

CHARACTERIZATION OF BIOMASS IN NEPAL USED FOR ENERGY PURPOSE
USING NEAR INFRARED SPECTROSCOPY



BIJENDRA SHRESTHA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF ENGINEERING IN FOOD AND AGRICULTURAL INTELLIGENCE ENGINEERING
SCHOOL OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2024
KMITL-2024-EN-D-108-275

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



COPYRIGHT 2024

SCHOOL OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

หัวข้อวิทยานิพนธ์	การกำหนดลักษณะของชีวมวลในประเทศเนาลซึ่งใช้เพื่อวัตถุประสงค์ด้านพลังงานโดยใช้สเปกโทรสโกปีอินฟราเรดย่านใกล้
นักศึกษา	Bijendra Shrestha
รหัสประจำตัว	63601254
ปริญญา	วิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชา	วิศวกรรมอาหารและเกษตรอัจฉริยะ
พ.ศ.	2567
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ศาสตราจารย์ ดร. ปานมนัส ศิริสมบุญ รองศาสตราจารย์ ดร. เจษฎา โพธิ์สม Prof. Dr. Bim Prasad Shrestha

บทคัดย่อ

ชีวมวลเป็นทรัพยากรหมุนเวียนที่เป็นกลางต่อคาร์บอนและหาได้ง่าย นำเสนอทางเลือกซึ่งยั่งยืนแทนเชื้อเพลิงฟอสซิล มีส่วนช่วยในการประหยัดพลังงานและรักษาสิ่งแวดล้อม โดยคิดเป็นสองในสามของการใช้พลังงานหมุนเวียนทั่วโลก โดยมีการใช้งานตั้งแต่การปรุงอาหารในครัวเรือนและการทำความร้อน ไปจนถึงการผลิตไฟฟ้า การขนส่ง และการใช้ในอุตสาหกรรม แม้จะมีการใช้อย่างแพร่หลาย แต่การใช้ชีวมวลแบบดั้งเดิมแบบไม่มีประสิทธิภาพยังคงแพร่หลายทั้งในครัวเรือนและในโรงงานอุตสาหกรรม นอกจากนี้ โดยทั่วไปชีวมวลจะมีการซื้อขายแลกเปลี่ยนโดยปริมาตรหรือโดยมวลมากกว่าโดยค่าสมบัติพลังงานแท้จริงของชีวมวล ด้วยเหตุนี้ การประเมินสมบัติชีวมวลที่รวดเร็ว เชื่อถือได้ และไม่ทำลาย ซึ่งปรับให้เหมาะกับลักษณะเฉพาะของพลังงาน ตลอดจนการจัดการและการใช้อย่างมีประสิทธิภาพผ่านเทคโนโลยีพลังงานที่เหมาะสม จึงมีความสำคัญอย่างยิ่งต่อการตอบสนองความต้องการพลังงานของภาคชนบทและอุตสาหกรรม ตลอดจนผู้ค้าอย่างยั่งยืน

วิทยานิพนธ์นี้จึงได้พัฒนา เปรียบเทียบ และเลือกแบบจำลองที่มีสมรรถนะการทำนายดีที่สุดโดยอาศัยการถดถอยกำลังสองน้อยที่สุดบางส่วน (PLSR) สำหรับการทำนายสมบัติพลังงานของต้นไม้โตเร็วและเศษเหลือทางการเกษตรในประเทศเนปาล การศึกษานี้ใช้สเปกโทรสโกปีฟูเรียร์ทรานสฟอร์มอินฟราเรดย่านใกล้ (NIRS) ในช่วงเลขคลื่น 3,595 ถึง 12,489 cm^{-1} เป็นวิธีทางเลือก สมบัติพลังงานเป้าหมาย ได้แก่ ค่าความร้อนที่สูงขึ้น

(HHV) ที่ได้รับจากบอมบ์แคลอรีมิเตอร์ พารามิเตอร์การวิเคราะห์ขั้นสูงสุด คือ คาร์บอน (C) ไฮโดรเจน (H) ไนโตรเจน (N) และออกซิเจน (O) ในหน่วยเปอร์เซ็นต์โดยน้ำหนัก (wt.%) วัดโดยใช้เครื่องวิเคราะห์ห้องประกอบ CHNO/S และพารามิเตอร์การวิเคราะห์โดยประมาณ ได้แก่ ปริมาณความชื้น (MC) สารระเหย (VM) คาร์บอนคงที่ (FC) และปริมาณเถ้า (A) เป็นเปอร์เซ็นต์ (%) โดยใช้เครื่องวิเคราะห์เทอร์โมกราวิเมตริก นอกจากนี้ การศึกษายังมีจุดมุ่งหมายเพื่อคาดการณ์ลักษณะการเผาไหม้ เช่น ดัชนีการจุดระเบิด (D_i) ดัชนีการเผาไหม้ (D_f) ดัชนีประสิทธิภาพการเผาไหม้ (S_i) และดัชนีความสามารถในการติดไฟ (C_i) โดยใช้ข้อมูลจากเครื่องวิเคราะห์เทอร์โมกราวิเมตริก การศึกษานี้ยังพยายามที่จะแสดงให้เห็นว่าแบบจำลองที่พัฒนาขึ้นนั้นพร้อมสำหรับการใช้งานทางอุตสาหกรรม โดยเฉพาะอย่างยิ่งในแง่ของ HHV

แบบจำลอง PLSR ต่างๆ ได้รับการพัฒนาโดยใช้แนวทางการจัดการสเปกตรัมเบื้องต้นแบบต่างๆ ได้แก่ ไม่มีการจัดการสเปกตรัมเบื้องต้น การจัดการสเปกตรัมเบื้องต้นแบบดั้งเดิม การจัดการสเปกตรัมเบื้องต้นโดยแบ่งช่วง กรณีนี้แบ่งเป็น 5 ช่วง และ 3 ช่วง มีการใช้อัลกอริธึมทางพันธุกรรม (GA) และอัลกอริธึมการฉายภาพต่อเนื่อง (SPA) ในการคัดเลือกเลขคลื่นอรรถลักษณะ มีการใช้แนวทางการจัดการสเปกตรัมเบื้องต้นหลายอย่างร่วมกันแบบใหม่ โดยสมมติว่าการใช้วิธีการจัดการสเปกตรัมเบื้องต้นหลายวิธีตลอดช่วงเลขคลื่นอินฟราเรดย่านใกล้ ทั้งหมดจะช่วยปรับปรุงสมรรถนะการทำนายของแบบจำลองได้ นอกจากนี้ การศึกษายังได้วิเคราะห์ผลกระทบของชีวมวลประเภทต่างๆ โดยเฉพาะไม้และพันธุ์ที่ไม่ใช่ไม้ ต่อการพัฒนาแบบจำลองผ่านการวิเคราะห์แผนภูมิการกระจายการวิเคราะห์นี้ถือเป็นครั้งแรกที่พยายามตรวจสอบความพร้อมของแบบจำลองที่พัฒนาขึ้นสำหรับการใช้งานทางอุตสาหกรรม โดยเฉพาะอย่างยิ่งในแง่ของ HHV

ในการศึกษานี้ โมเดลที่ดีที่สุดได้รับเลือกโดยพิจารณาจากค่าสัมประสิทธิ์การพิจารณาสูงสุดในชุดการตรวจสอบแบบจำลอง (R^2_p) รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนาย (RMSEP) ต่ำ และอัตราส่วนการทำนายต่อการเบี่ยงเบน (RPD) สูง แบบจำลองที่มีสมรรถนะการทำนายสูงสุดสำหรับ HHV ได้มาจากชีวมวลบดโดยใช้ GA-PLSR ที่มีอนุพันธ์อันดับหนึ่ง โดยได้ R^2_p ที่ 0.9574, RMSEP ที่ 170.3282 J/g, RPD ที่ 4.9 ซึ่งบ่งชี้ถึงความเหมาะสมสำหรับการใช้งานที่หลากหลาย รวมถึงการวิจัย สำหรับแบบจำลองทำนายพารามิเตอร์การวิเคราะห์ขั้นสูงสุด แบบจำลองที่ดีที่สุดสำหรับ C และ H ได้รับการพัฒนาจากชีวมวลบดโดยใช้ GA-PLSR โดยได้ค่า R^2_p เท่ากับ 0.7217 และ 0.7678, RMSEP เท่ากับ 170.3282 wt.% และ 0.1434 wt.% และ RPD เท่ากับ 1.9 และ 2.1 ตามลำดับ อย่างไรก็ตาม แบบจำลองเหล่านี้มีประโยชน์สำหรับการคัดกรองแบบคร่าวๆ เท่านั้น แบบจำลอง N ซึ่งพัฒนาโดยใช้ MP-PLSR ด้วยเทคนิค 5 ช่วงจากชีวมวลบด แสดง R^2_p เท่ากับ 0.8410, RMSEP เท่ากับ 0.0973 wt.% และ RPD เท่ากับ 2.7 ทำให้เชื่อถือได้สำหรับการใช้งานในวงกว้าง แต่ควรใช้ด้วยความ

ระมัดระวัง ในทางตรงกันข้าม แบบจำลองสำหรับ O จากชีวมวลแบบสับโดยใช้เทคนิค MP-PLSR เดียวกันให้ค่า R^2_p , RMSEP และ RPD เท่ากับ 0.7150, 1.3088 wt.% และ 1.9 ตามลำดับ จึงเหมาะสำหรับการคัดกรองเบื้องต้นเท่านั้น

สำหรับพารามิเตอร์การวิเคราะห์โดยประมาณ แบบจำลองสำหรับ MC และ FC ในชีวมวลแบบสับ แสดงสมรรถนะที่น่าพอใจ โดยสามารถนำไปใช้งานต่างๆได้แต่ต้องระมัดระวัง รวมถึงการวิจัยด้วย แบบจำลองที่ดีที่สุดสำหรับ MC และ FC สร้างโดยใช้ GA-PLSR พร้อมอนุพันธ์อันดับสองและ การใช้เลขคลื่นทั้งหมด (Full-PLSR) พร้อมออฟเซตคงที่ได้ค่า R^2_p สูงถึง 0.8654 และ 0.8773 ค่า RMSEP ต่ำ 0.85% และ 2.12% และ ค่า RPD สูงถึง 2.9 และ 3.0 ตามลำดับ ซึ่งบ่งบอกถึงความสามารถในการนำไปใช้งานได้ ในทางตรงกันข้าม แบบจำลอง VM ที่ดีที่สุดได้มาจากการใช้ MP-PLSR ของชีวมวลสับ พร้อมด้วยการแบ่งสเปกตรัมเป็น 5 ช่วง ซึ่งให้ผล R^2_p เท่ากับ 0.7937, RMSEP เท่ากับ 2.47% และ RPD เท่ากับ 2.2 ทำให้เหมาะสำหรับการคัดกรองหยาบเท่านั้น ในทำนองเดียวกัน แบบจำลองที่ดีที่สุดสำหรับปริมาณเถ้าได้มาจากชีวมวลสดโดยใช้ Full-PLSR ที่มีอนุพันธ์อันดับหนึ่ง ส่งผลให้ R^2_p เท่ากับ 0.7983, RMSEP เท่ากับ 1.26% และ RPD เท่ากับ 2.2 ซึ่งบ่งชี้ถึงความเหมาะสมสำหรับการคัดกรองคร่าวๆ เช่นกัน

สำหรับดัชนีสมรรถนะการเผาไหม้ (D_i , D_f , S_i และ C_i) แบบจำลองที่ดีที่สุดทั้งหมดได้มาจากการใช้มวลชีวภาพสด แม้ว่าจะมีเทคนิคในการจัดการสเปกตรัมเบื้องต้นที่แตกต่างกันก็ตาม แบบจำลองสำหรับ D_i ซึ่งพัฒนาโดยใช้ MP-PLSR ด้วยวิธีการจัดการสเปกตรัมเบื้องต้นโดยแบ่งสเปกตรัมเป็น 5 ช่วง ได้ R^2_p เท่ากับ 0.6782, RMSEP เท่ากับ 0.3879 wt.% min^{-4} และ RPD เท่ากับ 1.8 ซึ่งเหมาะสำหรับการคัดกรองแบบคร่าวๆ เท่านั้น ในทางตรงกันข้าม โมเดล D_f ซึ่งใช้ GA-PLSR กับอนุพันธ์อันดับหนึ่ง ได้ R^2_p ที่ 0.8426, RMSEP ที่ 0.4968 wt.% min^{-4} และ RPD ที่ 2.5 ทำให้เหมาะสำหรับการใช้งานในวงกว้าง รวมถึงการวิจัย สำหรับ S_i โมเดลที่เหมาะสมที่สุดจะใช้ Full-PLSR กับ MSC โดยได้ R^2_p เท่ากับ 0.8808, RMSEP เท่ากับ 0.1566 wt.% $^2 \text{min}^{-2} \text{ } ^\circ\text{C}^{-3}$ และ RPD เท่ากับ 3.1 ทำให้เหมาะสำหรับการใช้งานส่วนใหญ่ รวมถึงการวิจัย. สุดท้าย สำหรับ C_i ได้แบบจำลองที่เหมาะสมที่สุดโดยใช้ MP-PLSR ด้วยการจัดการสเปกตรัมเบื้องต้นแบบแบ่งสเปกตรัมเป็น 5 ช่วง ส่งผลให้ R^2_p เท่ากับ 0.7204, RMSEP เท่ากับ 0.3610 wt.% $\text{min}^{-1} \text{ } ^\circ\text{C}^{-2}$ และ RPD เท่ากับ 1.8 ทำให้แบบจำลองใช้ได้กับการตรวจคัดกรองแบบหยาบเท่านั้น

