful energy through various processes such as combustion, gasification, torrefaction, or fermentation, making it a suitable alternative to fossil fuels. However, its low energy density, high moisture content, and high oxygen-carbon ratio make it challenging to store, transport, and utilize effectively. Therefore, a deep understanding of biomass properties is necessary to design the best thermal conversion methods [3-5].

5]. In the current scenario, biomass is used mainly by the residential (cooking and heating) and industrial (combined heat and power) sectors through direct combustion, which negatively impacts health, the economy, energy, and the environment [6]. Research on bio-based energy technologies such as clean cooking stoves, gasifiers, biogas, bio-char, bio-briquettes, and pellets have yielded strong results in laboratory settings. However, due to inadequate and unreliable knowledge regarding the properties of biomass fuel, the overall efficiency and performance of these technologies remain only satisfactory. Additionally, various operation and maintenance challenges persist. Trading biomass based on volume and weight rather than its actual energy properties is still common. Therefore, rapid, reliable, and non-destructive assessment of biomass properties is of utmost importance for identifying the actual energy potential and for proper technical and monetary management and utilization [5].

Biomass can be assessed for energy usage by evaluating its HHV and ultimate analysis. The HHV is an important and standard indicator of the energy content of biomass [7]. A bomb calorimeter is used to measure the HHV, which is destructive in nature [8]. The ultimate analysis provides information on the elemental composition of biomass in terms of wt.% of C, H, N, S, and O. The heating value of the biomass is directly correlated with C, H, and O composition [9]. Biomass with higher C and H and/or O and H contents and lower N and S contents is recommended for energy usage as it improves the HHV of the biomass [9, 10].

Biomass is a good absorber of NIR radiation in the range of 3,595 to 12,489 cm⁻¹. It predominantly interacts with the bonds of nonsymmetrical molecules, including C, O, H, and N [11, 12], making it suitable for use in conjunction with NIRS and chemometrics for assessing the energy-related properties of biomass, including HHV and ultimate analysis parameters such as C, H, N, S, and O [13]. Several previous studies have utilized NIRS to develop models for rapid and accurate measurement of various biomass properties for energy usage. For instance, Posom et al. [14] developed a reliable online method for measuring the HHV of sugarcane using NIRS. Phuphaphud et al. [15] developed spectroscopic models using visible and shortwave NIR to predict and classify the energy content of growing cane stalks for breeding programs. Huang et al. [10] developed a prediction model for the HHV as well as the elemental composition (C, H, and N) of straw using NIRS. Posom et al. predicted the HHV [16] and elemental composition (C, H, N, O, and S) [17] of grounded bamboo using NIRS. Skvaril et al. [18] reviewed the application of NIRS in biomass energy conversion processes. Zhang et al. [19] studied the fast analysis of HHV and elemental composition of sorghum biomass using NIRS. Xue et al. [20] studied the use of an online NIRS system for measurement of crop straw fuel properties. These studies demonstrate the potential for NIRS to provide rapid, reliable, and non-destructive alternative methods for characterizing biomass for energy usage compared to traditional destructive thermal analysis techniques.

Despite NIRS being a rapid, reliable, and non-destructive analytical method, individual calibration models based on spectral data and each reference parameter must be developed for the NIR-based assessment of biomass properties. This procedure might be time-consuming and costly; however, in the long term, it will be beneficial for rapid and reliable evaluation procedures to assess biomass properties for their different applications.

In this study, a built-in code in MATLAB-R2020b was used to develop PLSR calibration models using spectral data from ten different biomass varieties (including five fast-growing tree varieties and five agricultural residue varieties), reference data obtained from a bomb calorimeter for HHV(J/g), and a CHNS/O elemental analyzer for wt.% of C, N, H, S, and O. The main objectives of this research are:

1. To develop PLSR models using no preprocessing, traditional preprocessing, multi-preprocessing 5-range and 3-range methods, GA, and SPA for assessing biomass properties for energy usage employing NIRS.
2. To compare the performance of the PLSR models based on R²_C, RMSEC, R²_P, RMSEP, RPD, and bias.
3. To select the better-performing PLSR-based model for each parameter and establish it as a reliable and non-destructive alternative method for rapidly assessing biomass properties for energy usage.

2. Materials and Methods

Figure 1 shows the overall research methodology for the evaluation of HHV and ultimate analysis parameters of grounded biomass for energy usage using NIRS combined with PLSR.

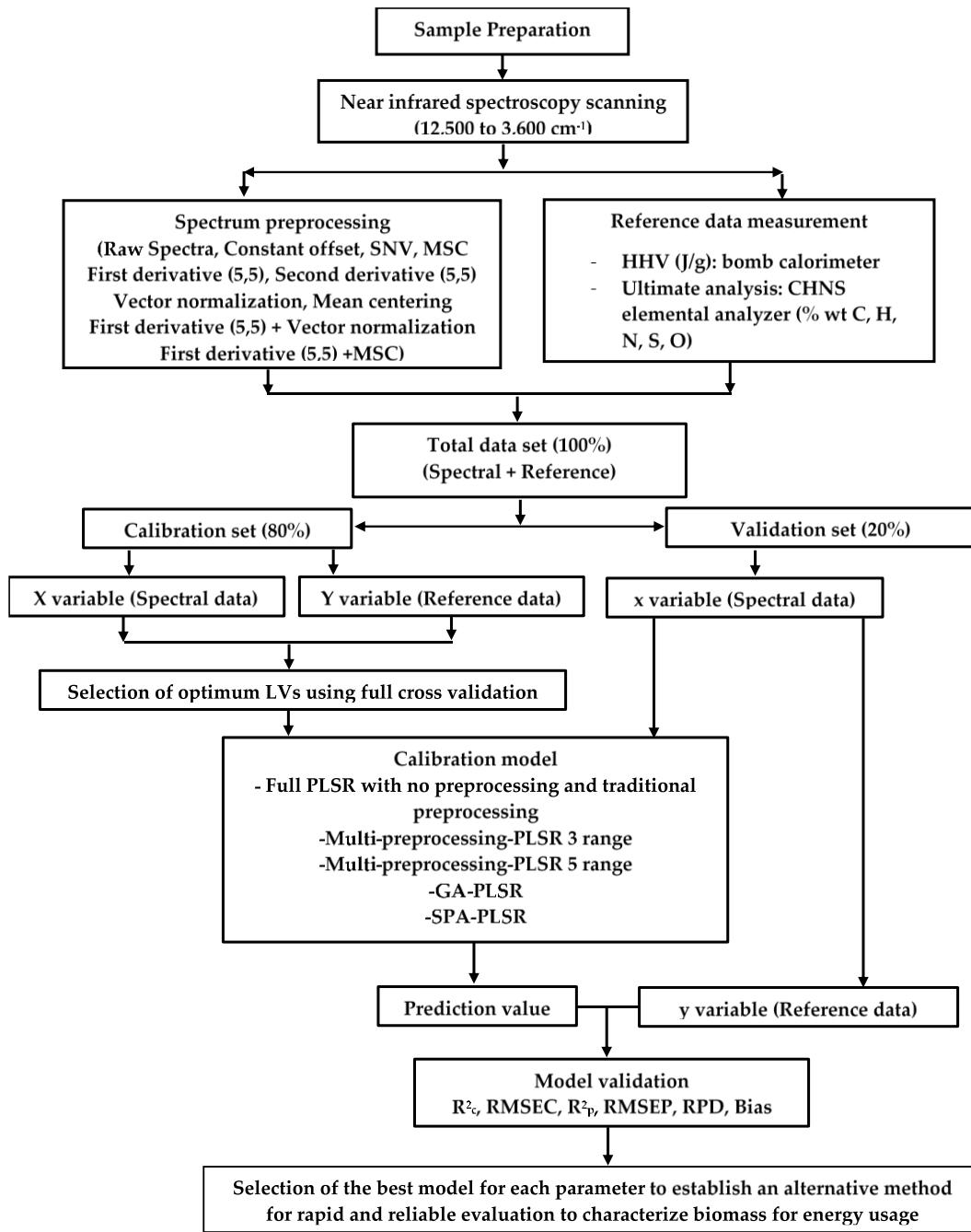


Figure 1. Flowchart of the overall research methodology for the evaluation of the HHV and ultimate analysis parameters of grounded biomass for energy usage using NIRS combined with PLSR.

2.1 Sample preparation

The biomass samples were collected from the Terai low flatland and mid-hill regions of Nepal, with altitudes ranging from 86 to 1,940 meters above sea level. The study included five fast-growing species: (1) *Alnus nepalensis*, (2) *Pinus roxburghii*, (3) *Bombusa vulagris*, (4) *Bombax ceiba*, and (5) *Eucalyptus camaldulensis*. Also included were five agricultural residues: (1) *Zea mays* (cob), (2) *Zea mays* (shell), (3) *Zea mays* (stover), (4) *Oryza sativa*, and (5) *Saccharum officinarum*. *Alnus nepalensis* and *Pinus*

roxburghii were collected from the mid-hill region; *Bombax ceiba*, *Eucalyptus camaldulensis*, and *Saccharum officinarum* were collected from the Terai region; and *Zea mays* (cob, shell, stover), *Bombusa vulagris* and *Oryza sativa* were collected from both Terai and mid-hill region of Nepal.

During preparation, all collected samples except for *Oryza sativa* were manually chopped into smaller pieces, i.e. less than 30 mm × 15 mm (refer to figure 2a); dried in the open sun; and stored in an airtight aluminum bag to maintain their biomass properties by preventing the exchange of air and moisture during transport to the Near-Infrared Spectroscopy Research Center for Agricultural Product and Food at School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand. The samples were ground using a multi-functional high-speed disintegrator (WF-04, Thai grinder, Thailand). The particle size of the grounded biomass was evaluated at Scientific and Technological Research Equipment Center (STREC) at Chulalongkorn University, Bangkok, Thailand, using the instrument Mastersizer 3000 (MAL1099267, Hydro MV). Figure 3 shows the representative particle size distribution of the ground biomass used in this research, ranging from 0.01 to 3080 μm . The ground samples were stored in airtight plastic zip lock bags before and during the experiment.

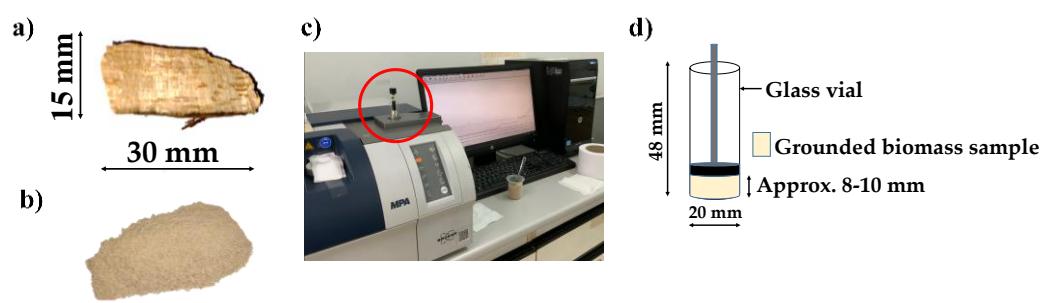


Figure 2. Nepal biomass in a) chips form ($> 30 \text{ mm} \times 15 \text{ mm}$), b) grounded form (1.88-3080 μm), c) FT-NIR (MPA, Bruker, Ettlingen, Germany) scanning between the wavenumber range 3,595 to 12,489 cm^{-1} , and d) ground sample presentation by transreflectance mode.