จากสมรรถนะของแบบจำลอง แบบจำลองที่ใช้ PLSR ที่สามารถนำไปใช้ได้ ได้มาจากชีวมวลสด ได้แก่ แบบจำลอง HHV (J/g), wt.% N, D_f , S_i และจากชีวมวลสับ ได้แก่ แบบจำลอง MC และ FC ให้ผลลัพธ์ที่รวดเร็ว เชื่อถือได้ และไม่มีการทำลายตัวอย่าง สำหรับการประเมินสมบัติชีวมวลเพื่อการใช้งานโดยใช้ NIRS นี้เหมาะ

สำหรับใช้ในการซื้อขายชีวมวลและพร้อมสำหรับการใช้งานทางอุตสาหกรรม อย่างไรก็ตาม เพื่อพัฒนาโมเดลระดับโลกที่ครอบคลุมมากขึ้น จำเป็นต้องมีการปรับปรุงอย่างต่อเนื่อง สิ่งนี้เกี่ยวข้องกับการรวบรวมตัวอย่างชีวมวลที่เป็นตัวแทนจำนวนมากขึ้น การตรวจสอบความถูกต้องของการทำนายของแบบจำลองอย่างต่อเนื่องกับตัวอย่างที่ไม่รู้จัก และการสำรวจและศึกษาอัลกอริธึมการเรียนรู้ของเครื่องที่เหมาะสมยิ่งขึ้นเพื่อให้แน่ใจว่าสามารถคาดการณ์สมบัติพลังงานชีวมวลได้อย่างแม่นยำและเที่ยงในทุสภาพารามิเตอร์

คำสำคัญ: สเปกโทรสโกปีอินฟราเรดย่านใกล้, ชีวมวล (สับและบด), การถดถอยกำลังสองน้อยที่สุดบางส่วน, ค่าความร้อนสูง, พารามิเตอร์การวิเคราะห์ขั้นสูงสุด, พารามิเตอร์การวิเคราะห์โดยประมาณ, ดัชนีสมรรถนะการเผาไหม้



Thesis	CHARACTERIZATION OF BIOMASS IN NEPAL USED FOR ENERGY PURPOSE USING NEAR INFRARED SPECTROSCOPY
Student	BIJENDRA SHRESTHA
Student ID.	63601254
Degree	Doctor of Engineering
Program	Food and Agricultural Intelligence Engineering
Year	2024
Thesis Advisors	Prof. Dr. Panmanas Sirisomboon Associate Prof. Dr. Jetsada Posom Prof. Dr. Bim Prasad Shrestha

ABSTRACT

Biomass, a carbon-neutral and readily available renewable resource, offers a sustainable alternative to fossil fuels, contributing to energy savings and environmental protection. It accounts for two-thirds of global renewable energy utilization, with applications ranging from household cooking and heating to electricity generation, transportation, and industrial use. Despite its widespread adoption, inefficient consumption of traditional biomass remains prevalent in both household and industrial settings. Additionally, biomass is commonly traded by volume or mass rather than its actual energy properties. Consequently, rapid, reliable, and non-destructive evaluation of biomass, tailored to its specific energy characteristics, and effective management and utilization through appropriate energy technologies are crucial for sustainably meeting the energy demands of rural and industrial sectors, as well as traders.

This thesis develops, compares, and selects the best-performing model based on partial least squares regression (PLSR) for predicting the energy properties of fast-growing trees and agricultural residues in Nepal. The study employs Fourier Transform (FT) near-infrared spectroscopy (NIRS) within a wavenumber range of 3595 to 12,489 cm^{-1} as an alternative method. The target energy properties include higher heating value (HHV) obtained from a bomb

calorimeter, ultimate analysis parameters (carbon (C), hydrogen (H), nitrogen (N), and oxygen (O) in weight percentages (wt.%) from a CHNO/S elemental analyzer, and proximate analysis parameters (moisture content (MC), volatile matter (VM), fixed carbon (FC), and ash content (A) in percentage (%)) from a thermogravimetric analyzer. Additionally, the study aims to predict combustion characteristics, such as the ignition index (D_i), burnout index (D_f), combustion performance index (S_i), and flammability index (C_i), using data from a thermogravimetric analyzer. This study also seeks to demonstrate that the developed models are ready for industrial application, particularly in terms of HHV.

Different PLSR models were developed using various preprocessing approaches: no preprocessing, traditional preprocessing, multi-preprocessing in 5-range and 3-range techniques. A genetic algorithm (GA), and a successive projection algorithm (SPA) were used for featured wavenumber selection. A novel multi-preprocessing approach was employed, assuming that using multiple preprocessing methods across the entire NIR wavenumber range would enhance model performance. Furthermore, the study analyzed the impact of different biomass types, specifically wood and non-wood species, on model development through scatter plot analysis. This investigation, the first of its kind, seeks to validate the readiness of the developed models for industrial application, especially in terms of HHV.

In this study, the best model was selected based on its high coefficient of determination in the validation set (R^2_p), low root mean square error of prediction (RMSEP), and the high ratio of prediction to deviation (RPD). The highest-performing model for HHV was derived from ground biomass using GA-PLSR with the first derivative, achieving an R^2_p of 0.9574, RMSEP of 170.3282 J/g, RPD of 4.9, indicating suitability for a wide range of applications including research. For elemental composition, the best models for C and H were developed from ground biomass using GA-PLSR, achieving R^2_p values of 0.7217 and 0.7678, RMSEP of 170.3282 wt.% and 0.1434 wt.%, and RPD of 1.9 and 2.1, respectively; however, these models are primarily useful only for rough screening. The N model, developed using MP-PLSR with a 5 range technique from ground biomass, showed an R^2_p of 0.8410, RMSEP of 0.0973 wt.%, and RPD of 2.7, making it reliable for broader applications but should be used with caution. In contrast, the model for O from chipped biomass using the same MP-PLSR technique yielded R^2_p , RMSEP, and RPD values of 0.7150, 1.3088 wt.%, and 1.9 respectively, making it suitable only for preliminary screening.

For proximate analysis parameters, the models for MC and FC in chip biomass show satisfactory performance, making them cautiously applicable in various applications, including research. The optimal models for MC and FC, constructed using GA-PLSR with the second derivative and Full-PLSR with a constant offset, achieved high R^2_p values of 0.8654 and 0.8773, low RMSEP values of 0.85% and 2.12%, and high RPD values of 2.9 and 3.0, respectively, indicating their applicative capabilities. In contrast, for VM, the best model was derived using MP-PLSR with a 5-range technique from chip biomass, yielding an R^2_p of 0.7937, an RMSEP of 2.47%, and an RPD of 2.2, making it suitable only for rough screening. Similarly, the best model for ash content was obtained from ground biomass using Full-PLSR with the first derivative, resulting in an R^2_p of 0.7983, an RMSEP of 1.26%, and an RPD of 2.2, indicating suitability for rough screening as well.

For combustion performance indices (D_i , D_f , S_i , and C_i), the best models were all derived using ground biomass, although with varying preprocessing techniques. The model for D_i , developed using MP-PLSR with a 5-range technique, achieved an R^2_p of 0.6782, an RMSEP of 0.3879 wt.%·min⁻⁴, and an RPD of 1.8, suitable only for rough screening. In contrast, the D_f model, utilizing GA-PLSR with the first derivative, secured an R^2_p of 0.8426, an RMSEP of 0.4968 wt.%·min⁻⁴, and an RPD of 2.5, making it appropriate for broader applications, including research. For S_i , the optimal model used Full-PLSR with MSC, achieving an R^2_p of 0.8808, an RMSEP of 0.1566 wt.%²·min⁻²·°C⁻³, and an RPD of 3.1, rendering it suitable for most applications, including research. Lastly, for C_i , the optimal model was obtained using MP-PLSR with a 5-range technique, resulting in an R^2_p of 0.7204, an RMSEP of 0.3610 wt.%·min⁻¹·°C², and an RPD of 1.8, making it applicable only for rough screening.

Based on model performance, the selected PLSR-based model—derived from ground biomass for HHV (J/g), wt.% N, D_f , S_i , and from chip biomass for MC and FC—offers a rapid, reliable, and non-destructive alternative method for assessing biomass properties for energy usage using NIRS. This model is suitable for use in biomass trading and is ready for industrial application. However, to develop a more comprehensive global model, continuous enhancement is necessary. This involves incorporating a larger number of representative biomass samples, consistently validating with unknown samples, and exploring more suitable machine learning algorithms to ensure accurate predictions of biomass energy properties across all parameters.

Key words: near-infrared spectroscopy, biomass (chip and ground), partial least squares regression, higher heating value, ultimate analysis parameters, proximate analysis parameters, combustion performance indices



ACKNOWLEDGEMENT

I express my heartfelt gratitude to my advisors: Prof. Dr. Panmanas Sirisomboon from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand; Associate Prof. Dr. Jetsada Posom from Khon Kaen University, Thailand; and Prof. Dr. Bim Prasad Shrestha from Kathmandu University, Nepal. Their unwavering guidance and continuous motivation have been invaluable throughout my doctoral journey in the Food and Agricultural Intelligence Engineering program.

I am also thankful to Kathmandu University, Nepal, for granting me an educational leave, which enabled me to pursue my doctoral degree at the Department of Agricultural Engineering, KMITL.

Special thanks go to Assistant Prof. Dr. Jiraporn Sripinyowanich Jongyingcharoen, Head of the Department of Agricultural Engineering, Assistant Prof. Dr. Wanput Saechua, also from the Department of Agricultural Engineering at KMITL, and Dr. Thitima Phanomsophon for their invaluable guidance and support during the experimentation and data analysis phases of my research at the Near-Infrared Spectroscopy Research Center for Agricultural Product and Food, KMITL.

Additionally, I would also like to thank Associate Prof. Dr. Pimpen Pornchaloempng from the Department of Food Engineering, School of Engineering, at KMITL; Dr. Axel Funke from the Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology, Germany; and Prof. Dr. Hidayah Ariffin from the Laboratory of Biopolymer and Derivatives, Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, Malaysia, for their valuable suggestions and guidance in my research and manuscript writing for publication.

I would like to thank the members of my thesis defense committee for their valuable feedback and guidance. Their expertise and constructive suggestions played a crucial role in shaping the final version of my thesis.

My stay at KMITL would not have been possible without the strong support and guidance of my dear friends: Dr. Bijay Bahadur Pradhan, Er. Aditya Joshi, Er. Alice Bahadur Shrestha, and Ms. Elle Kingdow Chanachote. Their encouragement and support always infused me with positive energy, significantly contributing to the completion of my doctoral degree.

Lastly, I dedicate all my achievements to date to my beloved family members – my father, Mr. Surendra Gopal Yogal; my mother, Ms. Mahal Sundari Shrestha; my brother, Dr. Rijendra Yogal; my sister-in-law, Dr. Sandhya Khuwaunju; my wife, Ms. Krishnaa Barun; and my sons, Mr. Hridayendra Yogal and Mr. Subiran Yogal.