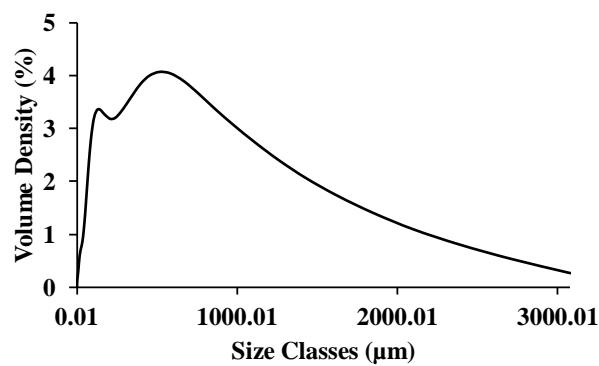


Figure 3. Representative particle size distribution of the ground biomass ranging from 0.01 to 3080 μm .

2.2. Spectral data collection

As shown in figures 2c) and 2d), the grounded biomass samples were placed in a glass vial (20 mm diameter and 48 mm height) and scanned using an FT-NIR spectrometer (MPA, Bruker, Ettlingen, Germany) in a transreflectance mode at the controlled temperature of $25 \pm 2^\circ\text{C}$. The spectrometer operates with a resolution of 16 cm^{-1} , with a background scan time and sample scan time of 32 scans (average), logging absorbance data - $\log(1/R)$ within wavenumber range of 3,595 to 12,489 cm^{-1} , where R is the diffuse reflectance detected from the grounded biomass sample. Prior to scanning, the FT-NIR spectrometer was normalized by performing a gold plate background scan. The primary

purpose of performing a background scan on every new ground sample was to compensate for instrumental drift and ambient environmental influences such as temperature, light, relative humidity, etc. on the measurement setup [12].

All the grounded samples were scanned twice without changing their positions, with no NIR leakage occurring during scanning. The average absorbance value for each sample, with respect to its wavenumber, was considered spectroscopic data for model development. Figure 4a) shows the raw spectrum of ten different grounded biomasses within the wavenumber range between 3,595 to 12,489 cm^{-1} , which were used to evaluate HHV and ultimate analysis parameters.

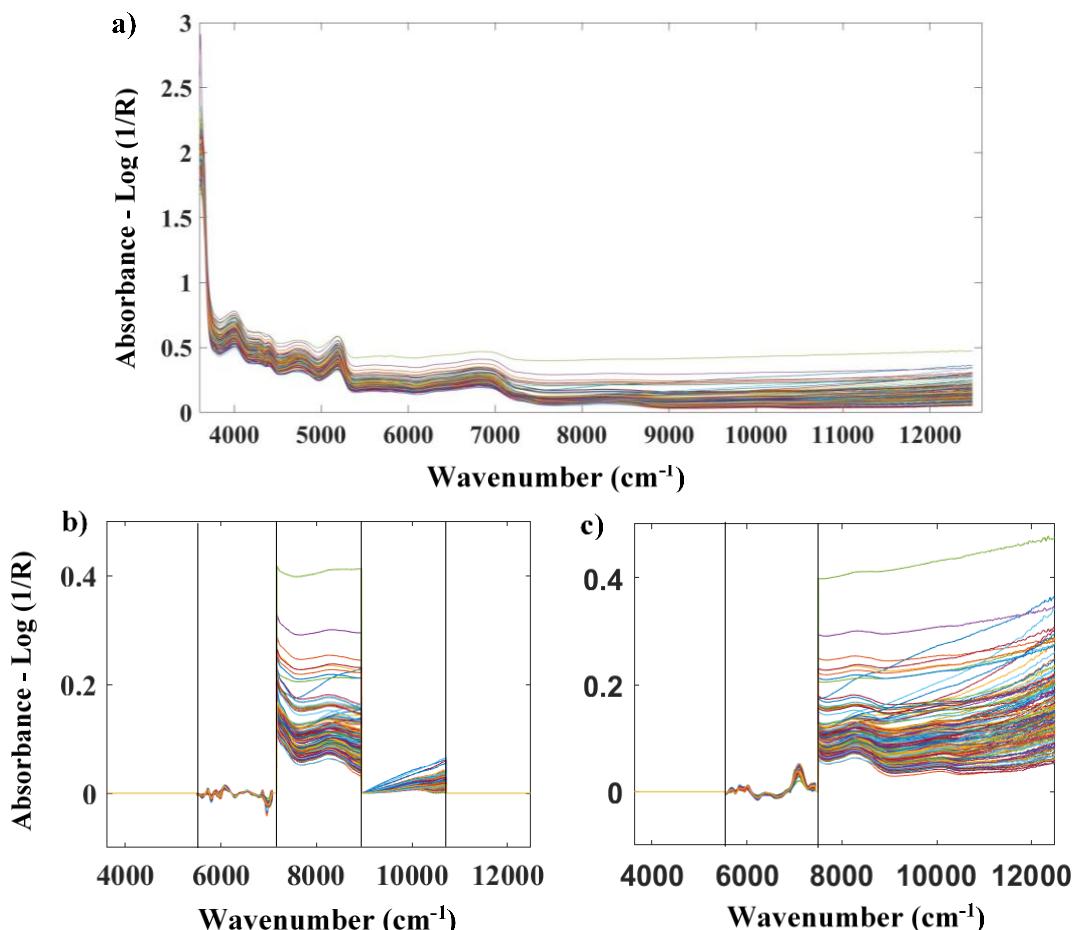


Figure 4. a) Raw spectra of grounded biomass. Pretreated spectra of the grounded biomass using b) the multi-preprocessing five range method (0, 5, 1, 6, and 0) and the multi-preprocessing 3-range method (0, 4, and 1).

2.3. Reference analysis

Due to the complex nature of NIR absorbance data, it must be correlated with reference values obtained using a standard laboratory method [21]. Thus, the reference data, which include HHV, C, H, N, S, and O, were evaluated after being scanned from a FT-NIR spectrometer.

2.3.1 Higher heating value (HHV)

The HHV of the grounded biomass is measured using the isoperibol method with an automatic bomb calorimeter (IKA C 200, Germany). Before the start of the experiment, the bomb calorimeter was calibrated with two tablets of benzoic acid (IKA C 723), each with a total weight of 1.0092 g and a gross calorific value of 26,462 J/g. To verify the calibration, the test was repeated with a single tablet of benzoic acid, and the results were compared. A cotton thread (IKA C 170.4) with a gross calorific value of 50 J/cotton twist was used for ignition in the bomb to measure the HHV of the grounded

sample. To ensure that the space in the bomb was saturated with water vapor throughout the entire experiment period, 2 ml of aqua pro (IKA C5003.1) were added into 1 liter of water and poured into the bomb calorimeter vessel [14]. The HHV (J/g) of each grounded sample was replicated twice, and the average value was considered as reference data for model development. A quantity of 0.5 ± 0.2 g of grounded sample was weighted using an electronic balance (Mettler Toledo JS1203C) with a resolution of 0.0001 g. Including preparation, the total experimental time to measure HHV for each sample was approximately 40 minutes.

2.3.2 Ultimate analysis

The ultimate analysis includes measurement of C, H, N, S, and O. The wt.% of C, H, N, and S in the ground sample were measured using the CHNS/O analyzer (Thermo ScientificTM FLASH 2000). The wt.% of O in the grounded biomass sample was calculated as a difference:

$$\text{wt.\% O} = 100 - (\text{wt.\% C} + \text{wt.\% H} + \text{wt.\% N} + \text{wt.\% S}) \quad (3)$$

Outliers for all the measured reference data were calculated using the following equation, where X_i is the measured value of sample i , and \bar{X} and SD are the average and standard deviation of the measured values of all samples, respectively:

$$\frac{(X_i - \bar{X})}{SD} \geq \pm 3 \quad (4)$$

If equation (4) is satisfied for any sample i , the sample is considered as an outlier and is not considered in the total data set for model development [22].

Similarly, the standard error of laboratory (SEL), which explains the precision of the reference method, was calculated for the bomb calorimeter and CHNS/O elemental analyzer using the following equation, where, y_1 and y_2 are the replicates of each sample reference value measurement and N_T is the total number of experiment samples:

$$SEL = \sqrt{\frac{\sum_{i=1}^{N_T} (y_1 - y_2)^2}{N_T}} \quad (5)$$

2.4. Spectral preprocessing

Spectral preprocessing is one of the important components of NIR calibration. Ten different varieties of grounded biomass samples were scanned to collect spectral data, whose physical, chemical, and biological properties may vary from sample to sample. Although the raw spectrum for all the biomass samples appears similar, instrumental errors, variations in light scattering during sample scanning, and a large number of redundant and interfering variables can introduce unwanted and harmful signals into the spectrum (refer to figure 4a). To improve spectral features, remove noise, address overlapping peaks and baseline shifts, handle collinearity within the spectral data, and enable easy data interpretation for calibration [23, 24], NIR spectral preprocessing is necessary before model development.

In this study, the raw spectrum was pre-treated with two approaches. The first approach was a traditional approach involving the entire raw spectrum, i.e. the full wavenumber range from 3,595 to 12,489 cm^{-1} using no preprocessing or traditional preprocessing methods. The second approach was a novel multi-preprocessing approach, where the entire wavenumber range was divided into different sections and pretreated using a combination set of various preprocessing methods.

For the traditional approach, ten different types of spectrum pretreatment methods were used for calibration models. These included (1) first derivative (segment=5 and gap=5), (2) second derivative (segment=5 and gap=5), (3) constant offset, (4) SNV, (5) MSC, (6) vector normalization, (7) min-max normalization, (8) mean centering, (9) first derivative (segment=5 and gap=5) + vector normalization, and (10) first derivative (segment=5 and gap=5) + MSC.

For the multi-preprocessing approach, the entire wavenumber range was divided into different sections and pre-treated with various pretreatment combination sets obtained from seven different preprocessing methods and marked as follows: 0 = empty (all the absorbance values = 0), 1 = raw spectra, 2 = SNV, 3 = MSC, 4 = first derivative (5,5), 5 = second derivative (5,5) and 6 = constant offset.

For the multi-preprocessing 5-range method, the entire wavenumber range was equally divided into five sections: 3,625.72–5,392.30 cm^{-1} , 5,400.02–7,166.59 cm^{-1} , 7,174.31–8,940.89 cm^{-1} , 8,948.60–10,715 cm^{-1} , and 10,722.9–12,489.48 cm^{-1} . However, the wavenumber range from 3,595 to 12,489 cm^{-1}

could not be equally divisible by 5, so the last four independent variables were removed from the total dataset, leaving 1150 out of 1154 considered for model development. Similarly, for the multi-preprocessing 3-range method, the entire wavenumber range was divided into three sections: 3594.87–5492.59 cm⁻¹, 7498.314–5500.30 cm⁻¹, and 7506.02–12489.48 cm⁻¹.

Figures 4b) and 4c) show the spectrum of the grounded biomass obtained from the multi-preprocessing method with a) 5-range and b) 3-range methods, respectively. In figure 4b), the raw spectrum was pre-treated with the preprocessing combination set of 0, 5, 1, 6, and 0, i.e. empty (zero absorbance) from 3,625.72–5,392.30 cm⁻¹, second derivative from 5,400.02–7,166.59 cm⁻¹, raw spectra from 7,174.31–8,940.89 cm⁻¹, constant offset from 8,948.60–10,715 cm⁻¹, and empty (zero absorbance) from 10,722.9–12,489.48 cm⁻¹. Similarly, in figure 4c), the raw spectrum was pre-treated with the preprocessing combination set of 0, 4, and 1, i.e. empty (zero absorbance) from 3594.87–5492.59 cm⁻¹, first derivative from 7498.314–5500.30 cm⁻¹, and raw spectrum from 7506.02–12489.48 cm⁻¹. The best combination set for multi-preprocessing is determined by the optimum LVs obtained from full cross-validation.

MATLAB-R2020b (MathWorks, USA) built-in code was used to select the optimal combination set of multi-preprocessing methods for developing a PLSR calibration model.

2.5. Model development

The accuracy of the model is one of the major concerns in the NIRS. Accuracy can be improved by using different spectral pretreatments and appropriate data analysis methods. Various research articles related to NIR modeling have concluded that PLSR is one of the effective and commonly used quantitative analysis techniques [14, 25-27]. Therefore, this study proposes PLSR-based models that can handle highly collinear spectroscopic data [28] for the assessment of grounded biomass properties. In this study, the following models were developed to match its objectives: (1) full wavenumber range – PLSR with no preprocessing and traditional preprocessing techniques, (2) multi-preprocessing PLSR 3-range method, (3) multi-preprocessing PLSR 5-range method, (4) GA-PLSR, and (5) SPA-PLSR.

To develop PLSR models using different methods, the total data set obtained after the removal of outliers was manually divided into an 80% calibration set and a 20% validation set as shown in Figure 1. The calibration set was designed to include the maximum and minimum reference values, thereby representing a wider range to generate a regression model [24]. The calibration set was first subjected to full cross-validation to select the optimal number of LVs. This number ensures the smallest possible standard error for data analysis; considering too few LVs leads to underfitting, and considering too many LVs leads to overfitting. If several LVs show similar or comparatively better model performance, the smallest number of LVs was selected for model development [29]. The PLSR models for assessing biomass properties for energy usage were created using in-house code in MATLAB-R2020b (Mathworks, USA).

GA and SPA are the wavelength selection methods that select the highly influential wavenumbers from the spectra and have been shown to provide better performance when combined with PLSR compared to PLSR with full wavenumber range only, thus avoiding overfitting [30-32]. SPA selects the variables with minimum collinearity and assesses them based on the value of the root mean square error obtained from the validation set. In SPA, uninformative variables are eliminated until the model's performance no longer increases [33]. GA selects variables with a minimum amount of redundant information, starting with one variable and adding a new one to the loop in each iteration, maximizing its fitness. The model developed with GA-PLSR shows the lowest prediction error as it maximizes the fitness and covariance between the spectral and reference data [34, 35]. In GA-PLSR and SPA-PLSR, the new calibration dataset was processed through full-cross validation to select the optimum LVs, which were then considered for PLSR model development.

The accuracy of the NIR model should be compared with the reference method. Therefore, the performance of the model was determined in terms of R^2_c , RMSEC, R^2_v , RMSEP, RPD, and bias [36]. These parameters can be calculated as follows, where y is the measured value, \hat{y} is the predicted value, i is subscript indicate number of sample, \bar{y} is the mean of the measured value, N_T is the number of samples, SD is the standard deviation of the measured values of the validation set and n is the number of samples in the validation set:

$$R^2_c, R^2_p = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (5)$$

$$RMSEC, RMSEP = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N_T}} \quad (6)$$

$$RPD = \frac{SD}{RMSEP} \quad (7)$$

$$Bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (8)$$

The better model was selected based on the tradeoff value between the highest R^2_c , R^2_p , and RPD and the lowest RMSEC, RMSEP, and bias. In this study, the performance results, namely R^2 and RPD value, were interpreted based on the recommendations of Williams et al. (2019) [37] and Zornoza et al. (2008) [38], respectively.

As per recommendations of Williams et al. (2019), R^2 up to 0.25 are not usable for NIRS calibration; 0.26-0.49 indicates poor calibration, and reasons for this should be researched; 0.50-0.64 is considered okay for rough screening; 0.66-0.81 is okay for rough screening and some other appropriate calibrations; 0.83-0.90 is usable with caution for most applications, including research; 0.92-0.96 is usable in most applications, including quality assurance; and 0.98+ is excellent and can be used in any application [37]. Similarly, according to Zornoza et al. (2008), an RPD value of less than 2 is considered insufficient for applications; RPD between 2 and 2.5 makes approximate quantitative predictions possible; RPD values between 2.5 and 3 are considered good for prediction, and RPD greater than 3 indicates an excellent prediction [38].

3.0. Result and Discussion

The energy potential and conversion efficiency of fast-growing trees and agricultural residues can be influenced by the composition of lignocellulosic matter [39]. Fast-growing trees tend to have a higher energy conversion efficiency than agricultural residues as their lignocellulosic matter contains less lignin and more cellulose and hemicellulose [40]. Figure 5 shows the near-infrared spectra of pure cellulose and pure lignin along with 90 samples of fast-growing trees and 110 samples of agricultural residues, all with average absorbance values. The raw spectra of the fast-growing trees and agricultural residues was compared with that of pure cellulose and pure lignin, revealing that both contain lignocellulosic matter as evidenced by the presence of the same distinct peaks. Significant positive or negative peaks in absorbance value relative to wavenumber can indicate important vibrations of particular bonds occurring at those frequencies, which could have a significant impact on the model's performance. Figure 5 clearly shows that the vibration band between approximately $5181-6150\text{ cm}^{-1}$ corresponds to the lignin band (with low absorbance for cellulose), while the range between approximately $6150-6800\text{ cm}^{-1}$ corresponds to cellulose band (with low absorbance for lignin) [25]. Additionally, important peaks were observed at approximately 4019 cm^{-1} , 4405 cm^{-1} , 4762 cm^{-1} , 5181 cm^{-1} and 6897 cm^{-1} . The peak at 4019 cm^{-1} results from the combination of C-H stretching and C-C stretching in cellulose, whereas the peak at 4405 cm^{-1} corresponds to the combination of O-H stretching and C-O stretching in cellulose. The peak at 4762 cm^{-1} corresponds to the combination of O-H bending and C-O stretching in polysaccharides. The peak at 5181 cm^{-1} corresponds to the combination of O-H stretching and HOH bending in polysaccharides. The peak at 6897 cm^{-1} corresponds to the first overtone of fundamental O-H stretching band in water and starch [41].

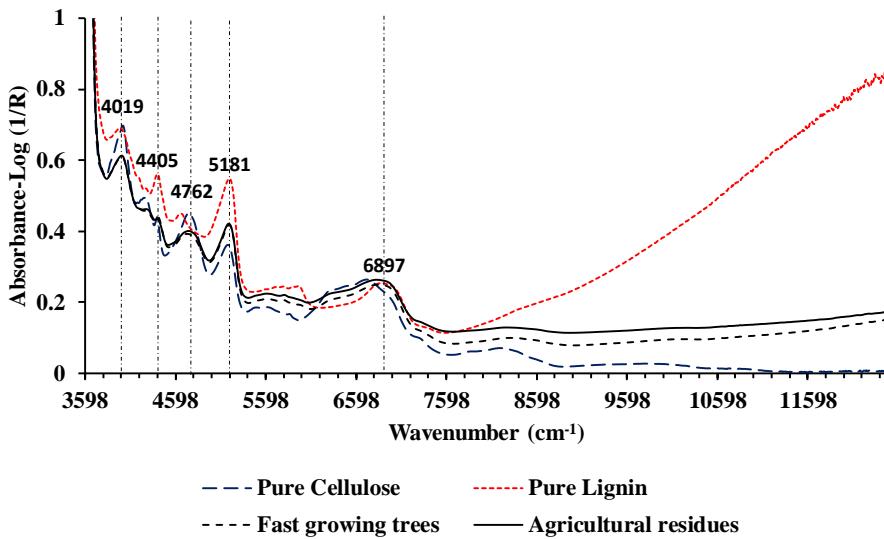


Figure 5. Spectra of fast-growing tree and agricultural residues compared to pure cellulose and pure lignin.

Table 1 presents the average HHV and ultimate analysis parameters (C, N, H and O) of different fast-growing trees and agricultural residues included as reference data for developing the model. As per previous research, the HHV of biomass is positively correlated with C and H contents, while it is negatively correlated with O and N contents [42]. Table 1 indicates that fast-growing trees have higher average values of HHV, C, and H contents and lower O and N contents compared to agricultural residues. These results are consistent with the correlation observed between measured data of the HHV and elemental composition of ground biomass.

Table 1. Average reference value of HHV and Ultimate analysis parameter of fast-growing trees and agricultural residues.

Category	Particular	HHV (J/g)	C (wt.%)	N (wt.%)	H (wt.%)	O (wt.%)
Fast growing tree	<i>Alnus nepalensis</i>	17932	45.9115	0.3115	5.7255	48.0515
	<i>Pinus roxburghii</i>	18349	46.8367	0.0606	5.8283	47.2744
	<i>Bombusa vulgaris</i>	17310	45.6132	0.2327	5.7536	48.4005
	<i>Eucalyptus camaldulensis</i>	17105	44.5536	0.0896	5.6164	49.7404
	<i>Bombax ceiba</i>	17077	44.8557	0.3162	5.8179	49.0102
Agricultural residue	<i>Zea mays</i> (cob)	17297	44.7794	0.2488	5.7619	49.2100
	<i>Zea mays</i> (shell)	16409	45.6518	0.4318	6.2113	47.7050
	<i>Zea mays</i> (stover)	16753	44.3988	0.7069	5.6697	49.2245
	<i>Oryza sativa</i>	15417	40.4261	0.4996	5.3042	53.7701
	<i>Saccharum officinarum</i>	17029	43.6413	0.1047	5.7047	50.6827

Table 2 shows the statistical summary data for the HHV (J/g) and ultimate analysis parameters, i.e. wt.% of C, N, H, and O used in the calibration set and validation set for the model development.

Table 2. Statistical data of the HHV and ultimate analysis parameters of the grounded biomass used in PLSR model development.