Bijendra Shrestha



TABLE OF CONTENTS

บทคัดย่อ.....	I
ABSTRACT.....	V
ACKNOWLEDGEMENT.....	IX
TABLE OF CONTENTS.....	XI
LIST OF TABLES.....	XVII
LIST OF FIGURES.....	XX
ABBREVIATIONS.....	XXVI
Chapter 1 Introduction.....	1
1.1 Statement and Significance of the problem.....	1
1.2 Goal and objective.....	4
1.3 Novelty of thesis research.....	5
1.4 Hypothesis to be tested.....	6
1.5 Scope of the study.....	6
1.6 Research mapping.....	7
1.7 Process of the study.....	8
References.....	10
Chapter 2 Literature Review.....	13
2.1 Near Infrared Spectroscopy (NIRS).....	13
2.2 Sample preparation.....	16
2.2.1 Transmission.....	16
2.1.2 Reflection.....	17

2.2.3 Transflection.....	17
2.2.4 Interaction.....	18
2.3 Interferometer or Fourier transform (FT) Near Infrared Spectrometer	18
2.4 Factor effecting spectrum generation	19
2.5 General procedure of NIRS	21
2.6 Recent application of NIRS	22
2.7 Biomass.....	24
2.8 Physical properties	25
2.8.1 Particle size –	25
2.8.2 Density.....	26
2.8.3 Grindability.....	26
2.8.4. Flowability.....	27
2.9 Calorific value.....	27
2.9.1 Bomb calorimeter.....	28
2.10 Proximate analysis.....	30
2.11 Ultimate analysis	32
2.12 Combustion	34
2.13 Thermogravimetric Analyzer.....	36
2.14 Spectroscopy Quantitative Modeling	38
2.14.1 Sample collection covering future sample and its appropriate storage	39
2.14.2. NIR sample scanning and reference data measurement.....	39
2.14.3 Data separation for calibration set and validation set	41

2.14.4. Spectral data preprocessing.....	41
2.14.5 Modelling	44
2.14.6 Validation, performance analysis and model updating	46
References	52
Chapter 3 Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques	60
Abbreviations.....	62
3.1. Introduction	64
3.2. Materials and Methods.....	67
3.2.1. Sample Preparation	69
3.2.2. Spectral Data Collection.....	70
3.3. Reference Analysis.....	71
3.3.1. Higher Heating Value (HHV).....	72
3.3.2. Ultimate Analysis.....	72
3.3.3. Outlier and Standard Error of Laboratory.....	72
3.4. Spectral Preprocessing.....	73
3.5. Model Development.....	76
3.6. Results and Discussion.....	78
3.6.1. Comparison of Near-Infrared Spectra of Ground Biomass with Pure Cellulose and Hemicellulose.....	78
3.6.2. Higher Heating Value and Ultimate Analysis in Ground Biomass	79
3.7. Comparison with Previous Work.....	97

3.8. Conclusions.....	99
References	100
Chapter 4 Evaluating limit of detection and quantification for higher heating value and ultimate analysis of fast-growing trees and agricultural residues biomass using NIRS	107
Abbreviations.....	109
4.1. Introduction	110
4.2. Materials and methods.....	112
4.2.1 Samples.....	112
4.2.2 Instrument.....	112
4.2.3. Spectral preprocessing and Model development.....	113
4.2.4. Limit of detection and limit of quantification.....	116
4.3. Results and Discussion.....	116
4.4. Conclusions.....	120
References	121
Chapter 5 Effect of combined non-wood and wood spectra of biomass chips on rapid prediction of ultimate analysis parameters using near infrared spectroscopy	125
Abbreviations.....	127
5.1 Introduction	128
5.2 Materials and Methods.....	132
5.2.1 Sample Preparation	133
5.2.2 Spectral Data Collection.....	133
5.2.3 Reference Analysis	133

5.2.4 Outlier and Standard Error of Laboratory	134
5.2.5. Spectral Preprocessing and Model Development	134
5.2.6 Limit of Quantification (LOQ)	136
5.3 Results and Discussion.....	136
5.3.1 wt.% of C.....	138
5.3.2 wt.% of H	141
5.3.3 wt.% of O	143
5.3.4 wt.% of N	144
5.3.5 Effect of Non-Wood and Wood Samples on Model Performance.....	146
5.3.6 Comparison of Model Performance between Using Chipped and Ground	
Biomass Spectra	155
5.4 Conclusions.....	157
References.....	159
Chapter 6 NIR spectroscopy as an alternative to thermogravimetric analyzer for	
biomass proximate analysis : Comparison of chip and ground biomass models.....	163
Abbreviations.....	165
6.1 Introduction	167
6.2 Materials and Methods.....	171
6.2.1 Sample Preparation	172
6.2.2 FT-NIRS Scanning	172
6.2.3 Reference Analysis (Proximate Analysis).....	174
6.2.4 Outlier Identification.....	176

6.2.5 Spectral Preprocessing	177
6.2.6 Partial Least Squares Regression Model Development	178
6.2.7 Limit of Quantification (LOQ)	180
6.3 Results and Discussion.....	180
6.3.1 Spectral Preprocessing	180
6.3.2 Reference Data.....	181
6.3.3 Prediction of Proximate Analysis Data	184
6.3.4 Limit of Quantification (LOQ)	197
6.3.5 Effect of Combined Wood Samples with Non-Wood Samples in Developed Models.....	199
6.3.6 Comparison of the Model Performance between Chip and Ground Biomass	200
6.3.7 Comparison with Previous Work.....	201
6.4 Conclusions.....	202
References	204
Chapter 7 Near-Infrared Spectroscopy Modeling of Combustion Characteristics in Chip and Ground Biomass from Fast-Growing Trees and Agricultural Residue.....	209
Abbreviations.....	211
7.1. Introduction	213
7.2. Materials and Methods	216
7.2.1. Sample Preparation	218
7.2.2. FT-NIRS Scanning	219
7.2.3. Thermogravimetric Analysis Experiment.....	219

7.2.4. Reference Data Calculation.....	220
7.2.5. Outlier Identification	220
7.2.6. Partial Least Squares Regression Modeling	221
7.3. Results and Discussion.....	222
7.3.1. NIR Spectra of Fast-Growing Trees and Agricultural Residues.....	222
7.3.2. Combustion Characteristic Parameters and Combustion Performance Indices from TGA.....	223
7.3.3. Modeling for Combustion Performance Indices.....	226
7.3.4. Comparison with Previous Work.....	242
7.3.5. Benefit of Combined Agricultural Residue with Fast-Growing Trees in Model Development.....	244
7.4. Conclusions.....	245
References	247
Chapter 8 Conclusion and recommendations	252
8.1 Conclusion.....	252
8.2 Recommendations.....	254
Appendix.....	256
Author Biography	257

LIST OF TABLES

Table 2.1 Engineering application of physical properties of lignocellulosic biomass [37].....	27
---	----

Table 2.2 Guideline for the interpretation of R^2 (Williams et al. (2019))	48
Table 3.1 Statistical data of the HHV and ultimate analysis parameters of the grounded biomass used in PLSR model development.....	82
Table 3.2 Average reference value of HHV (J/g), ash content (wt.%), and ultimate analysis parameter of fast-growing trees and agricultural residues on a dry basis (wt.%).	83
Table 3.3 Results of the PLSR-based model for the HHV (J/g) and ultimate analysis (wt.%) of grounded biomass, bolded model showing the best performance.....	87
Table 4.1 Analysis result of the HHV and ultimate analysis parameters of the ground biomass for assessment of LOD and LOQ	118
Table 5.1 The number of non-wood samples and wood samples in calibration set and validation set.....	136
Table 5.2 The statistical data of the ultimate analysis parameters of the chip biomass obtained using CHNS/O elemental analyzer used in PLSR model development.....	137
Table 5.3 Results of the PLSR-based model for ultimate analysis (wt.%) of chip biomass, bolded model showing the best performance.....	138
Table 5.4 The range of wt.% of C, H, N, and O of non-wood and wood samples in calibration and validation sets.....	147
Table 5.5 The trend line characteristics of the wood and non-wood species in scatter plots of the best models for C, H, N, and O.	149
Table 5.6 The trend line characteristics of specific biomass species for Carbon evaluation optimized model.....	152
Table 5.7 The trend line characteristics of specific biomass species for Nitrogen evaluation optimized model.....	153
Table 5.8 The trend line characteristics of specific biomass species for Hydrogen evaluation optimized model.....	154
Table 5.9 The trend line characteristics of specific biomass species for Oxygen evaluation optimized model.....	154

Table 6.1 Statistical data of the proximate analysis parameters of the chip and ground biomass used in the PLSR model development.	185
Table 6.2 Details on the number of identified outliers in chip and ground biomass evaluated before the development of the PLSR model.	186
Table 6.3 Results of the PLS regression-based model for the proximate analysis (%) of chip biomass, with the bolded model showing the best performance.	187
Table 6.4 Results of the PLS regression-based model for the proximate analysis (%) of ground biomass, with the bolded model showing the best performance.	189
Table 6.5 The dominant peaks on the regression coefficient plot obtained from the best-performing PLSR-based model.	192
Table 6.6 LOQ results of PLSR models for the % of MC, VM, FC, and ash content in chip and ground biomass.	198
Table 6.7 The range of % of MC, VM, FC, and ash content of wood and non-wood samples in calibration and validation sets.	199
Table 7.1 Combustion parameters and performance indices for fast-growing trees and agricultural residue subjected to direct combustion using TGA.	228
Table 7.2 Statistical data of combustion performance parameters for ground and chipped biomass of fast-growing trees and agricultural residue used in model development (after outliers were removed from the total of 115 samples).	229
Table 7.3 Results of the PLSR-based model for the combustion performance indices of chip biomass, with the model showing the best performance highlighted in bold.	230
Table 7.4 Results of the PLSR-based model for the combustion performance indices of ground biomass, with the model showing the best performance highlighted in bold.	231
Table 7.5 The dominant peaks on the regression coefficient plot and average absorbance plot obtained via the best-performing PLSR-based model [44].	236

LIST OF FIGURES

Figure 1.1. Overall experimental flowchart to predict the energy properties and combustion characteristics of fast-growing trees and agricultural residues in Nepal for energy usage using near-infrared spectroscopy.	7
Figure 2.1: Electromagnetic spectral regions.....	13
Figure 2.2: Basic NIR spectrometer configuration [1].....	15
Figure 2.3: Transmission.....	16
Figure 2.4: Diffuse reflection.....	17
Figure 2.5 Transflection.....	17
Figure 2.6 Interaction.....	18
Figure 2.7 Optical diagram of Michelson interferometer.....	18
Figure 2.8 Construction of bomb calorimeter.....	28
Figure 3.1 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.....	68
Figure 3.2 Nepal biomass in (a) chips form (>30 mm × 15 mm), (b) grounded form (1.88–3080 μm), (c) FT-NIRS (MPA, Bruker, Ettlingen, Germany) scanning between the wavenumber range 3595 to 12,489 cm ⁻¹ , and (d) ground sample presentation by transreflectance mode.	70
Figure 3.3 Representative particle size distribution of the ground biomass ranging from 0.01 to 3080 μm.	70
Figure 3.4 (a) Raw spectra of grounded biomass. (b) Preprocessing of raw spectra using traditional approach (MSC). (c) Pretreated spectra by the multi-preprocessing method with a 5-range method in five equal sections: MSC, empty, raw spectra, empty, and raw spectra (from left to right), within the wavenumber range of 3625.72–12,489.48 cm ⁻¹ . (d) Pretreated spectra by the multi-preprocessing method with a 3-range method in three sections:	

Second derivative, first derivative, and constant offset (from left to right) within the wavenumber range of 3594.87–12,489.48 cm ⁻¹	71
Figure 3.5 Spectra of fast-growing trees and agricultural residues compared to pure cellulose and pure lignin.	79
Figure 3.6 Histogram of reference values in calibration and validation sets for the (a) HHV (J/g), (b) wt.% C, (c) wt.% of H, (d) wt.% of O, and (e) wt.% of N.....	81
Figure 3.7 Measured versus predicted value in calibration and validation sets for the (a) HHV (J/g), (b) wt.% of C, (c) wt.% of H, (d) wt.% of O, and (e) wt.% of N.	85
Figure 3.8 The average absorbance value of HHV (J/g) obtained using the first derivative preprocessing with a selection of important wavenumbers obtained from GA, within the full wavenumber range of 3594.87–12,489.5 cm ⁻¹	87
Figure 3.9 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 3594.87–12,489.5 cm ⁻¹	91
Figure 3.10 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 3594.87–12,489.5 cm ⁻¹	93
Figure 3.11 The regression coefficient for the wt.% O of grounded biomass using the multi-preprocessing PLSR 5-range method.....	95
Figure 3.12 The regression coefficient for the wt.% of N of grounded biomass using the multi-preprocessing PLSR 3-range method.....	97
Figure 4.1 Representative particle size distribution of the ground biomass ranging from 0.01 to 3080 μm.	113
Figure 4.2 The average raw spectra of the fast-growing trees and agricultural residues of Nepal grounded biomass spanning the range of 12489.48 – 3594.87 cm ⁻¹ [15].....	115
Figure 4.3 Scatter plot of measured vs predicted values from the training set for: a) HHV, b) wt.% of carbon, c) wt.% of oxygen, d) wt.% of hydrogen, and e) wt.% of nitrogen.	119