Parameter	Experimental Method	N _T	Calibration Set					Validation Set				
			N _c	Max	Min	Mean	SD	N _p	Max	Min	Mean	SD
HHV (J/g)	Bomb Calorimeter	196	157	18616	14682	16962	848	39	18553	14965	17049	836
C (wt.%)	CHNS/O	108	87	48.0000	38.3950	44.3278	2.1161	21	47.7400	40.8550	44.9039	1.8910
N (wt.%)	CHNS/O	95	76	0.8300	0.0000	0.2807	0.1870	19	0.8200	0.0000	0.3187	0.2506
H (wt.%)	CHNS/O	93	74	6.4800	4.9500	5.7448	0.3044	19	6.2550	5.1850	5.6911	0.3053
O (wt.%)	CHNS/O	96	77	54.3600	46.2600	49.5235	1.9326	19	53.5500	46.6167	49.3122	1.8671

3.1. Higher heating value

Out of the 200 samples, four were identified as outliers and were removed from the total data set to develop PLSR-based models for evaluating the HHV. The SEL for the bomb calorimeter used to evaluate HHV was calculated to be 255.7708 J/g. Table 3 displays the optimum result of various PLSR-based models using the full wavenumber range (3,594.87–12,489.5 cm⁻¹) to evaluate the HHV of the ground biomass from the fast-growing tree and agricultural residues.

Figure 6a) shows the scatter plot of HHV measured and predicted value from the calibration and validation sets using GA-PLSR. The GA-PLSR with LVs 14 and spectral pretreatment first derivative using 692 important wavenumbers yielded the best performance result, with R²_C of 0.9505, RMSEC of 188.0117 J/g, R²_P of 0.9574, RMSEP of 170.3282 J/g, RPD of 4.89, and bias of -21.9648 J/g. The model included a sufficient number of homogenous samples, from both fast-growing tree and agricultural residue, for model development and had a wider HHV range, resulting in higher R²_C, R²_P, and RPD, and lower RMSEC, and RMSEP values compared to other models. Compared to the full-PLSR model performance, the GA improved the PLSR model accuracy by 8.5069%. Similarly, the multi-preprocessing 5 range method improved the accuracy of the full-PLSR model by 4.1839%. According to Williams et al. (2019) [37] and Zornoza et al. (2008) [38], the GA-PLSR model for evaluating HHV is acceptable for most applications with excellent prediction, including quality assurance.

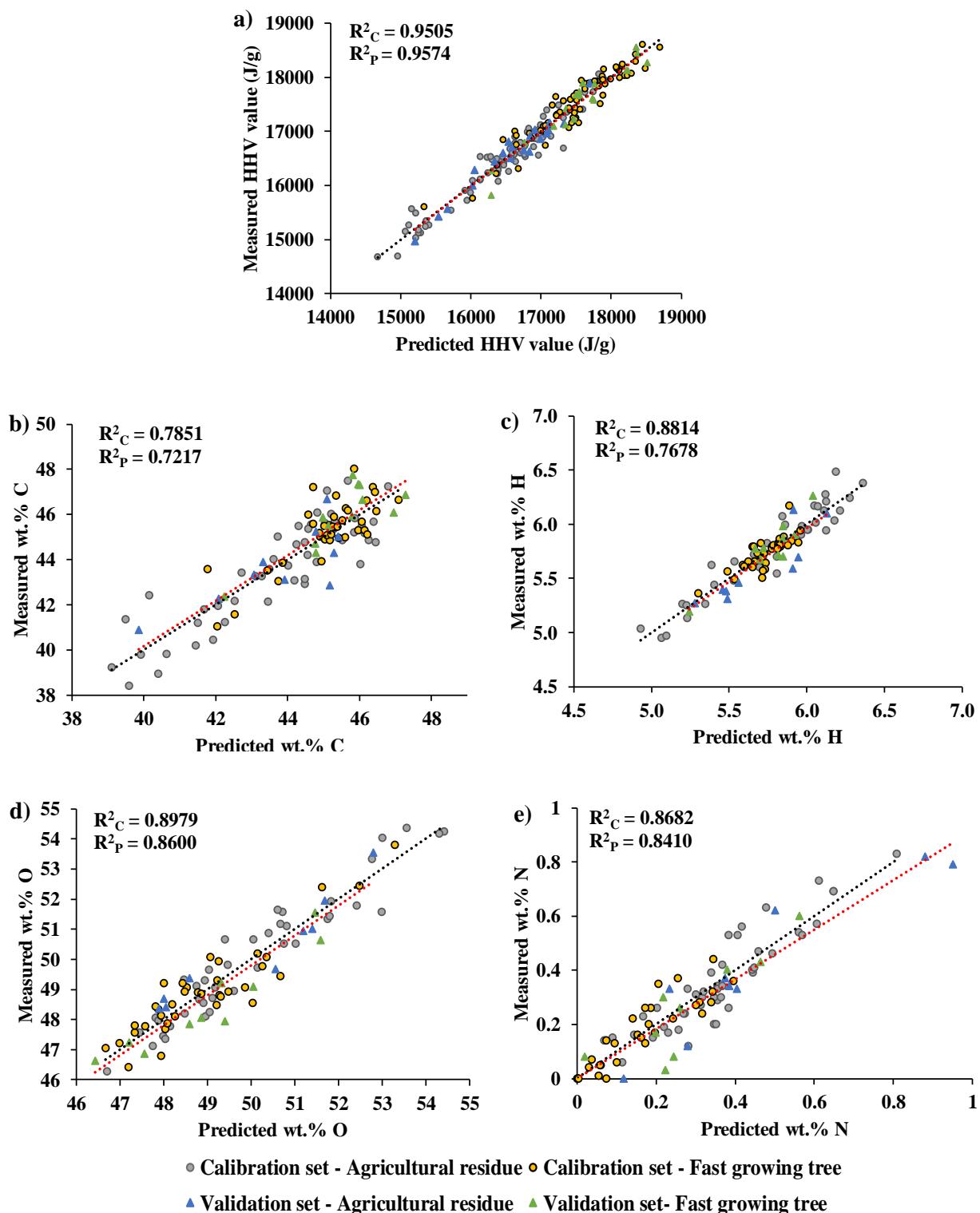


Figure 6. Measured versus predicted value in calibration and validation sets for the a) HHV, b) wt.% of C, c) wt.% of H, d) wt.% of O, and e) wt.% of N.

Figure 7 shows the average absorbance values obtained after preprocessing with the first derivative, highlighting the 692 selected wavenumbers (marked in red) obtained from GA, within the full spectral range of $3594.87\text{--}12489.5\text{ cm}^{-1}$. The figure highlights important peaks in the following ranges: $4003.73\text{--}4111.73\text{ cm}^{-1}$, $4366.3\text{--}4451.16\text{ cm}^{-1}$, $5091.45\text{--}5114.59\text{ cm}^{-1}$, and $5130.02\text{--}5292.02\text{ cm}^{-1}$, which may significantly influence the model performance.

In the range of 4003.73–4111.73 cm^{-1} , the wavenumber 4019 cm^{-1} represents the combination of C-H stretching and C-C stretching in cellulose and is used as a reference. Similarly, the range of 4366.3–4451.16 cm^{-1} includes the reference wavenumber 4405 cm^{-1} , which corresponds to the combination of O-H and C-O stretching in cellulose. Polysaccharides are characterized by the combination of O-H stretching and HOH bending, which is represented by the reference wavenumber 5102 cm^{-1} in the range of 5091.45–5114.59 cm^{-1} . Additionally, the range of 5130.02–5292.02 cm^{-1} includes the reference wavenumber 5200 cm^{-1} , which corresponds to the combination of O-H stretching and HOH deformation of O-H molecular water [41]. Lignocellulosic biomass derives its primary energy from cellulose, hemicellulose, and lignin [43, 44]. As can be seen in Figure 7, the important peaks with vibration bonds of C-H, C-C, O-H, and C-O stretching and HOH deformation of O-H molecular water correspond to the structure of cellulose and lignin. Therefore, they are likely to have the greatest influence on the assessment of the HHV of ground fast-growing trees and agricultural residues. This study is in line with previous studies by Sirisomboon et al. [45] and Lestander et al. [46], in which the authors reported that vibration bonds C-H, C-C, and O-H stretching contribute significantly to the HHV of bamboo and biofuels, respectively. Additionally, Zhang et al. [19] reported that the vibration bond of C-H stretching in aromatic and CH_3 structure can be used to assess the HHV of sorghum biomass. Posom et al. [5] indicated in their study that the vibration of C-H stretching highly influences the prediction of the HHV of *leucaena Leucocephala* pellets.

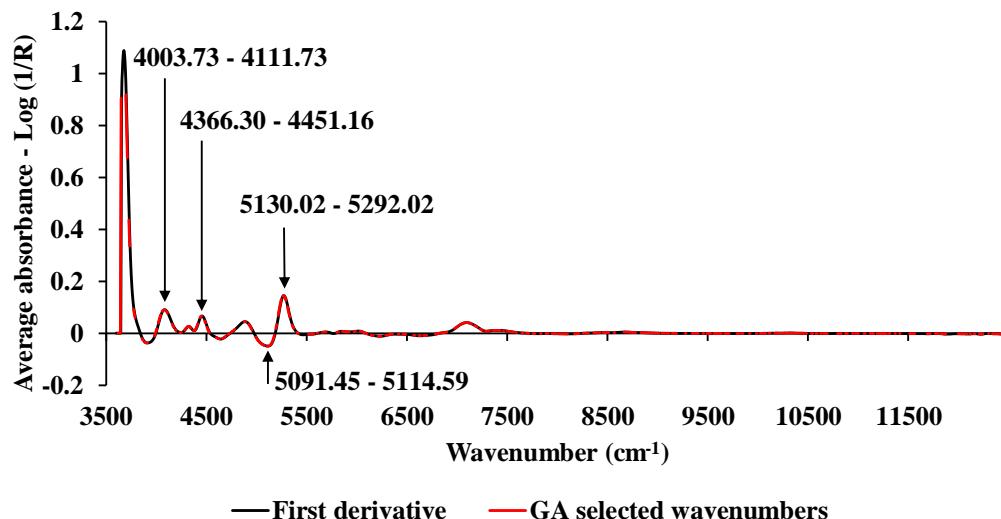


Figure 7. The average absorbance value of HHV (J/g) obtained using the first derivative pre-processing with a selection of important wavenumbers obtained from GA, within the full wavenumber range of 3594.87–12489.5 cm^{-1} .

Table 3. Results of the PLSR-based model for the HHV (J/g) and ultimate analysis (wt.%) of grounded biomass, bolded model showing the best performance.