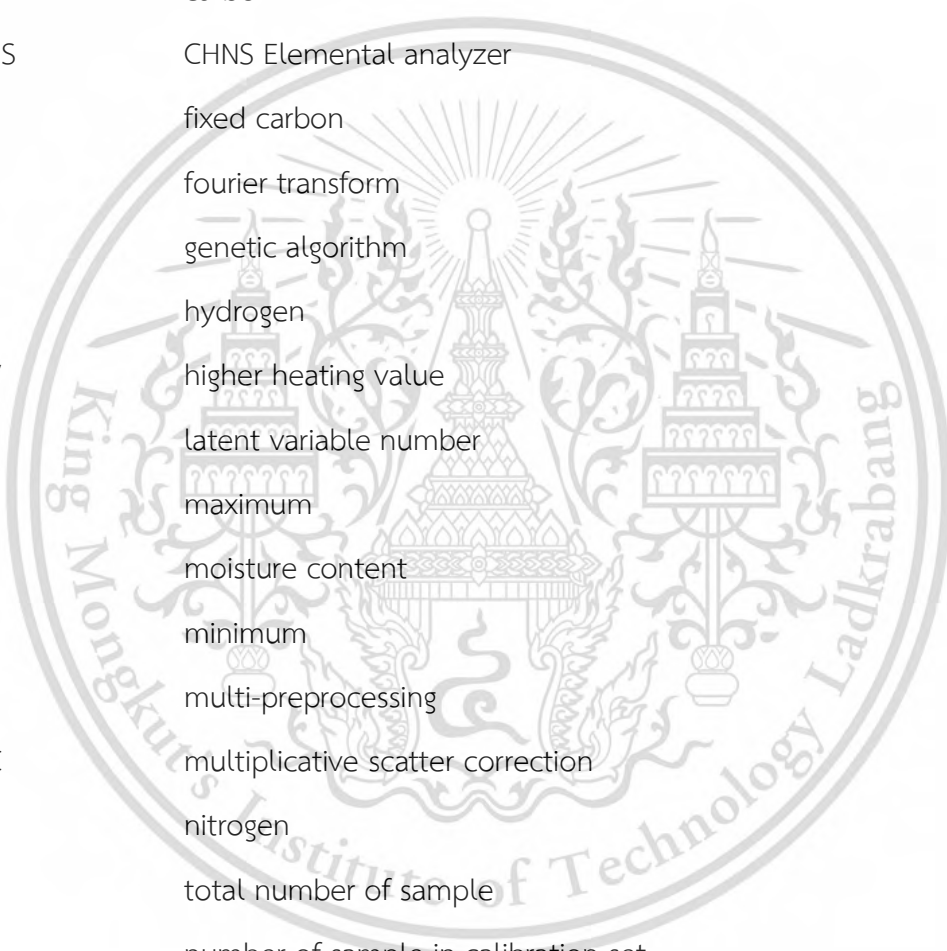
Figure 5.1 Flowchart of the overall research methodology for the rapid prediction of the ultimate analysis parameters of chip biomass for energy usage by NIRS using PLSR.	132
Figure 5.2 Measured versus predicted value in calibration and prediction sets for (a) wt.% of C, (b) wt.% of H, (c) wt.% of O, and (d) wt.% of N.....	140
Figure 5.3 The second derivative absorbance value of studied biomass obtained using the sd2 preprocessing with a selection of important wavenumber obtained from GA for prediction of wt.% of C, within the full wavenumber range of 3594.87–12489.48 cm ⁻¹	141
Figure 5.4 The vector normalization absorbance value of studied biomass obtained using the vector normalization preprocessing with a selection of important wavenumber obtained from GA for prediction of wt.% of H, within the full wavenumber range of 3594.87–12489.48 cm ⁻¹	142
Figure 5.5 The regression coefficient for the wt.% O of chip biomass using the MP PLSR 5-range method.....	144
Figure 5.6 The regression coefficient for the wt.% N of chip biomass using the MP PLSR 3-range method.....	146
Figure 5.7 The scatter plots of optimized model for wt.% of (a) C, (b) H, (c) O, and (d) N where the simple regression lines of non-wood group and wood group illustrated both in calibration set and validation set.....	150
Figure 6.1 Flowchart depicting the overall research methodology for the rapid assessment of biomass proximate analysis parameters using TGA, while using NIRS with PLSR as an alternative.	171
Figure 6.2 <i>FT-NIR spectrometer scanning of (a) chip biomass in diffuse reflectance and sphere macro sample rotating mode using a quartz-bottom sample cup, and (b) ground biomass in transfectance mode using a glass vial.</i>	173
Figure 6.3 Raw spectra of (a) chip biomass obtained from diffuse reflectance mode of FT-NIRS scanning and (b) ground biomass obtained from transfectance mode of FT-NIRS scanning across the entire wavenumber range from 3594.87 to 12,489.48 cm ⁻¹	173

Figure 6.4 Representative particle size distribution of the ground biomass ranging from 0.01 to 3080 μm	174
Figure 6.5 <i>The TG and DTG curves obtained from TGA analysis of Nepalese biomass</i>	175
Figure 6.6 (a) TG and (b) DTG profiles of fast-growing tree and agricultural residue biomass obtained using TGA within the temperature range from 35 to 700 $^{\circ}\text{C}$ with a heat flow rate of 10 $^{\circ}\text{C}/\text{min}$	176
Figure 6.7 The average pretreated spectra from the best performing model for: (a) MC with the second derivative, (b) VM with a multi-preprocessing combination set of 4, 4, 3, 2, and 0, (c) FC with a constant offset, and (d) ash content with the first derivative.	181
Figure 6.8 <i>Overall proximate analysis results of ten different biomass varieties, including four wood varieties of fast-growing trees and six non-wood varieties (five agricultural residues and one fast-growing tree) obtained from the TGA analysis.</i>	182
Figure 6.9 Measured versus predicted value in calibration set and validation set for chip biomass (a) MC, (b) VM, (c) FC, and (d) ash and for ground biomass (e) MC, (f) VM, (g) FC, and (h) ash.	183
Figure 6.10 The average absorbance value of MC (%) of chip biomass obtained using second derivative preprocessing with a selection of important wavenumbers obtained from GA.	190
Figure 6.11 The regression coefficient for the MC (%) of ground biomass using the Full-PLSR with spectral preprocessing of mean centering.	190
Figure 6.12 Regression coefficient for the % of VM in chip biomass using the MP PLSR 5-range method with a spectral multi-preprocessing combination set of 4, 4, 3, 2, and 0.	193
Figure 6.13 Regression coefficient for the % of VM in ground biomass using the MP PLSR 5-range method with a spectral multi-preprocessing combination set of 3, 5, 4, 5, and 0.	194
Figure 6.14 Regression coefficient for the % of FC in chip biomass using the Full-PLSR method with a spectral preprocessing of constant offset.	195

Figure 6.15 Regression coefficient for the % of FC in ground biomass using Full-PLSR method with a spectral preprocessing of first derivative.....	196
Figure 6.16 Regression coefficient for the % of ash content in chip biomass using the MP PLSR 3-range method with a spectral multi-preprocessing combination set of 5, 4, and 4.....	196
Figure 6.17 Regression coefficient for the % of ash content in ground biomass using Full-PLSR method with a spectral preprocessing of D1.....	197
Figure 7.1 Flowchart illustrating the comprehensive research methodology for evaluating combustion performance indices of fast-growing trees and agricultural residues, using TGA in conjunction with NIRS combined with PLSR.....	217
Figure 7.2 Average raw spectra of fast-growing trees and agricultural residue for (a) chip biomass obtained through diffuse reflectance mode and (b) ground biomass obtained through transmittance mode of FT-NIRS scanning.....	223
Figure 7.3 TG and DTG curves obtained via TGA within the temperature range of 35 to 700 °C for (a) <i>Alnus nepalensis</i> , (b) <i>Pinus roxiburghii</i> , (c) <i>Bombusa vulagris</i> , (d) <i>Eucalyptus camaldulensis</i> , (e) <i>Bombax ceiba</i> , (f) <i>Zea mays</i> (cob), (g) <i>Zea mays</i> (shell), (h) <i>Zea mays</i> (stover), (i) <i>Oryza sativa</i> , and (j) <i>Saccharum officinarum</i>	224
Figure 7.4 The average spectrum for (a) D_i in chip biomass is obtained by using the D2 (5,5). (b) D_i in ground biomass is obtained by using the multi-preprocessing PLSR-5 range with a combination set of 3,5,3,1,0. (c) D_f in chip biomass is achieved by using min-max normalization. (d) D_f in ground biomass is obtained by using D1 (5,5). (e) S_i in chip biomass is acquired by using D1 (5,5). (f) S_i in ground biomass is acquired by MSC. (g) C_i in chip biomass is calculated by using the multi-preprocessing PLSR-3 range method with a combination set of 2,5,0. (h) C_i in ground biomass is determined by the multi-preprocessing PLSR-5 range method with a combination set of 3,2,1,1,4.....	232
Figure 7.5 Measured versus predicted value in calibration set and validation set for chip biomass: (a) ignition index, (b) burnout index, (c) combustion performance index, and (d)	

flammability index; and for ground biomass: (e) ignition index, (f) burnout index, (g) combustion performance index, and (h) flammability index.....	233
Figure 7.6 The regression coefficient for the D_i of chip biomass using the full-PLSR approach with spectral preprocessing of D2.....	234
Figure 7.7 The regression coefficient for the D_i of ground biomass using the multi-preprocessing PLSR-5 range method with a spectral preprocessing combination set of 3, 5, 3, 1, 0. ...	234
Figure 7.8 The average absorbance value of D_f in chip biomass using min-max normalization with selection of important wavenumbers obtained through GA.....	237
Figure 7.9 The average absorbance value of D_f in ground biomass using D1 with selection of important wavenumbers obtained via GA.....	238
Figure 7.10 The average absorbance value of S_i in chip biomass using D1 with selection of important wavenumbers obtained through GA.	239
Figure 7.11 The regression coefficient for the S_i of ground biomass using the full-PLSR model with spectral preprocessing of MSC.....	240
Figure 7.12 The regression coefficient for the C_i of chip biomass using the multi-preprocessing PLSR-3 range method with a spectral preprocessing combination set of 2, 5, 0.	241
Figure 7.13 The regression coefficient for the C_i of ground biomass using the multi-preprocessing PLSR-5 range method with a spectral preprocessing combination set of 5, 0, 4, 2, 5. ...	242

ABBREVIATIONS



%	percentage
A	ash content
C	carbon
CHNS	CHNS Elemental analyzer
FC	fixed carbon
FT	fourier transform
GA	genetic algorithm
H	hydrogen
HHV	higher heating value
LVs	latent variable number
Max	maximum
MC	moisture content
Min	minimum
MP	multi-preprocessing
MSC	multiplicative scatter correction
N	nitrogen
N_T	total number of sample
N_c	number of sample in calibration set
NIRS	near infrared spectroscopy
N_p	number of sample in validation set
O	oxygen
PLSR	partial least squares regression
R^2	coefficient of determination

R^2_c	coefficient of determination of calibration set
R^2_p	coefficient of determination of validation set
RMSEC	root mean square error of calibration set
RMSEP	root mean square error of prediction set
RPD	ratio of prediction to deviation
S	sulfur
SD	standard deviation
SEC	standard error of calibration set
SEL	standard error of laboratory
SEP	standard error of validation set
SNV	standard normal variate
SPA	successive projection algorithm
SW	selected wavenumber
TGA	thermogravimetric analysis
VM	volatile matter
wt. %	weight percentage



Chapter 1

Introduction

1.1 Statement and Significance of the problem

Energy is a fundamental driver of social and economic development for nations. Sustainable energy sources, both in the short and long term, are essential for economic growth, improved quality of life, and environmental protection. The relentless increase in global energy demand has prompted a search for diverse energy sources that can meet this demand while minimizing detrimental effects on energy availability and the environment. According to the International Energy Agency (IEA), as of 2020, 80% of global primary energy consumption is attributed to fossil fuels, resulting in a substantial carbon footprint [1]. The predominant reliance on non-renewable fossil fuels contributes to environmental degradation through heightened greenhouse gas emissions and raises compelling concerns regarding future energy security due to finite reserves. The financial burdens associated with fossil fuels, compounded by their price volatility and geopolitical uncertainties, have triggered an urgent pursuit of alternative energy solutions.

A balanced and sustainable energy portfolio requires the promotion of renewable energy sources and technological mix, primarily including hydro, wind, solar, biomass, and geothermal. Among these, biomass energy emerges as a promising solution, constituting 10% of the global energy supply and 13% of energy consumption [2]. Biomass, as a renewable, carbon-neutral, abundantly available, and sustainable energy source, has the potential to reduce or replace reliance on fossil fuels, leading to energy savings and environmental preservation [3]. Studies indicate that global lignocellulosic biomass, including crop residues, wood, and grass, generates approximately 181.5 billion tonnes each year [4]. This underscores the significant potential of utilizing agricultural residues and wood from fast-growing trees in combination with efficient combustion energy technologies for bio-conversion, with the added benefit of substantial reductions in harmful emissions into the atmosphere, making biomass a viable alternative to fossil fuels. Additionally, it contributes to waste management and rural development.

Nepal, a landlocked agricultural country covering a total area of 147,516 km², is ecologically divided into three regions: terai, hill, and mountain. The country is heavily forested, and over 33.83% of its population resides in rural municipalities [5]. Nepal's energy sources fall into three broad categories: traditional, commercial, and alternative energy resources. Traditional sources encompass biomass fuels, particularly fuelwood, agricultural residues, and animal dung, used in the traditional way through direct combustion. Commercial sources include fossil fuels and electricity. Alternative sources consist of solar photovoltaics and thermal energy, biomass energy, wind energy, biofuel, biogas, and mini/micro hydropower [6]. In 2017, Nepal's total energy consumption reached 544 PJ (Peta joule). Biomass resources alone supplied approximately 76%, equivalent to 413.5 PJ, of the total energy requirements, while commercial energy sources (electricity, LPG, gasoline, jet fuel, diesel, and coal) accounted for the remaining 24% [7]. Fuelwood and agricultural residues contributed 65% (353.6 PJ) and 4% (21.8 PJ), respectively, to the overall energy consumption [7]. In the fiscal year 2018/19, total energy consumption reached 14.014 million toe (tons of oil equivalent), with traditional, commercial, and alternative energy resources making up 68.5%, 29.4%, and 2.1% of the total energy consumption, respectively [8].

According to the 2021 census, over half (51%) of Nepalese households primarily use wood/firewood for cooking. Other main sources of cooking fuel include liquefied petroleum gas (44.3%), cow dung (2.9%), biogas (1.2%), electricity (0.5%), kerosene (0.05%), and other sources (0.1%) [5]. The Nepal Energy Sector Vision 2050 A.D. report predicts that total energy consumption in Nepal will reach 626,000 TJ, and 1,011,000 TJ in 2030, and 2050, respectively. This indicates a substantial demand for solid biomass for the country's energy needs in the coming years, followed by an increased demand for petroleum products [9].

The Government of Nepal, through the Alternative Energy Promotion Center, the national focal point for the promotion and development of renewable energy in Nepal, as well as its development partners, private sectors, and non-governmental organizations, is promoting various biomass energy technologies under renewable energy subsidy policy of 2016 across all three ecological regions of Nepal. This initiative aims to ensure quality and enhance beneficiary trust in these technologies [10]. Despite the robust nature of biomass energy technologies and its efficient performance in laboratory setting, the real-world application and overall performance remain satisfactory due to lack of sufficient knowledge regarding biomass energy properties. There is also

a lack of adequate information regarding the potential energy generation from locally available firewood and various agricultural residues[11]. Proper identification of biomass, considering its actual characteristics, and its effective management and utilization through appropriate energy technologies have not been implemented sustainably. Consequently, all investments, whether from the government, public, or private sectors, in the research, development, and promotion of biomass energy technologies have not been able to meet the actual energy needs and demands of rural populations and industries. Biomass traders and energy companies often engage in biomass trading based on mass or volume rather than its actual energy values. Hence, there is a national priority for research focused on the rapid, reliable, and non-destructive characterization and prediction of biomass properties as a fuel. This will enable the effective and efficient promotion of biomass energy technologies in residential, commercial, and industrial sectors.

To date, to the best of our knowledge, no research or report has been published on the evaluation and prediction of Nepal biomass energy properties using FT- NIRS. Biomass has the unique quality of being a strong absorber of NIR radiation within the range of 3595 to 12,489 cm^{-1} . It predominantly interacts with the bonds of non-symmetrical molecules, such as C, O, H, and N [12, 13], which makes it highly suitable for use in conjunction with NIRS and chemometrics to assess various energy-related properties of biomass. These properties include the HHV, ultimate analysis parameters, proximate analysis parameters, as well as the combustion performance indices. Several research has successfully utilized NIRS to develop models for the rapid and accurate measurement of various biomass properties for energy utilization. For instance, Posom et al. [14] established a reliable online method for measuring the HHV of sugarcane bagasse 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. [16] created a prediction model for the HHV and the elemental composition (C, H, and N) of straw using NIRS. Posom et al. predicted the HHV [17] and elemental composition (C, H, N, O, and S) [18] of grounded bamboo using NIRS. Skvaril et al. [19] reviewed the application of NIRS in biomass energy conversion processes. Zhang et al. [20] explored the fast analysis of HHV and elemental composition of sorghum biomass using NIRS. Xue et al. [21] investigated the use of an online NIRS system for measurement of crop straw fuel properties. These studies collectively demonstrate the potential of NIRS to provide rapid, reliable, and non-destructive alternative methods for

characterizing biomass for energy usage compared to traditional destructive thermal analysis techniques. This approach offers several advantages, including assisting biomass traders in determining a fair price based on the actual energy properties rather than relying solely on mass or volume. Furthermore, industries reliant on biomass for energy can optimize system efficiency and cost-effectiveness through informed biomass selection. Most importantly, it was observed that no paper reported NIRS models developed using combined data from ten different types of biomass, especially fast-growing trees and agricultural residues, but rather specific types. Therefore, this study proposes NIRS-based research to characterize fast-growing trees and agricultural residues biomass in Nepal for energy purposes using NIRS.

1.2 Goal and objective

The main objective of this thesis was to predict the energy properties and combustion characteristics of fast-growing trees and agricultural residues in Nepal for energy usage using near-infrared spectroscopy as an alternative method. This prediction was based on reference data obtained from a bomb calorimeter for higher heating value, CHNO/S elemental analyzer for ultimate analysis parameters, and a thermogravimetric analyzer for proximate analysis parameters and combustion performance indices.

The specific objectives of this research were as follows:

- To develop and compare the performance of PLSR-based models correlating the NIR spectral characteristics of the fast growing trees and agricultural residues with the higher heating value, ultimate analysis parameters, and proximate analysis parameters in both chip and ground biomass forms using near infrared spectroscopy.
- To develop and compare the performance of PLSR-based models correlating the NIR spectral characteristic of the fast growing trees and agricultural residues with the combustion characteristics, specifically the ignition index, burnout index, comprehensive combustion performance index, and flammability index, in both the chip and ground biomass forms using near infrared spectroscopy.
- To select the best performing PLSR-based model and demonstrate the potential of NIR spectroscopy as a rapid, reliable, and accurate alternative method for evaluating the biomass higher heating value, ultimate analysis parameters, proximate parameters, and

combustion performance indices and to prove whether the models developed is ready for industrial use e.g. in case of higher heating value.

1.3 Novelty of thesis research

Spectral preprocessing is a crucial component of NIR calibration [22-24]. Traditionally, the development of NIRS models for predicting and evaluating sample properties has relied solely on the pretreatment of the entire spectrum using traditional preprocessing techniques. However, there has been a lack of exploration regarding the pretreatment of raw spectra by employing different preprocessing techniques on distinct sections of the entire wavelength. It is thought that a multi-preprocessing approach, i.e., a novel spectral preprocessing technique that divides the entire spectrum into different sections using various spectral preprocessing methods based on random pairs, will improve the assessment of biomass properties for energy usage using NIRS. Based on this assumption, this study introduces two methods: the "multi-preprocessing 5-range method" and the "multi-preprocessing 3-range method," as unique components to enhance the assessment of biomass properties using NIRS. The research outcome from the multi-preprocessing technique with PLSR is expected to serve as a pivotal milestone in the research and development of NIRS. The effect of biomass species especially wood and non-wood species to be included in the model development on the model performance through the scatter plot analysis is first time proposed and investigated by this thesis.

The novel concept of a spectral multi-preprocessing approach was first introduced in the research titled "Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques" (refer to Chapter 3) [22]. This approach has subsequently been applied in the research described in Chapters 4-7. Similarly, the impact of including spectra from various biomass species (both non-wood and wood) in the model, analyzed using scatter plot techniques, is introduced in the research titled "Effect of Combined Non-Wood and Wood Spectra of Biomass Chips on Rapid Prediction of Ultimate Analysis Parameters Using Near-Infrared Spectroscopy" (refer to Chapter 5).

1.4 Hypothesis to be tested

The hypothesis is that near-infrared spectroscopy technique can be a rapid, reliable, accurate, and non-destructive alternative method for evaluating the higher heating value, ultimate analysis parameters, proximate analysis parameters, and combustion performance indices of fast growing trees and agricultural residues.

Null hypothesis (H_0): NIRS cannot be alternative to conventional methods for evaluating the higher heating value, ultimate analysis parameters, proximate analysis parameters, and combustion performance indices of fast growing trees and agricultural residues.

Alternative hypothesis (H_1): NIRS can be alternative method to conventional methods for evaluating the higher heating value, ultimate analysis parameters, proximate analysis parameters, and combustion performance indices of fast growing trees and agricultural residues.

1.5 Scope of the study

In this research, ten different varieties of biomass samples that are locally available and commonly used for energy purpose were collected from hill and terai region of Nepal. These include five varieties of fast-growing trees, namely *Alnus nepalensis*, *Pinus roxburghii*, *Bombusa vulagris*, *Bambax ceiba*, *Eucalyptus camaldulensis*, and five varieties of agricultural residues, namely *Zea mays* (cob), *Zea mays* (shell), *Zea mays* (stover), *Oryza sativa*, and *Saccharum officinarum*. This sample were used to assess the potential of NIR spectroscopy in characterizing biomass for energy usage, specifically in terms of HHV, ultimate analysis, proximate analysis, and combustion performance index, both in chip and ground biomass form.

For spectral data, an FT-NIR spectrometer (MPA, Bruker, Ettlingen, Germany) with a wavelength range of 3594.87 to 12,489.48 cm^{-1} , a resolution of 16 cm^{-1} , and a background scan time and sample scan time of 32 scans (average) was utilized. To measure HHV, an automatic bomb calorimeter (IKA C 200, staufen, Baden – Württemberg, Germany) was used. For ultimate analysis, a CHNS/O analyzer (Thermo Scientific™ FLASH 2000, Waltham, Massachusetts, United States) was employed. Additionally, for proximate analysis and combustion performance indices, a thermogravimetric analyzer (TG 209 F3 Tarsus, Netzsch, Bavaria, Germany) with a microbalance sensitivity of 0.1 μg resolution was used.

1.6 Research mapping

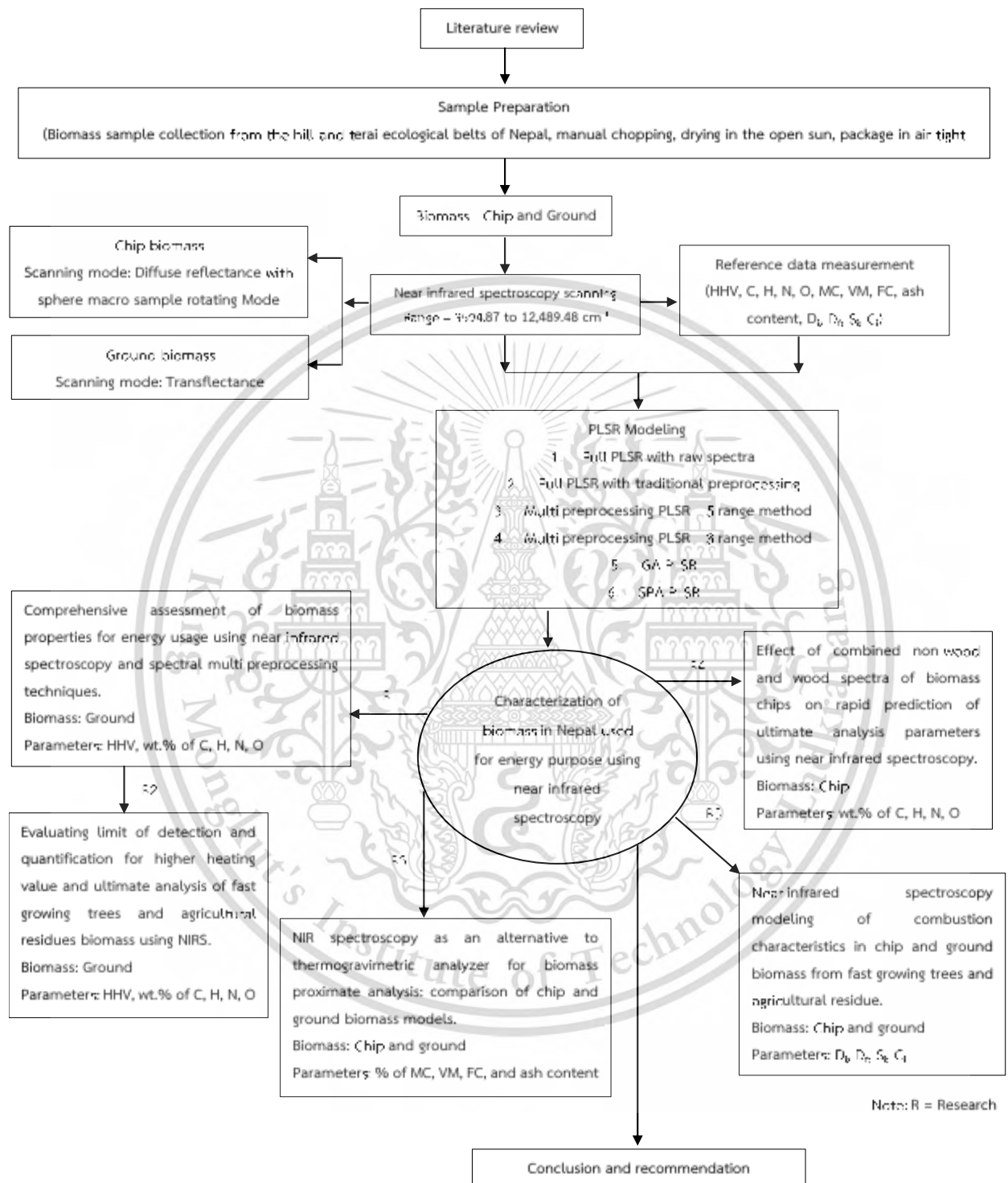


Figure 1.1. Overall experimental flowchart to predict the energy properties and combustion characteristics of fast-growing trees and agricultural residues in Nepal for energy usage using near-infrared spectroscopy.

1.7 Process of the study

To achieve the goals and objectives of the research, the thesis is organized into eight chapters.