Parameter	Algorithm	Pre-processing	LVs	Calibration Set		Validation set			
				R^2_c	RMSEC	R^2_p	RMSEP	RPD	bias
HHV (J/g)	Full-PLSR	First derivative(5,5) + Vector normalization	14	0.9527	183.6910	0.9491	186.1651	4.44	-13.5781
	SPA-PLSR	First derivative (SW = 479)	15	0.9469	194.7442	0.9486	187.0927	4.41	-0.1578
	GA-PLSR	First derivative (SW = 692)	14	0.9505	188.0117	0.9574	170.3282	4.89	-21.9648
	MP-PLSR: 3 range	Combination set: 5,5,4	15	0.9538	181.5744	0.9470	189.9072	4.35	-12.3121
	MP-PLSR: 5 range	Combination set: 4,4,5,4,3	13	0.9546	180.0513	0.9533	178.3761	4.72	-35.5676
wt.% C	Full-PLSR	Raw spectra	13	0.8433	0.8326	0.6129	1.1488	1.67	0.3193
	SPA-PLSR	First derivative + vector normalization (SW=70)	9	0.8001	0.9405	0.6316	1.1207	1.76	0.4010
	GA-PLSR	First derivative (g = 5, s = 5) (SW=50)	9	0.7851	0.9753	0.7217	0.9740	1.93	0.1877
	MP-PLSR: 3 range	Combination set: 5,0,6	13	0.8791	0.7315	0.6765	1.0502	1.78	0.1739
	MP-PLSR: 5 range	Combination set: 3,0,1,3,0	12	0.8451	0.8280	0.6737	1.0548	1.88	0.3807
wt.% N	Full-PLSR	First derivative(g = 5, s= 5)	9	0.8457	0.0730	0.8284	0.1011	2.63	-0.0403
	SPA-PLSR	Second derivative (g = 5, s= 5)(SW=601)	11	0.9091	0.0560	0.7691	0.1173	2.20	-0.0381
	GA-PLSR	Second derivative (g = 5, s= 5)(SW=990)	10	0.9026	0.0580	0.8010	0.1089	2.36	-0.0338
	MP-PLSR: 3 range	Combination set: 4,4,5	10	0.9196	0.0527	0.7960	0.1102	2.36	-0.0383
	MP-PLSR: 5 range	Combination set: 4,4,5,3,4	9	0.8682	0.0675	0.8410	0.0973	2.65	-0.0309
wt.% H	Full-PLSR	SNV	14	0.8335	0.1234	0.7678	0.1434	2.10	-0.0234
	SPA-PLSR	Raw (SW=1148)	14	0.8286	0.1252	0.6439	0.1776	1.68	0.0014
	GA-PLSR	SNV (SW=457)	14	0.8814	0.1041	0.7678	0.1434	2.14	-0.0356
	MP-PLSR: 3 range	Combination set: 4,5,6	13	0.8800	0.1047	0.6422	0.1780	1.68	-0.0145
	MP-PLSR: 5 range	Combination set: 4,0,6,4,6	13	0.8864	0.1019	0.6040	0.1872	1.60	-0.0254
wt.% O	Full-PLSR	MSC	14	0.8902	0.6363	0.6589	1.0612	1.76	-0.2534
	SPA-PLSR	Mean Centering (SW: 1082)	14	0.8942	0.6245	0.5945	1.1570	1.58	-0.1242
	GA-PLSR	Raw (SW:113)	14	0.8610	0.7160	0.8329	0.7426	2.45	-0.0267
	MP-PLSR: 3 range	Combination set: 5,4,6	13	0.8918	0.6316	0.7183	0.9643	1.94	-0.2229
	MP-PLSR: 5 range	Combination set: 3,0,1,0,1	14	0.8979	0.6136	0.8600	0.6798	2.78	-0.1909

3.2. Ultimate analysis

The sulfur content in the ground biomass samples of fast-growing trees and agricultural residue was not detected using the CHNS/O analyzer (Thermo Scientific™ FLASH 2000). This may be because the S content in the biomass is too low to be detected[47]. Therefore, PLSR-based models for the wt.% of S were not developed in this study.

3.2.1. wt.% of C

The SEL for the CHNS/O elemental analyzer used to evaluate the wt.% of C content in grounded biomass was calculated as 1.6936 wt.%. Table 3 shows the overall optimum results of PLSR-based models for the evaluation of wt.% of C content in the grounded biomass within the full wavenumber range of 3,594.87–12,489.5 cm⁻¹. Out of the 120 samples, 11 samples were identified as an outlier and removed from the total data set for model development. The model developed through GA-PLSR with spectrum preprocessing of first derivative (gap=5 and segment=5) and LVs 9 provided better results with R^2_c , RMSEC, R^2_p , RMSEP, RPD, and bias value as 0.7851, 0.9753 wt.%, 0.7217, 0.9740 wt.%, 1.93, and 0.1877 wt.%, respectively. Compared with full-PLSR, the GA-PLSR method improved the model accuracy by 8.5069%. Similarly, the multi-preprocessing 5 range method improved the PLSR model by 8.1842%. The scatter plot of the GA-PLSR method for the wt.% of C content in grounded biomass is shown in Figure 6b). According to the recommendation by Williams et al. (2019) [37], the PLSR model with GA method is usable for rough screening and some other appropriate calibrations,

based on the obtained R^2 value. Similarly, considering the RPD value, as suggested by Zornoza et al. (2008) [38], the model is acceptable for the prediction of wt.% C content in the grounded biomass.

Figure 8 shows the average absorbance values obtained after preprocessing with the first derivative, highlighting the 50 selected wavenumbers (marked in red) obtained from GA, within the full spectral range of 3,594.87–12,489.5 cm^{-1} . The high peaks with positive values marked in red at a specific wavenumber indicate the functional group, spectra-structure, and material type, which might be significant in the assessment of wt.% of C. In Figure 8, significant peaks can be noticed at 3650, 4019, 4405, 4878, and 7042 cm^{-1} , respectively.

The peak at 3650 cm^{-1} corresponds to the functional group of O-H, the spectral structure with the fundamental stretching vibrational absorption band of O-H (-CH₂-OH), and material type of primary alcohols. The peak at 4019 cm^{-1} corresponds to the functional group of C-H/C-C, the spectral structure of the C-H stretching and C-C stretching combination, and the material type of cellulose. The positive peaks at 4405 cm^{-1} and 4878 cm^{-1} are associated with the functional group O-H/C-H, and combination of N-H/C-N/N-H amide II and amide III; spectral structure O-H stretching and C-O stretching; N-H in-plane bend, C-N stretching and N-H in plane bend combination with material type cellulose and amides/proteins, respectively. The peak at 7042 cm^{-1} corresponds to O-H aromatic with the spectral structure of O-H first overtone of the fundamental stretching band, and the material type of hydrocarbons [41]. Lignin contains a high carbon content [48]. According to Zhang et al. [19], vibration bands related to C-H stretching, CH₂, C-H aromatics, O-H stretching, and HOH deformation are essential for predicting the C content of sorghum biomass. Similarly, Posom and Sirisomboon [49] found that N-H stretching, N-H deformation, C-N stretching, O-H stretching, and C-O stretching of starch significantly contribute to the model development of C content in bamboo. The average absorbance plot for wt.% of C shows the peaks at 3650, 4019, 4405, 4878, and 7042 cm^{-1} , which complement vibration bands reported in previous studies and also the spectra of pure lignin and pure cellulose. While these observed vibration bands at different peaks may have a significant impact on the overall performance of the model, this study suggests that the FT-NIRS may not provide sufficiently high resolution spectra to create an accurate prediction model for wt.% of C.

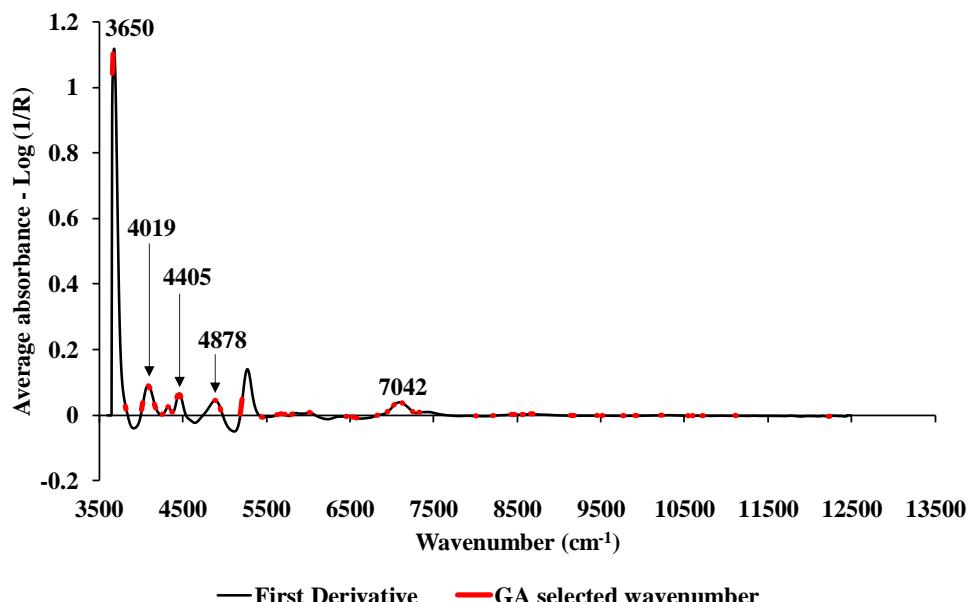


Figure 8. The average absorbance value of wt.% of C obtained using the first derivative preprocessing with a selection of important wavenumbers obtained from GA, within the full wavenumber range of 3594.87–12489.5 cm^{-1} .

3.3.2. wt.% of H

The SEL for the CHNS/O elemental analyzer to evaluate wt.% of H content in grounded biomass was calculated as 0.3206 wt.%. The optimum results of different PLSR-based models for evaluation of wt.% of H within the full wavenumber range were presented in Table 3. Before modeling, outliers

from the reference values were calculated, and 27 out of the 120 samples were detected as outliers. Therefore, 93 grounded biomass samples were used for the model development. The best model was developed from the wavelength selection method, GA-PLSR, within the wavenumber range of 3,595–12,489 cm^{-1} and spectral preprocessing from SNV. The best performing model for the evaluation of wt.% of H content in the grounded biomass produced R^2_c of 0.8814, RMSEC of 0.1041 wt.%, R^2_p of 0.7678, RMSEP of 0.1434 wt.%, RPD of 2.14, and bias of -0.0356 wt.%. The GA-PLSR model exhibits a minimal improvement in model accuracy of 0.0092% compared to the full-PLSR model.

Figure 6c) shows the scatter plot of measured versus predicted wt.% of H content in the grounded biomass obtained using GA-PLSR. According to William et al. (2019) [37], based on the R^2 value, the model can be used for rough screening and some other appropriate calibrations. To improve the performance of the model, it is recommended to include additional representative biomass samples with high concentration and wide range of wt.% of H content that are uniformly and representatively distributed in both the calibration and validation sets and are obtained from both fast-growing trees and agricultural residue varieties.