Chapter 1 focuses on outlining the statement and significance of the research problems, the research objectives, the novelty of the thesis, the hypothesis to be tested, the scope of this thesis, and the overall experimental research flow chart.

Chapter 2 consists of a literature review on NIRS, biomass, combustion, and quantitative modeling, specifically PLSR. These topics are discussed in alignment with the research objectives.

Chapter 3 provides a comprehensive assessment of the HHV and ultimate analysis parameters of ground biomass for energy usage using NIRS. The main purpose of this chapter is to develop, compare, and select the best-performing PLSR-based model (Full-PLSR, the multi-preprocessing PLSR-5 range method, the multi-preprocessing PLSR-3 range method, GA-PLSR, SPA-PLSR) for each parameter and establish it as a reliable and non-destructive alternative method that could have a practical application in real life. The chapter concludes with the development of the PLSR-based model for the assessment of HHV and ultimate analysis parameters in ground biomass.

Chapter 4 presents an assessment of the limit of detection and limit of quantification in the recommended models for HHV and the ultimate analysis parameters of ground biomass from chapter 3. The main purpose of this study is to demonstrate the model's sensitivity for each parameter, based on the best training model's performance.

Chapter 5 examines the effects of combined non-wood and wood biomass chip spectra on the rapid prediction of ultimate analysis parameters using NIRS with PLSR, following the same algorithms described in Chapter 3. The primary goal is to analyze the model's performance for each parameter in chip biomass, compare it to ground biomass, and select the most suitable model. Additionally, this chapter investigates the impact of mixed non-wood and wood species on model performance through scatter plot analysis and evaluates the model's sensitivity by assessing the limit of quantification for each parameter in the proposed calibration set.

Chapter 6 evaluates the performance of PLSR-based models for proximate analysis parameters (MC, VM, FC, and ash content) in chipped and ground biomass, using the same algorithms as those for HHV and ultimate analysis. The main goal is to determine these proximate

analysis parameters as alternatives to TGA. This chapter concludes with recommendations for selecting the best model for each parameter and calculates the limit of quantification for both chipped and ground biomass.

Chapter 7 presents the results on biomass combustion characteristics parameters and combustion performance indices obtained from a TGA. Using the spectral data corresponding to each combustion performance index, PLSR-based models were developed using the same algorithm as that used for HHV, ultimate analysis parameters, and proximate analysis parameters for both chipped and ground biomass forms. Their performances were compared, and the best-performing model for each index was recommended, establishing it as a rapid and non-destructive alternative method for assessing combustion performance indexes.

Chapter 8 summarizes the results of the thesis and provides recommendations for further study based on the findings of this study.



References

1. IEA Greenhouse Gas Emissions from Energy Data Explorer. <https://www.iea.org/data-and-statistics/data-tools/greenhouse-gas-emissions-from-energy-data-explorer> (11 August 2023),
2. Popp, J.; Kovács, S.; Oláh, J.; Divéki, Z.; Balázs, E., Bioeconomy: Biomass and biomass-based energy supply and demand. *New Biotechnology* **2021**, *60*, 76-84.
3. 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).
4. Nimmanterdwong, P.; Chalermsoonsuwan, B.; Piumsomboon, P., Prediction of lignocellulosic biomass structural components from ultimate/proximate analysis. *Energy* **2021**, *222*, 119945.
5. *National Population and Housing Census 2021 - National Report*; Office of the Prime Minister and Council of Ministers, Government of Nepal: Nepal, May 2023, 2021.
6. *Electricity Demand Forecast Report (2015-2040)*; Water and Energy Commission Secretariat, Government of Nepal: Kathmandu, January 2017.
7. Bhusal, S.; Nakarmi, A. M., Sustainable Energy Planning for Nepal in The Federal Structure. *Journal of Advanced College of Engineering and Management* **2019**, *5*, 127-145.
8. *Economic Survey 2019/20*; Government of Nepal: Singh Durbar, Kathmandu, 2020.
9. *Nepal's Energy Sector Vision 2020 A.D.*; Water and Energy Commission Secretariat, Government of Nepal: Singha Durbar, Kathmandu, Nepal, 2013.
10. *Renewable Energy Subsidy Policy, 2073 BS*; Ministry of Population and Environment, The Government of Nepal: May 2016.
11. *Biomass Energy Strategy 2017*; Ministry of Population and Environment, The Government of Nepal: Kathmandu, January 2017.
12. 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>.
13. Roger, J.-M.; Mallet, A.; Marini, F., Preprocessing NIR Spectra for Aquaphotomics. *Molecules* **2022**, *27*, (20), 6795. <https://doi.org/10.3390/molecules27206795>.

14. Posom, J.; Phuphaphud, 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. Phuphaphud, 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. 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>.
17. Jetsada, P.; 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.
18. 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.
19. 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>.
20. 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>.
21. 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>.
22. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B. P., Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* **2023**, *16*, (14), 5351.

23. Schoot, M.; Kapper, C.; van Kollenburg, G. H.; Postma, G. J.; van Kessel, G.; Buydens, L. M.; Jansen, J. J., Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size. *Chemometrics and Intelligent Laboratory Systems* **2020**, *204*, 104105.
24. Rinnan, Å.; Van Den Berg, F.; Engelsen, S. B., Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* **2009**, *28*, (10), 1201-1222.



Chapter 2

Literature Review

2.1 Near Infrared Spectroscopy (NIRS)

Spectroscopy is the subject related to the measurement and analysis of electromagnetic radiation that is absorbed or emitted by the materials. Figure 2.1 shows the electromagnetic spectral regions, which includes various types of electromagnetic radiation with different wavelengths (λ) and frequencies (ν).

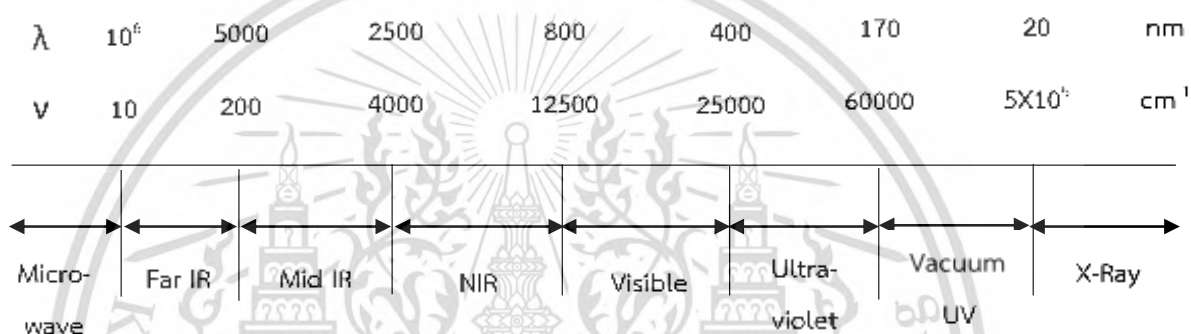


Figure 2.1: Electromagnetic spectral regions

As shown in Figure 2.1, the near-infrared (NIR) region of the electromagnetic spectrum is a portion of the infrared region located just beyond the visible light spectrum. NIRS is based on the absorption of electromagnetic radiation at the wavelength range approximately from 800 to 2500 nm corresponding to the wavenumber range 12500-4000 cm^{-1} [1]. In NIRS, the wavelength is measured in nanometer (nm), frequency in hertz (per second), and wavenumber in per centimeter (cm^{-1}). NIRS offers significant benefits, including rapid analysis, cost-effectiveness per test, environmental friendliness (as it eliminates the need for purchasing or disposal of chemicals), non-destructive, and versatility as an analytical method [2]. NIRS affords the user a level of precision equal to, but cannot be surpassing, that of conventional reference methods. Additionally, NIRS is entirely safe and adaptable for laboratory and field use [3].

Hydrogen, being the lightest atom, vibrates with a large amplitude during hydrogenic vibration. Thus, NIR radiation, typically ranging from 780 to 2500 nm, primarily interacts with hydrogen bonds like O-H, C-H, N-H, and S-H [4]. NIR radiation incident on a sample can be transmitted, absorbed, or reflected, and this information is captured by NIRS [5]. The portion of

the incoming light energy that exits the sample and reaches the detectors is the energy utilized in NIRS analysis. Accuracy in NIRS analysis refers to its proximity to the reference value. NIRS excels in predicting compositions based on constituents such as water, fat, protein, starch, cellulose, hemicellulose, lignin, and other abundant absorbers like O-H, C-H, N-H, and S-H [6, 7].

According to the law of conservation of energy, the total radiant intensity incident on the sample (I_o) must equate to the combined intensity absorbed (I_A), transmitted (I_T), and reflected (I_R) by the sample [8].

$$I_o = I_A + I_T + I_R \quad \text{-----}[1.1]$$

Absorption of radiation is explained with the equation

$$A = abc \quad \text{-----}[1.2]$$

where 'a' is a constant called absorptivity, 'b' is the thickness of the sample with which the radiation passes, and 'c' is the concentration of molecules in the sample [9].

Similarly, reflection of radiation by the sample is explained with the equation:

$$\text{Log } 1/R \sim ac / s \quad \text{-----}[1.3]$$

Where 'a' is a constant called absorptivity, 'c' is the concentration of molecules in the sample, and 's' is a scattering constant [9].

The basic elements of NIR instruments are a radiation light source, a monochromator, and a detector, allowing for transmittance or reflectance measurement. When monochromatic radiation is incident in a sample, the sample might have absorbed, transmitted or reflected the incident beams (refer to Figure 2.2).

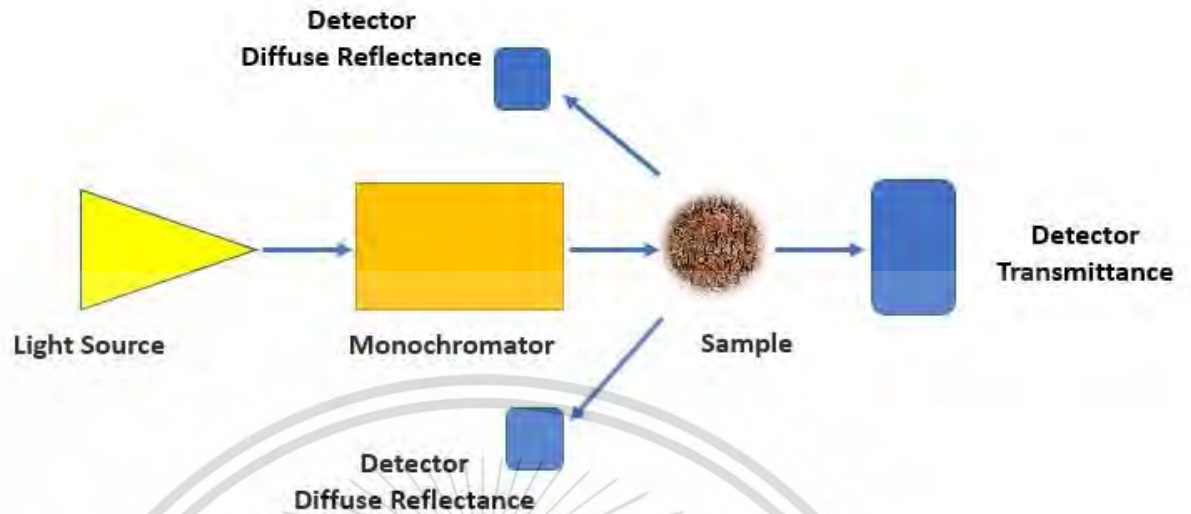


Figure 2.2: Basic NIR spectrometer configuration [1]

In an NIR instrument, the light source is essentially divided into two groups: broad waveband (such as tungsten halogen lamps) and narrow waveband (Light Emitting Diodes (LEDs) and laser diodes). The most commonly used light source in NIR instruments is the tungsten halogen lamp.

There are different available technologies for separating polychromatic light (such as sunlight or light from a lamp) into monochromatic light for both qualitative and quantitative analysis. These technologies include filter photometers, diffracting gratings, interferometers, diode arrays, and acoustic-optical tunable filters (AOTFs). Filter photometers or LEDs cover a narrow spectral range of 10-100 nm, while diffraction gratings, interferometers, diode arrays, or AOTFs provide full spectral coverage.

Detectors are utilized to convert the intensity of incident-radiated energy on samples into electrical signals. Detectors are classified based on their operational principle into thermal detectors and photon detectors. A thermal detector absorbs infrared energy through a temperature-sensitive surface. It exhibits uniform sensitivity across all wavelengths but has a relatively low response time. On the other hand, a photon detector converts incident light into an equivalent current in an external circuit in proportion to the light's intensity. The most common types of detectors include silicon, lead sulfide (PbS), and indium gallium arsenide (InGaAs). Silicon detectors are fast, low-noise, small in size, and highly sensitive from the visible region to 1100 nm. PbS detectors are slower but sensitive from 1100 to 2500 nm and offer good signal-to-noise

properties. InGaAs is an expensive detector that combines the speed and size characteristics of a silicon detector with the wavelength range of a PbS detector.

The selection of an NIR instrument is primarily based on the following criteria: wavelength precision, resolution, signal-to-noise ratio, and cost.

2.2 Sample preparation

The term “sample preparation” in NIRS refers to the sample placed in the spectrometer for scanning. Typically, there are four different types of sample preparations in NIR:

1. Transmission
2. Reflection
3. Transflection
4. Interaction

2.2.1 Transmission

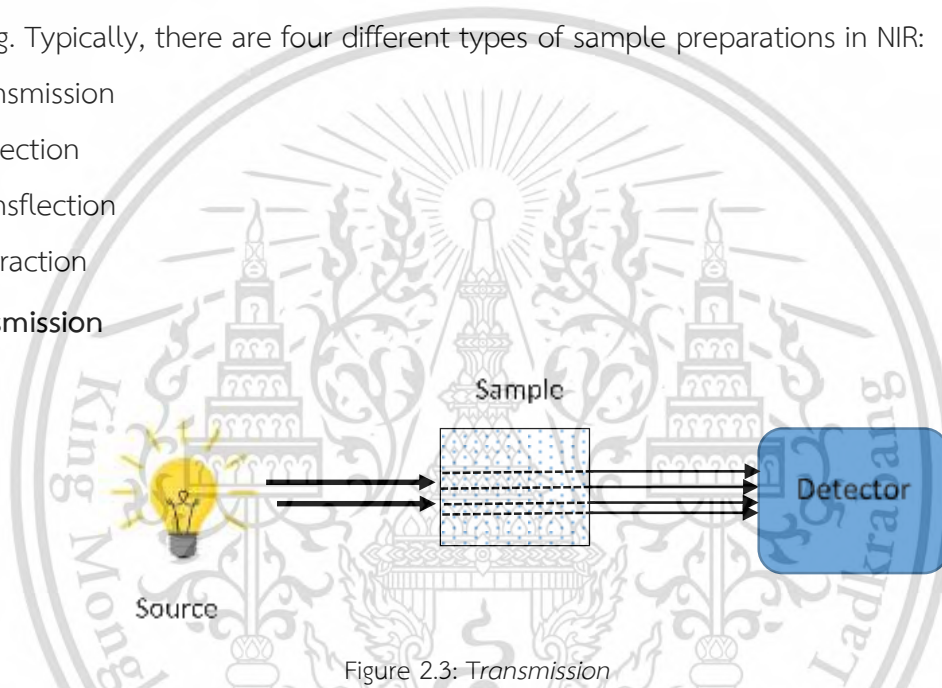


Figure 2.3: Transmission

In transmission, radiation passes through the sample and is detected by the detector, as shown in figure 2.3. The distance the radiation travels through the sample is called the path length. Light scatters within the sample due to its molecules. This scattering is more pronounced in solid samples and less so in liquids. If the path length is excessively long, the light absorbed within the sample becomes too high, preventing transmission or reflection. In these instances, the detector cannot capture the radiation, resulting in no spectrum being generated.

2.1.2 Reflection

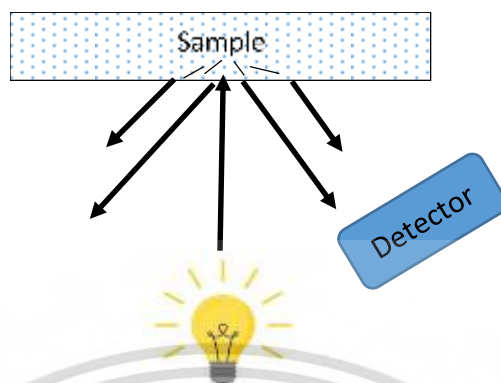


Figure 2.4: Diffuse reflection

As shown in the Figure 2.4, when the light beam reflects towards the source and the detector picks it up, it's referred to as reflection. This method is typically suitable for solid and semi-solid samples. Reflectance manifests in two forms: "specular" reflectance, which is surface-based and lacks object-specific information, and "diffuse" reflectance, originating within the object. Solid samples, exhibiting significant diffusion, convey valuable information about their composition and other attributes. Reflection, a prevalent sample preparation mode, finds extensive use in modern NIR instruments.

2.2.3 Transflection

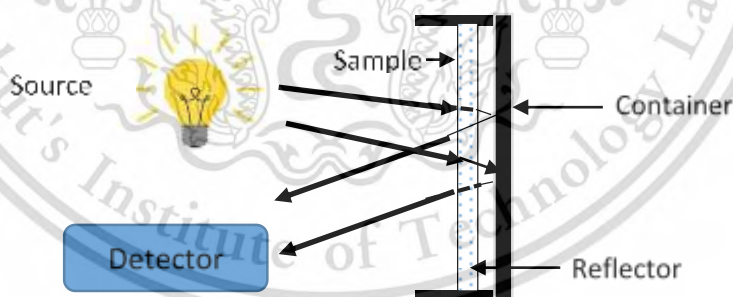


Figure 2.5 Transflection

Transflection, as shown in Figure 2.5, combines transmission and reflection methods. A thin reflector is placed under the sample cup, typically made of non-NIR absorbing materials like ceramic, aluminum, gold, or any substance capable of reflecting Near Infrared. In this process, the light beam initially passes through the sample, undergoing transmission. Subsequently, this transmitted light hits the reflector, leading to reflection that the detector picks up. Transflection is predominantly used for liquids (both thick and thin) and thin solid samples.

2.2.4 Interaction

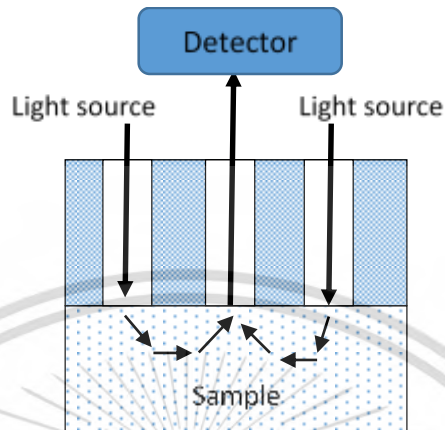


Figure 2.6 Interaction

In interaction mode, as shown in Figure 2.6, an optical fiber is employed. Light passes to the sample via the optical fiber and scatters within it. The scattered light is then reflected into the optical fiber and detected by the radiation sensor. This method is applicable to solid, semi-solid, and liquid samples.

2.3 Interferometer or Fourier transform (FT) Near Infrared Spectrometer

The interferometer is an optical device that controls the generation of interference patterns between two light beams. The schematic of one of the most common types of interferometers, the Michelson interferometer, is shown in the figure 2.7 below:

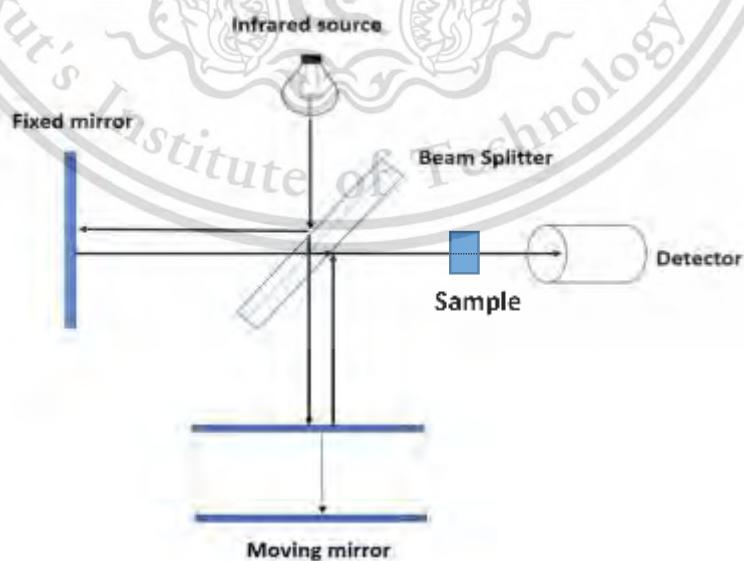


Figure 2.7 Optical diagram of Michelson interferometer

The Michelson interferometer consists primarily of a light source, a beam splitter, a fixed mirror, a movable mirror, and a detector. In NIRS, the emitted light falls within the NIR range. A fixed beam splitter, positioned at a 45° angle, divides the incoming light into two beams. These beams then reunite into a single beam after reflecting off two mirrors—one fixed and the other movable. The resulting combined light is directed towards the detector.

The fixed mirror and moving mirror are positioned perpendicular to each other. The movable mirror moves along an axis parallel to the light source. The beam splitter partially directs light to the fixed mirror and the remainder to the moving mirror. The reflected light from both mirrors returns to the beam splitter. When the mirrors are equidistant, the reflected beams interfere constructively, resulting in the highest beam intensity. Conversely, when the two light beams from the mirrors are out of phase, they interfere destructively, while fixed mirror is fixed and moving mirror is moving generating an interferogram. The fluctuation in NIR intensity while the moving mirror displaces contains spectral information obtained through a Fourier Transform. This transformation converts the signal from the time-domain into the frequency domain [10] creating spectrum.

2.4 Factor effecting spectrum generation

When a broad NIR radiation beam interacts with a sample molecule at energy state E1, the sample absorbs energy from the beam, inducing a transition from E1 to E2. This absorption results in a reduction in radiation intensity detected by the detector, described by $\nu = \Delta E / h$, where, ν represents frequency, ΔE indicates the energy change, and h denotes Planck's constant. This process generates a spectrum.

Different molecules within a material, each possessing distinct hydrogen bonds, absorb NIR radiation at different frequencies, leading to varied absorption intensities and multiple peaks in the spectrum. A molecule with just one hydrogen bond will absorb only a single peak. Variations in the scanning environment, such as ambient temperature and humidity, or the intrusion of water from the surrounding air into the spectrometer, have the potential to modify the NIR spectra acquired from scanned samples.

NIR spectra are highly sensitive to temperature variations. Elevated sample temperatures indicate a higher energy level within a molecule. Consequently, molecules in high-temperature

samples require less energy for absorption compared to those in low-temperature samples, even when the structure and concentration remain the same. Several studies have advised against temperature fluctuations when utilizing NIR spectra, primarily due to the susceptibility of water's weaker hydrogen bonding to temperature changes. Such variations can impact model accuracy.

Similarly, it is highly advisable to perform background scanning before scanning new samples in NIRS. This is essential because the instrument gauges the intensity of NIR light interacting with the molecular vibrations of the sample. Nevertheless, elements like the material of the sample holder, air, or other components along the instrument's optical path may absorb or scatter light to some extent. Thus, to compensate ambient influences and instrument-related issues in the measurement setup and to obtain accurate and meaningful spectral data from the sample, it is crucial to mitigate the impact of these background factors.

Likewise, NIRS encounters limitations when scanning raw materials of varying particle sizes, even when their chemical compositions are identical. Biomass, for instance, presents varying shapes and sizes that pose challenges in precise measurement. The variance in particle sizes leads to air gaps between them, causing NIR radiation absorption by trapped air during scanning. Consequently, this variance results in differing spectra for identical raw materials despite their identical chemical composition. Therefore, maintaining uniform and homogeneous particle sizes is strongly advised. Finer particles enable NIRS to accurately scan their actual absorbance values.

To achieve the highest accuracy in NIRS, the following measures should be considered:

1. Sample preparation: Ensure proper collection and storage of samples
2. Sample temperature: Maintain stable and uniform sample temperatures during sample scanning (experiments).
3. Particle shape and size: Consider the shape and size of sample particles.
4. Prevent NIR leakage: Ensure no NIR light leakage during sample scanning to avoid information loss.
5. Follow reference lab Methodology: Adhere to standard lab methodologies.
6. Regular maintenance: Perform regular calibration, and validation with unknown samples, and maintain the NIR instrument.
7. Careful instrument operation: Operate the instrument with precision and care.

8. Utilize appropriate modelling techniques: Employ suitable modeling techniques for analysis.

2.5 General procedure of NIRS

Samples designated for NIRS are chosen for scanning, acquiring optical data in either reflectance or transmittance mode. Subsequently, the scanned sample undergoes constituent measurement through a reference laboratory method. The optical and reference data correspond to X and Y matrices, respectively, and are used in suitable chemometric techniques to establish a calibration equation. NIRS requires precise sampling with minimal or no sample preparation. Spectra data pre-treatments are performed on sample data to enhance the calibration equation's performance. Absorbance values and reference data are determined for both the calibration and prediction sets.

Regression methods leverage absorbance values and reference values from the calibration set to derive the calibration equation. The computed reference values are based on the calibration equation of the calibration set, while absorbance values of the prediction set are used to validate the calibration equation's performance.

Expert analytical involvement is crucial in the NIRS measurement process for evaluation and method maintenance. Regular calibration of methods is necessary to ensure expected accuracy and validation of chemometric methods using independent test sets at regular intervals ensures reliability.

Changes in sample quality, environmental conditions (like temperature, humidity), or gradual instrument alterations can deteriorate analysis quality over time. Timely recognition of such changes through vigilant observation is vital, requiring planned maintenance methods. Experts regularly review and compare spectra with earlier calibration spectra to identify outliers, comprehend reasons behind them, and develop effective implementation and maintenance strategies [11].