Figure 9 shows the average absorbance spectrum pre-treated with the SNV and highlights with red marks the important wavenumbers obtained using GA. The important peaks selected at 4019, 4608, 5155, 6897, and 8163 cm^{-1} may have a significant influence on the performance of the model for the evaluation of wt.% of H content in the grounded biomass samples. The peak at 4019 cm^{-1} is associated with the functional group of C-H/C-C, the spectral structure of C-H stretching and C-C stretching combination, with material type cellulose. The peak at 4608 cm^{-1} is associated with the combination of C-H stretching and C-H deformation in alkenes. Similarly, the peak at 5155 cm^{-1} corresponds to combination of O-H stretching and HOH bending in water. The peak at 6897 cm^{-1} corresponds to spectral structure O-H, arising from the first overtone of fundamental stretching band, with material type starch/polymeric alcohol. The peak at 8163 cm^{-1} is associated with second overtone of C-H fundamental stretching band and material type hydrocarbons [41]. The selected peaks mostly fall within a similar range compared to the study conducted by Posom and Sirisomboon [49]. This finding supports the results of the current study, indicating that these selected peaks are likely to have a significant influence on the performance of the models.

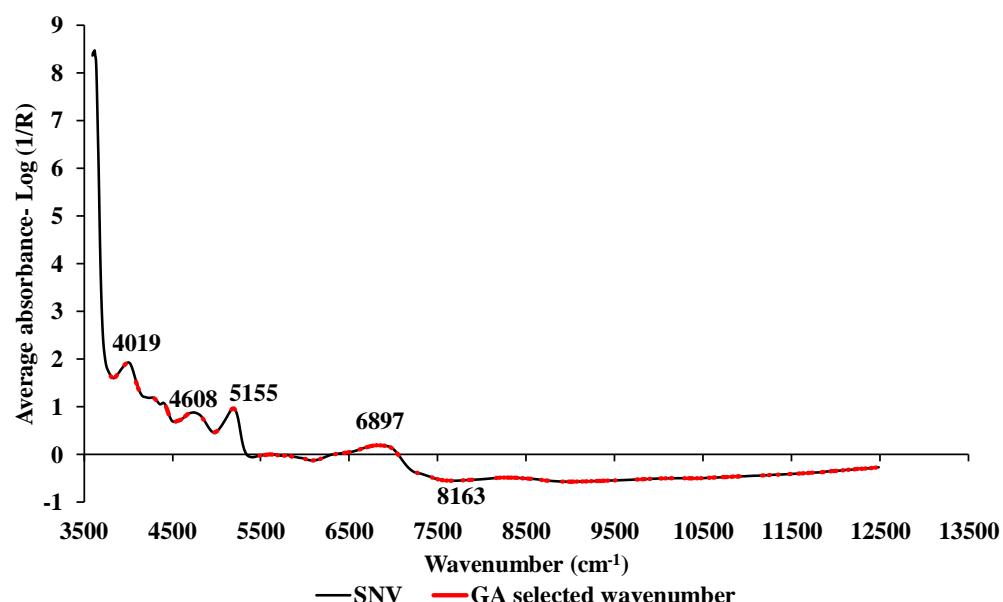


Figure 9. The average absorbance value of wt.% of H obtained using SNV preprocessing with a selection of important wavenumbers obtained from GA within full wavenumber range of 3594.87–12489.5 cm^{-1} .

3.3.3. wt.% of O

Based on the assumption that the sulphur content in biomass is zero, as its wt.% is too low to be detected by instrument, the wt.% of O in biomass is calculated using equation (3). The optimum results of the PLSR-based models for predicting the wt.% of O content in the grounded biomass are shown in Table 3. The best result was obtained from the multi-preprocessing PLSR 5-range method with a spectral preprocessing combination set of 3, 0, 1, 0, and 1, i.e. MSC, empty, raw spectra, empty, and raw spectra, respectively, from the range 3,625.72–12,489.5 cm^{-1} , which are equally divided into five sections. Figure 6d) shows the scatter plot for measured and predicted wt.% of O. With LVs 14, the best performing model for evaluating the wt.% of O content in the grounded biomass produced an R^2_c of 0.8979, RMSEC of 0.6136 wt.%, R^2_p of 0.8600, RMSEP of 0.6798 wt.%, RPD of 2.78, and a bias of -0.1909 wt.%. Compared with full-PLSR, the multi-preprocessing 5-range method improved the model accuracy by 35.9404%. Based on Williams et al. (2019) [37] and Zornoza et al. (2008) [38], the model with multi-preprocessing PLSR 5-range method is usable for good prediction with caution for most applications, including research.

Figure 10 shows the regression plot versus the entire wavenumber range for identifying the important wavenumbers that might play a significant role in producing better model results for evaluating wt.% of O. Important peaks were noticed at 3650, 4307, 4405, 5495, 8163, 8754, 11655, and 12300 cm^{-1} . The peak at 3650 cm^{-1} corresponds to the functional group of O-H with material type primary alcohols. Similarly, the peaks at 4307 cm^{-1} and 4405 cm^{-1} are associated with the spectra-structure of the C-H stretching and CH_2 deformation combination and the O-H stretching and C-O stretching combination, respectively, with material types polysaccharides and cellulose, respectively. The peaks at 5495 cm^{-1} and 8163 cm^{-1} might correspond to second overtone O-H stretching and C-O stretching combination, and C-H stretching, respectively, with cellulose and hydrocarbon as the respective materials. The peaks at 8754 cm^{-1} and 11655 cm^{-1} corresponds to the second overtone of the fundamental stretching band of C-H and the third overtone of the fundamental stretching band of C-H, with material type for both peaks as hydrocarbons and aromatic. Similarly, the peak at 12300 cm^{-1} is associated with the spectra-structure of C-H combination, with material type as hydrocarbon and aliphatic. A previous study by Posom and Sirisomboon [49] showed peaks at similar wavenumbers with vibration bands of C-H aromatic, C-H aliphatic, and O-H stretching of alcohol, which supports the finding of this study. Therefore, these vibration bands may have significant influence on the development of the model for the assessment of wt.% of O in ground biomass.

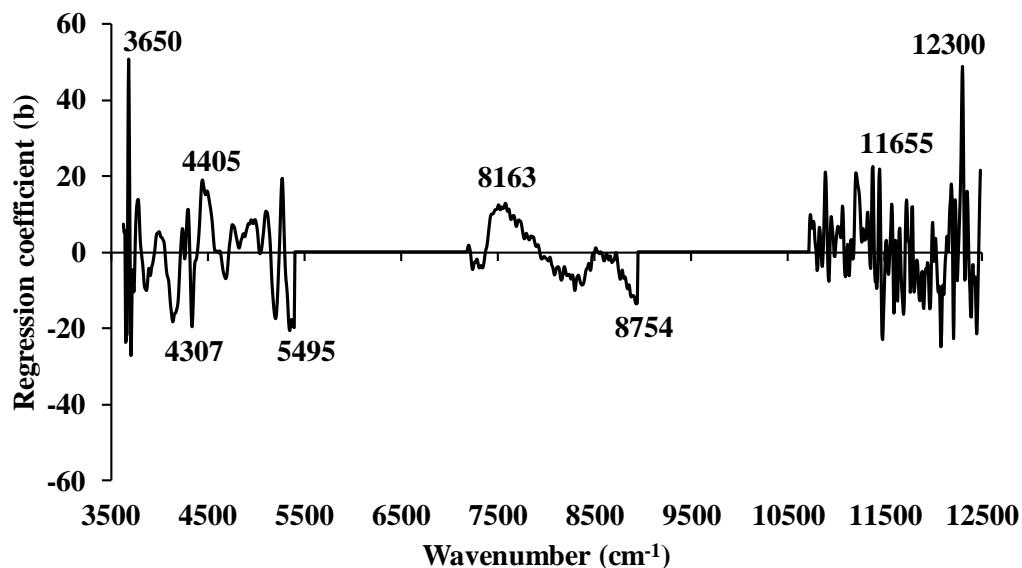


Figure 10. The regression coefficient for the wt.% O of grounded biomass using the multi-preprocessing PLSR 5 range method.

3.3.4. wt.% of N

The SEL of the CHNS elemental analyzer for evaluating wt.% of N content in grounded biomass was calculated as 0.0761 wt.%. Table 3 shows the optimal outcomes of the PLSR-based models for predicting the wt.% of N content in grounded biomass. The best prediction result of the wt.% of N in grounded biomass was obtained using the multi-preprocessing PLSR 5-range method with a spectral preprocessing combination set of 4, 4, 5, 3, and 4, which included the first derivative followed by the first derivative, second derivative, MSC, and first derivative, respectively, in five equally divided sections from 3,625.72–12,489.48 cm^{-1} . Figure 6e) shows the scatter plot of measured versus predicted wt.% of N content in the grounded sample using the multi-preprocessing PLSR 5-range method. The best performance for evaluating wt.% of N content in the grounded biomass resulted in an R^2_{C} of 0.8682, RMSEC of 0.0675 wt.%, R^2_{P} of 0.8410, RMSEP of 0.0973 wt.%, RPD of 2.65, and bias of -0.0309 wt.%. Compared with full-PLSR, the multi-preprocessing 5-range method improved the model accuracy by 3.7587%. According to William et al. (2019) [37], the model is suitable for most applications, including research. Based on the recommendation of Zornoza et al. (2008) [38], the prediction of wt.% of N content from the multi-preprocessing PLSR 5-range method with RPD value of 2.65 is considered good for prediction.

Figure 11 shows the regression coefficient plot for wt.% of N content in the grounded biomass obtained from the multi-preprocessing PLSR 5-range method. The figure displays numerous positive and negative high and low peaks. The high peaks at 4019, 4307, 4673, 5200, 5952, 6711, and 12453 cm^{-1} might significantly contribute to the evaluation of wt.% of N content. The negative peak at 4019 cm^{-1} might correspond to C-H stretching and C-C stretching combination with material type as cellulose. The positive peaks at 4307 cm^{-1} , 4673 cm^{-1} , 5200 cm^{-1} , and 5952 cm^{-1} might be associated with the structure of C-H stretching and CH_2 deformation combination (material: polysaccharides), C-H stretching and C=O stretching combination and C-H deformation combination (material: lipids), O-H stretching and HOH deformation combination (material: O-H molecular water), and C-H (first overtone of fundamental stretching band), aromatic C-H (material: hydrocarbons, aromatic), respectively. The peak at 6711 cm^{-1} might be associated with O-H (first overtone of fundamental stretching band) with material type as starch/polymeric alcohol. Common natures of peaks were noticed in the range between 11500 and 12500 cm^{-1} , for which 12453 cm^{-1} is described as a reference, which might correspond to the spectral structure of C-H combination, with material type as hydrocarbon and aliphatic [41]. The selected regression coefficient peaks show similar peaks compared to the study performed by Posom and Sirisomboon [49], with vibration bands of C-H stretching, C-C stretching, O-H stretching, and H-O-H deformation combination. This supports the findings of our study and suggests that these peaks are likely to have a vital influence on the performance of the model.

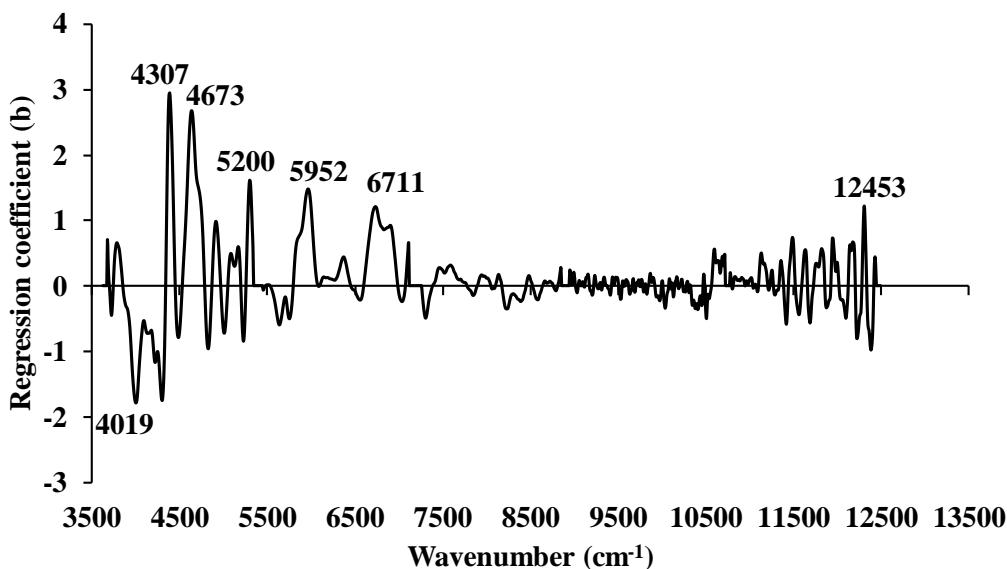


Figure 11. The regression coefficient for the wt.% of N of grounded biomass using the multi-preprocessing PLSR 3 range method.

3.4. Comparison with previous work

Although various studies have been conducted on the development of models for the evaluation of HHV and ultimate analysis parameters using NIRS with a similar wavenumber range and reference mean value combined with chemometrics, no research or reports on fast-growing trees and agricultural residues of Nepalese biomass with ten different biomass varieties using NIRS have been reported to date.

In a previous study, Nakawajana et al. [50] evaluated the HHV of grounded cassava rhizome using PLSR and achieved an R^2 of 0.90. Similarly, Nakawajana et al. [25], Posom et al. [3], Zhang et al. [51], and Posom et al. [5] developed PLSR models for rick husk, grounded bamboo, sorghum biomass, and *Leucaena leucocephala* pellets, respectively, with R^2 0.79, 0.92, 0.96, and 0.96. All the studies used NIRS scanning of biomass on diffuse reflectance mode. However, the GA-PLSR model in this study outperformed previous research by using NIRS scanning of biomass in transreflectance mode for evaluating HHV.

The PLSR-based models developed from multi-preprocessing 5-range methods for ultimate analysis showed better performance in evaluating oxygen content compared to the PLS model developed by Jetsada et al [49] for bamboo, which had R^2_P values of 0.52 for oxygen. However, the results of this study for the evaluation of C, N, and H contents were lower, with Jetsada et al. [49] for bamboo showing R^2_P values of 0.80 for C, 0.85 for H, and 0.97 for N. Similarly, the models developed by Zhang et al. [51] for sorghum biomass with R^2_P values of 0.96 for wt.% of C, 0.87 for wt.% of H, 0.86 for wt.% of N, and 0.83 for wt.% of O and by Huang et al. [10] for straw with R^2_P values of 0.97 for wt.% of C, 0.77 for wt.% of H, and 0.87 for wt.% of N showed better results than the PLSR-based model in this study. Nhuchhen [52] predicted the ultimate parameters of torrified biomass with respect to proximate analysis, resulting in R^2 values of 0.83 for wt.% of C, 0.70 for wt.% of H, and 0.84 for wt.% of O, respectively. The proposed model in this study showed better performance for H and O, but the performance of C content in the grounded biomass could be improved.

In general, having a sufficient number of homogenous biomass samples with a wider range of reference values and low SEL from bomb calorimeter and CHNS/O elemental analyzer could have played a catalytic role in achieving higher model performance when evaluating HHV, N, and O. However, lower model performance for evaluating C and H content may be due to a lower number of relevant variables or the selected variables in the calibration set not having a strong correlation with C and H content in biomass. To enhance the model performance for evaluating C and H content, the number of representative samples with high concentration of C and H should be increased, and possible contamination during sample preparation should be avoided. Additionally, the ambient environment of the laboratory should be properly controlled, and possible NIR radiation leakage during sample scanning should be rechecked. Outliers should be addressed properly, instrumental and analysis errors should be monitored correctly, or alternative modeling techniques should be considered for accurate evaluation.

Based on comparison with previous studies, this research provides strong evidence that the model's performance can be enhanced by conducting NIRS scanning of ground biomass in transreflectance mode rather than diffuse reflectance mode. To update the model for robust application, the number of ground biomass samples must be increased and validated using unknown samples.

4.0 Conclusion

PLSR-based models were developed and compared using NIRS to evaluate HHV, and ultimate analysis, i.e. wt.% of C, H, N, and O content in the grounded biomass in transreflectance mode, was employed for assessing the biomass properties for energy usage. The model with the optimum performance was selected based on the parameters R^2_C , RMSEP, R^2_P , RMSEP, RPD, and bias. The models for HHV, N, and O are suitable for most applications, including research, while the model for wt.% of C and wt.% of H was only fair and usable for rough screening. The performance of fair models could be improved by incorporating more representative samples collected from various geographical locations in Nepal, considering the wide statistical range of the reference values.

This study showed that multi-preprocessing 5-range method, a new approach to spectral preprocessing for PLSR model development, improved model accuracy compared to the traditional method of preprocessing NIR spectra across the entire wavenumber range with a single process. Therefore, this research provides foundation in NIRS, indicating that preprocessing the entire wavenumber range with various preprocessing techniques could enhance model accuracy. The recommended models can serve as a reliable and non-destructive alternative method for rapidly assessing biomass properties for energy usage employing NIRS. However, to create a robust model, it is necessary to expand the model with data from various samples and validate it with unknown samples. Adopting these models could significantly reduce the economic gap between biomass traders for energy usage and other applications. Furthermore, the research outcomes could guide academic and research institutions and policy-making think tanks in planning for the proper identification, management, and utilization of bio-resources to meet future energy demand and supply.

Author Contributions: B.S.; conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing the original draft, writing – review & editing, visualization. J.P.; conceptualization, software, formal analysis, data curation, writing – review & editing, supervision. B.P.S.; writing – review & editing, conceptualization and validation. P.S.; conceptualization, data curation, writing – review & editing, supervision, project administration, funding acquisition.

Funding: This research was funded by the King Mongkut's Institute of Technology Ladkrabang, Thailand (KMITL doctoral scholarship KDS 2020/52).

Data Availability Statement: The data will be made available upon request from the corresponding author.

Acknowledgments: The authors express their gratitude to the Near-Infrared Spectroscopy Research Center for Agricultural Product and Food (www.nirsresearch.com), Department of Agricultural Engineering, School of Engineering at King Mongkut's Institute of Technology Ladkrabang, Thailand, for providing instruments, funding, and space for the experiment. The Department of Agricultural Engineering, Khon Kaen University, Thailand is also acknowledged for providing technical support for this research. Additionally, the authors extend their appreciation to the Department of Mechanical Engineering, Kathmandu University, Nepal, for their assistance in collecting biomass samples from various locations in Nepal.

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

References

1. Zhang, Y.; Wang, H.; Sun, X.; Wang, Y.; Liu, Z., Separation and Characterization of Biomass Components (Cellulose, Hemicellulose, and Lignin) from Corn Stalk. *BioResources* **2021**, 16, (4).
2. Díez, D.; Urueña, A.; Piñero, R.; Barrio, A.; Tamminen, T., Determination of hemicellulose, cellulose, and lignin content in different types of biomasses by thermogravimetric analysis and pseudocomponent kinetic model (TGA-PKM method). *Processes* **2020**, 8, (9), 1048. <https://doi.org/10.3390/pr8091048>.
3. Posom, J.; Sirisomboon, P., Evaluation of the higher heating value, volatile matter, fixed carbon and ash content of ground bamboo using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* **2017**, 25, (5), 301-310. 10.1177/0967033517728733.
4. Mierzwa-Hersztek, M.; Gondek, K.; Jewiarz, M.; Dziedzic, K., Assessment of energy parameters of biomass and biochars, leachability of heavy metals and phytotoxicity of their ashes. *Journal of Material Cycles and Waste Management* **2019**, 21, (4), 786-800. 10.1007/s10163-019-00832-6.
5. Posom, J.; Shrestha, A.; Saechua, W.; Sirisomboon, P., Rapid non-destructive evaluation of moisture content and higher heating value of Leucaena leucocephala pellets using near infrared spectroscopy. *Energy* **2016**, 107, 464-472. <https://doi.org/10.1016/j.energy.2016.04.041>.
6. Demirbas, A., Hazardous Emissions from Combustion of Biomass. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* **2007**, 30, (2), 170-178. <https://doi.org/10.1080/00908310600712406>.
7. Onifade, M.; Lawal, A. I.; Aladejare, A. E.; Bada, S.; Idris, M. A., Prediction of gross calorific value of solid fuels from their proximate analysis using soft computing and regression analysis. *International*

Journal of Coal Preparation and Utilization 2019, 42, (4), 1170-1184. <https://doi.org/10.1080/19392699.2019.1695605>.

8. Sheng, C.; Azevedo, J. L. T., Estimating the higher heating value of biomass fuels from basic analysis data. *Biomass and Bioenergy* 2005, 28, (5), 499-507. <https://doi.org/10.1016/j.biombioe.2004.11.008>.

9. El-Sayed, S. A.; Mostafa, M., Pyrolysis characteristics and kinetic parameters determination of biomass fuel powders by differential thermal gravimetric analysis (TGA/DTG). *Energy conversion and management* 2014, 85, 165-172. <https://doi.org/10.1016/j.enconman.2014.05.068>.

10. Huang, C.; Han, L.; Yang, Z.; Liu, X., Ultimate analysis and heating value prediction of straw by near infrared spectroscopy. *Waste Manag* 2009, 29, (6), 1793-7. <https://doi.org/10.1016/j.wasman.2008.11.027>.

11. Adnan, A.; Horsten, D. V.; Pawelzik, E.; Morlein, A. D., Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy. *Foods* 2017, 6, (5). <https://doi.org/10.3390/foods6050038>.

12. Roger, J.-M.; Mallet, A.; Marini, F., Preprocessing NIR Spectra for Aquaphotomics. *Molecules* 2022, 27, (20), 6795. <https://doi.org/10.3390/molecules27206795>.

13. Maraphum, K.; Saengprachatanarug, K.; Wongpichet, S.; Phuphuphud, A.; Posom, J., Achieving robustness across different ages and cultivars for an NIRS-PLSR model of fresh cassava root starch and dry matter content. *Computers and Electronics in Agriculture* 2022, 196. <https://doi.org/10.1016/j.compag.2022.106872>.

14. Posom, J.; Phuphuphud, A.; Saengprachatanarug, K.; Maraphum, K.; Saijan, S.; Pongkan, K.; Srimai, K., Real-time measuring energy characteristics of cane bagasse using NIR spectroscopy. *Sensing and Bio-Sensing Research* 2022, 38. <https://doi.org/10.1016/j.sbsr.2022.100519>.

15. Phuphuphud, A.; Saengprachatanarug, K.; Posom, J.; Taira, E.; Panduangnate, L., Prediction and Classification of Energy Content in Growing Cane Stalks for Breeding Programmes Using Visible and Shortwave Near Infrared. *Sugar Tech* 2022, 24, 1497-1509. <https://doi.org/10.1007/s12355-021-01075-2>.

16. Posom, J.; Sirisomboon, P., Evaluation of the higher heating value, volatile matter, fixed carbon and ash content of ground bamboo using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* 2017, 25, (5), 301-310.

17. Posom, J.; Saechua, W. In *Prediction of elemental components of ground bamboo using micro-NIR spectrometer*, IOP Conference Series: Earth and Environmental Science, 2019; IOP Publishing: 2019; p 012063.

18. Skvaril, J.; Kyprianidis, K. G.; Dahlquist, E., Applications of near-infrared spectroscopy (NIRS) in biomass energy conversion processes: A review. *Applied Spectroscopy Reviews* 2017, 52, 675-728. <https://doi.org/10.1080/05704928.2017.1289471>.

19. Zhang, K.; Zhou, L.; Brady, M.; Xu, F.; Yu, Jianming; Wang, D., Fast analysis of high heating value and elemental compositions of sorghum biomass using near-infrared spectroscopy. *Energy* 2017, 118, 1353-1360. <https://doi.org/10.1016/j.energy.2016.11.015>.

20. Xue, J.; Yang, Z.; Han, L.; Chen, L., Study of the influence of NIRS acquisition parameters on the spectral repeatability for on-line measurement of crop straw fuel properties. *Fuel* 2014, 117, 1027-1033. <https://doi.org/10.1016/j.fuel.2013.10.017>.

21. Jiao, Y.; Li, Z.; Chen, X.; Fei, S., Preprocessing methods for near-infrared spectrum calibration. *Journal of Chemometrics* 2020, 34, (11). <https://doi.org/10.1002/cem.3306>.

22. Posom, J.; Maraphum, K.; Phuphuphud, A., Rapid Evaluation of Biomass Properties Used for Energy Purposes Using Near-Infrared Spectroscopy. In *Renewable Energy-Technologies and Applications*, IntechOpen: 2020.

23. Yun, Y.-H.; Li, H.-D.; Deng, B.-C.; Cao, D.-S., An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends in Analytical Chemistry* 2019, 113, 102-115. <https://doi.org/10.1016/j.trac.2019.01.018>.

24. Broad, N.; Graham, P.; Hailey, P.; Hardy, A.; Holland, S.; Hughes, S.; Lee, D.; Prebble, K.; Salton, N.; Warren, P., Guidelines for the development and validation of near-infrared spectroscopic methods in the pharmaceutical industry. *Handbook of vibrational spectroscopy* 2002, 5, 3590-3610.

25. Nakawajana, N.; Posom, J.; Paeoui, J., The prediction of higher heating value, lower heating value and ash content of rice husk using FT-NIR spectroscopy. *Engineering Journal* 2018, 22, (5), 45-56. <https://doi.org/10.4186/ej.2018.22.5.45>.

26. Assis, C.; Ramos, R. S.; Silva, L. A.; Kist, V.; Barbosa, M. H. P.; Teofilo, R. F., Prediction of Lignin Content in Different Parts of Sugarcane Using Near-Infrared Spectroscopy (NIR), Ordered Predictors Selection (OPS), and Partial Least Squares (PLS). *Appl Spectrosc* 2017, 71, (8), 2001-2012. <https://doi.org/10.1177/0003702817704147>.

27. Li, Z.; Song, J.; Ma, Y.; Yu, Y.; He, X.; Guo, Y.; Dou, J.; Dong, H., Identification of aged-rice adulteration based on near-infrared spectroscopy combined with partial least squares regression and characteristic wavelength variables. *Food Chemistry*: X 2023, 17, 100539. <https://doi.org/10.1016/j.fochx.2022.100539>.

28. Shetty, N.; Gislum, R., Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR. *Field Crops Research* 2011, 120, (1), 31-37. <https://doi.org/10.1016/j.fcr.2010.08.008>.

29. Conzen, J., Multivariate Calibration: A practical guide for developing methods in the quantitative analytical chemistry. *Ettlingen, Germany: BrukerOptik GmbH* 2006.

30. Pitak, L.; Sirisomboon, P.; Saengprachatanarug, K.; Wongpichet, S.; Posom, J., Rapid elemental composition measurement of commercial pellets using line-scan hyperspectral imaging analysis. *Energy* 2021, 220. <https://doi.org/10.1016/j.energy.2020.119698>.

31. Xia, Z.; Zhang, C.; Weng, H.; Nie, P.; He, Y., Sensitive wavelengths selection in identification of Ophiopogon japonicus based on near-infrared hyperspectral imaging technology. *International journal of analytical chemistry* 2017, 2017. <https://doi.org/10.1155/2017/6018769>.

32. Zhang, C.; Liu, F.; Kong, W.; Zhang, H.; He, Y., Fast identification of watermelon seed variety using near infrared hyperspectral imaging technology. *Transactions of the Chinese Society of Agricultural Engineering* 2013, 29, (20), 270-277.

33. Liu, D.; Sun, D.-W.; Zeng, X.-A., Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food and Bioprocess Technology* 2014, 7, (2), 307-323.

34. Santos-Rufo, A.; Mesas-Carrascosa, F.-J.; García-Ferrer, A.; Meroño-Larraíva, J. E., Wavelength selection method based on partial least square from hyperspectral unmanned aerial vehicle orthomosaic of irrigated olive orchards. *Remote Sensing* 2020, 12, (20), 3426. <https://doi.org/10.3390/rs12203426>.

35. Maraphum, K.; Ounkaew, A.; Kasemsiri, P.; Hiziroglu, S.; Posom, J., Wavelengths selection based on genetic algorithm (GA) and successive projections algorithms (SPA) combine with PLS regression for determination the soluble solids content in Nam-DokMai mangoes based on near infrared spectroscopy. *Engineering and Applied Science Research* 2022, 49, (1), 119-126. <https://ph01.tci-thaijo.org/index.php/easr/article/view/245217>.

36. Jiang, Q.; Chen, Y.; Guo, L.; Fei, T.; Qi, K., Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS-NIR spectroscopy. *Remote Sensing* 2016, 8, (9), 755. <https://doi.org/10.3390/rs8090755>.

37. Williams, P.; Manley, M.; Antoniszyn, J., *Near infrared technology: getting the best out of light*. African Sun Media: 2019.

38. Zornoza, R.; Guerrero, C.; Mataix-Solera, J.; Scow, K. M.; Arcenegui, V.; Mataix-Beneyto, J., Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biology and Biochemistry* 2008, 40, (7), 1923-1930. <https://doi.org/10.1016/j.soilbio.2008.04.003>.

39. Alzagameem, A.; Bergs, M.; Do, X. T.; Klein, S. E.; Rumpf, J.; Larkins, M.; Monakhova, Y.; Pude, R.; Schulze, M., Low-input crops as lignocellulosic feedstock for second-generation biorefineries and the potential of chemometrics in biomass quality control. *Applied Sciences* 2019, 9, (11), 2252. <https://doi.org/10.3390/app9112252>.

40. Den, W.; Sharma, V. K.; Lee, M.; Nadadur, G.; Varma, R. S., Lignocellulosic biomass transformations via greener oxidative pretreatment processes: access to energy and value-added chemicals. *Frontiers in chemistry* 2018, 6, 141.

41. Weyer, L., *Practical guide to interpretive near-infrared spectroscopy*. CRC press: 2007.

42. Hasan, M.; Haseli, Y.; Karadogan, E., Correlations to predict elemental compositions and heating value of torrefied biomass. *Energies* 2018, 11, (9), 2443. <https://doi.org/10.3390/en11092443>.

43. Zoghiami, A.; Paës, G., Lignocellulosic biomass: understanding recalcitrance and predicting hydrolysis. *Frontiers in chemistry* 2019, 7, 874. <https://doi.org/10.3389/fchem.2019.00874>.

44. Ge, X.; Chang, C.; Zhang, L.; Cui, S.; Luo, X.; Hu, S.; Qin, Y.; Li, Y., Conversion of lignocellulosic biomass into platform chemicals for biobased polyurethane application. In *Advances in bioenergy*, Elsevier: 2018; Vol. 3, pp 161-213.

45. Sirisomboon, P.; Funke, A.; Posom, J., Improvement of proximate data and calorific value assessment of bamboo through near infrared wood chips acquisition. *Renewable Energy* 2020, 147, 1921-1931. <https://doi.org/10.1016/j.renene.2019.09.128>.

46. Lestander, T. A.; Rhén, C., Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. *Analyst* 2005, 130, (8), 1182-1189. <https://doi.org/10.1039/B500103J>.

47. Han, K.; Gao, J.; Qi, J., The study of sulphur retention characteristics of biomass briquettes during combustion. *Energy* 2019, 186, 115788. <https://doi.org/10.1016/j.energy.2019.07.118>.

48. Cagnon, B.; Py, X.; Guillot, A.; Stoeckli, F.; Chambat, G., Contributions of hemicellulose, cellulose and lignin to the mass and the porous properties of chars and steam activated carbons from various lignocellulosic precursors. *Bioresource Technology* **2009**, *100*, (1), 292-298. <https://doi.org/10.1016/j.biortech.2008.06.009>.

49. Posom, J.; Sirisomboon, P., Evaluation of lower heating value and elemental composition of bamboo using near infrared spectroscopy. *Energy* **2017**, *121*, 147-158. <https://doi.org/10.1016/j.energy.2017.01.020>.

50. Nakawajana, N.; Posom, J. In *Comparison of analytical ability of pls and svm algorithm in estimation of moisture content, higher heating value, and lower heating value of cassava rhizome ground using FT-NIR spectroscopy*, IOP Conference Series: Earth and Environmental Science, 2019; IOP Publishing: 2019; p 012032.

51. Zhang, K.; Zhou, L.; Brady, M.; Xu, F.; Yu, J.; Wang, D., Fast analysis of high heating value and elemental compositions of sorghum biomass using near-infrared spectroscopy. *Energy* **2017**, *118*, 1353-1360. <https://doi.org/10.1016/j.energy.2016.11.015>.

52. Nhuchhen, D. R., Prediction of carbon, hydrogen, and oxygen compositions of raw and torrefied biomass using proximate analysis. *Fuel* **2016**, *180*, 348-356. <https://doi.org/10.1016/j.fuel.2016.04.058>.