NIRS technology offers several advantages [12]:

1. Swift results: Delivers outcomes within seconds rather than hours or days.
2. Comparable accuracy to reference methods.
3. Precision (reproducibility) on par or often superior to reference methods.

4. Cost-effectiveness: Low labor expenses and no need for purchasing or disposing of chemicals.
5. Versatility: Simultaneous testing of numerous constituents.
6. Environmentally friendly: No use of chemicals.
7. Easy and inexpensive installation: No need for drainage or exhaust systems.
8. Minimal or no sample preparation required.
9. Durability: Instruments function reliably for over 10 years.
10. User-friendly and safe operation.
11. Networking capability: Multiple instruments, even remotely located, can be networked to utilize the same calibration, managed from a single control center to monitor performance.

The drawbacks linked with NIRS technology [12] include:

1. Requiring separate calibration for each commodity and constituent.
2. Necessity for continuous monitoring of accuracy and reproducibility (similarly required in reference analysis).
3. High initial costs of instrument purchase, although they rapidly compensate for themselves in saved labor, installation, time, and chemical expenses.
4. Limited understanding of the most efficient operation methods for these instruments.

2.6 Recent application of NIRS

NIRS has broad multi-disciplinary applications in quality control, quality assurance, and real-time screening across various fields, including agriculture, energy, food, medicine, and many other sectors. Recent studies in diverse fields have successfully demonstrated the practical applications of NIRS for rapidly detecting and evaluating desired reference values in an environmentally friendly manner, effectively assisting in decision-making and problem-solving.

In the field of agriculture, Gorji et al. confirmed the determination of moisture content in Genovese basil (*Ocimum basilicum* L.) using handheld NIRS, offering valuable insights for irrigation and crop management [13]. Zhao et al. identified zearalenone content in wheat, affirming the potential of integrating a colorimetric sensor array with NIRS to evaluate the safety of wheat and

potentially other grains [14]. Brien et al. predicted the ripening of mangoes using NIRS to improve postharvest management practices, thus reducing commercial losses in the mango industry [15]. Yang et al. assessed the quality of peach kernel samples in situ using Micro-NIRS and FT-NIRS [16]. Phanomsophon et al. swiftly classified primary micronutrient levels in durian (*Durio zibethinus* Murray CV. Monthong) leaves using FT-NIRS to regulate the random application of fertilizers, aiming to enhance both the quantity and quality of durian fruit [17].

In the energy field, particularly in biomass, Yan et al. developed a model using NIRS and a deep learning algorithm for the online prediction of moisture content in corn stalks, corn cobs, rice straw, wheat straw, bark, and branches at power plant sites [18]. Skvaril et al. review the application of NIRS in the biomass energy conversion process, mainly in pelleting, combustion, gasification, pyrolysis, to monitor its physical, thermochemical, biochemical, and physiochemical properties [19]. Machini and Rinnan have developed NIRS as a tool for rapidly assessing waste wood quality, focusing mainly on moisture content and net calorific value, whether for the production of panel boards or as feedstock in combustion units. This promotes the reuse of waste products and increases the share of renewable energy for energy applications [20].

In the food sector, Pan et al. developed a calibration model aimed at predicting the moisture, fat, and protein content of pork, thereby ensuring the maintenance of its quality in the meat industry through the utilization of NIRS [21]. Millatine et al. detected and quantified the percentage of adulteration in cocoa powder, primarily intended for application within the food industry for the preparation of bread, biscuits, pastry filling, and ice cream [22].

In medical science, NIRS finds wide application across various domains. It is utilized in clinical chemistry for sensing and monitoring changes in the body's homeostasis, analyzing biofluids and tissues from numerous organs [23]. Moreover, it facilitates continuous monitoring of cerebral oxygenation, particularly during cardiac surgery [24]. Additionally, NIRS has been extensively used as a tool for quality control in traditional Chinese medicine [25].

Recent studies also highlight the application of NIRS in the aviation industry. For instance, Li et al. conducted rapid detection of multiple physicochemical properties of jet fuel, directly influencing the safety and operational efficiency of aircraft [26].

2.7 Biomass

Biomass, an organic material generated by plants and animals, serves as a carbon-neutral fuel and is renewable [27]. It stands as the sole renewable energy source available in solid, liquid, or gaseous states [28]. This resource harbors stored chemical energy from the sun, which can be directly combusted for various thermal applications, ranging from household woodstoves to industrial combustion technologies, predominantly for power generation and co-firing with fossil fuels [29].

Despite its potential, biomass faces challenges hindering its smooth application, such as moisture content, low energy density, and a complex structure. Biomass is composed of holocellulose (cellulose + hemicellulose), lignin, extractives, mineral materials, and moisture [30]. Cellulose, hemicelluloses, and lignin form the macromolecular cell wall components, while extractives and minerals are minor low molecular weight components in wood biomass [31]. Biomass molecules consist of elemental components like C, H, O, N, S, etc [32]. Cellulose comprises glucose monomers, hemicellulose consists of hexose and pentose sugars, and lignin contains ether and carbon-carbon linkages [33]. The proportions and chemical compositions of lignin and hemicelluloses differ in lignocellulosic biomasses, while cellulose is a relatively consistent component [34]. Typically, biomass contains cellulose, hemicellulose, and lignin in the range of 40-60 wt%, 20-24 wt%, and 10-25 wt%, respectively, on a dry basis [31].

The rate of decomposition is faster in biomass with higher cellulose and hemicellulose content, characterized by narrower temperature ranges, while higher lignin levels lead to slower degradation with wider temperature ranges (ambient to about 600°C) [32]. Lignin, the most stable and complex among these components, is considered the primary controlling factor in the thermochemical decomposition process [35]. Acting as a good antioxidant and cross-linker, lignin facilitates bonding between cellulose and hemicellulose. Studies suggest that an increase in lignin content enhances the unsaturated carbon content and elevates the heating value [34]. However, understanding the influence of cellulose and hemicellulose on the heating value remains a significant challenge in identifying the entire lignocellulosic component impacting plant thermal properties [31].

The physical and chemical attributes of biomass critically influence the design and operation of associated biomass conversion facilities. These properties vary with plant species,

growth location, conditions, harvesting, and storage, posing challenges in accurately and consistently measuring biomass properties. Elemental analysis, proximate analysis, and thermal analysis regarding heat content stand as crucial data for biomass characterization [34].

Biomass, absorbing carbon dioxide during growth and emitting it during combustion, aids in atmospheric carbon dioxide recycling and does not significantly contribute to a net greenhouse effect [36]. Profound knowledge of biomass and its properties can save energy, reduce fossil fuel consumption, and limit the emission of CO₂, SO_x, NO_x, and heavy metals into the environment.

2.8 Physical properties

The physical properties of biomass primarily encompass particle size, density (both particle and bulk), grindability, flowability, and thermal properties [37]. These values vary significantly, where attributes like density, porosity, and internal surface area are specific to biomass species, while bulk density, particle size, and shape distribution are associated with fuel preparation methods [38].

2.8.1 Particle size –

Biomass particles come in various irregular shapes and lack uniformity in size, exhibiting forms like round, leaf-shaped, needle-like, triangular, etc. This diversity poses challenges in accurately measuring their dimensions, including length, width, and thickness. The aspect ratio (length/diameter) serves as a critical parameter defining the particle and is mathematically expressed as:

$$\text{Aspect ratio} = \frac{b}{l} \text{ ----- [1.4]}$$

Here, ‘b’ represents the width, defined as the minimum distance between two parallel lines tangential to the particle's projected outline, while ‘l’ denotes the length, corresponding to the maximum distance between two parallel lines perpendicular to the tangents for the width.

Different shapes and sizes of biomass particles, which correspond to different surface areas, significantly impact the efficiency of conversion and energy requirements—particularly concerning heat and mass transfer during thermal conversion [39]. Particle size serves as a crucial determinant of biomass quality and density. In the production of biomass end products such as pellets, particle size profoundly influences mechanical strength. During compression, the initial layer of particles tightly interacts with the subsequent layer, pressing against each other and filling

the successive gaps. As multiple layers compress and accumulate, the particles oppose one another, leading to increased inter-particle stress [40]. The characterization of particle size commonly employs two extensively used analysis methods: sieve shaker and digital imaging particle analysis [39].

2.8.2 Density

Density is defined as mass of the substance divided by its volume (kg/m^3).

$$\rho_p = m_p / V_p \quad \text{-----}[1.5]$$

Where, ρ_p = the particle density, m_p = particle mass and V_p = the particle true volume.

Biomass density is influenced by factors such as moisture content, particle shape and size, and surface properties. In the bioenergy domain, two types of densities are considered: 1) Particle density and 2) Bulk density. Particle density signifies the total mass of all particles divided by the volume they occupy, excluding pore space volume. On the other hand, bulk density (ρ_b) represents the mass of particles relative to the total volume they occupy, encompassing the pore space volume. Bulk density is contingent upon biomass particle characteristics like shape, size, moisture content, particle density, surface features, loading and emptying methods, and operating conditions. Moreover, the bulk density of biomass significantly hinges on tapping, whether it's in a loose fill or tight fill state [39].

Pore spaces in bulk sample is explained by the term porosity, which is define by the following formula

$$\epsilon_o = 1 - (\rho_b / \rho_p) \quad \text{-----}[1.6]$$

Where, ϵ_o = porosity of bulk biomass, ρ_b = bulk density and ρ_p = Particle density

The density of biomass varies according to its lignin content. Biomass tends to have lower density with lower lignin content and vice versa [37].

2.8.3 Grindability

Grindability denotes the energy needed to reduce an infinite particle size to 80% passing 100 μm . Due to their fibrous cellulose and lignin content, biomass presents challenges in grinding, and thus, biomass grindability primarily indicates its resistance to grinding. This process consumes energy and relies heavily on factors like the initial and final biomass particle size, moisture content, material properties, mass feed rate, and the types of grinding machines used [37].

2.8.4. Flowability

Flowability describes the flow behavior of biomass, indicating how efficiently biomass moves from one point to another. This property holds significance for the transportation, storage, and handling of biomass. Assessing biomass flowability involves parameters such as the angle of repose, cohesion coefficient, compressibility index, and flow index, all of which are influenced by the sphericity or shape factor of biomass particles.

Table 2.1 Engineering application of physical properties of lignocellulosic biomass [37]

Physical properties	Engineering application
Particle size	Design parameter for feeding, grinding and storage facilities
Density	Design parameter for handling, storage and transportation facilities
Grindability	Design parameter for grinding facilities
Flowability	Design parameter for handling, feeding and storage facilities

2.9 Calorific value

Calorific value (CV), also referred to as heating value, signifies the total energy released during combustion. It measures the heat energy liberated per unit mass of the fuel when completely burned. Expressed in Megajoules per kilogram (MJ/kg), Joules per kilogram (J/kg), or calories per gram (cal/g), CV is a crucial parameter characterizing a substance's combustibility and is instrumental in the design, operation, and calculations within thermal power plants [41, 42]. Understanding the CV aids in optimizing biomass usage for process efficiency. A higher CV denotes superior fuel quality. Consequently, the CV of biomass becomes a crucial trading parameter, which varies based on, not limited to, proximate parameters such as moisture content (MC), volatile matter (VM), fixed carbon (FC), and ash content (A), as well as ultimate analysis parameters like C, N, H, O, and S.

During combustion, a crucial liquid-vapor phase change occurs. Before a liquid fuel droplet burns, it vaporizes, and if adequately cooled, water vapor can condense from the combustion products. This leads to the division of CV into two types: higher heating value (HHV) or gross calorific value (GCV) and lower heating value (LHV) or net calorific value (NCV). HHV includes the latent heat from condensed water vapor in the final combustion products of biomass. When the steam is not condensed, the resulting value is the LHV from biomass combustion. LHV represents

the heat stored in biomass, excluding the latent heat of water vaporization, while HHV combines LHV with the latent heat of vaporization of water [39].

The relationship between HHV and LHV is expressed as;

$$\text{LHV}_{\text{wb}} = \text{HHV}_{\text{wb}} - 2.443 \times \text{MC\%} \quad \text{-----[1.7]}$$

where, MC is percent moisture content expressed on a wet basis (wb).

HHV of samples can be theoretically determined using the modified Dulong's formula as mentioned in equation 1.8.

$$\text{HHV (MJ/kg)} = [(33.5 \times \text{wt. \% C})/100] + [(142.3 \times \text{wt. \% H})/ 100] - [(15.4 \times \text{wt. \% O})/ 100] \text{ --} \\ \text{---[1.8]}$$

where, C, H and O are carbon, hydrogen and oxygen in weight percentage of biomass.

The HHV of biomass can be determined using a bomb calorimeter [43], a process that is both destructive and time-consuming, requiring more than 30 minutes per sample, including sample preparation and preprocessing [44].

2.9.1 Bomb calorimeter

The bomb calorimeter stands as the standard method for determining the HHV of biomass, typically requiring about 15 to 30 minutes [44] to yield values in kJ/kg, cal/g, etc., and it operates through a destructive process. Figure 2.8 show the construction of bomb calorimeter.

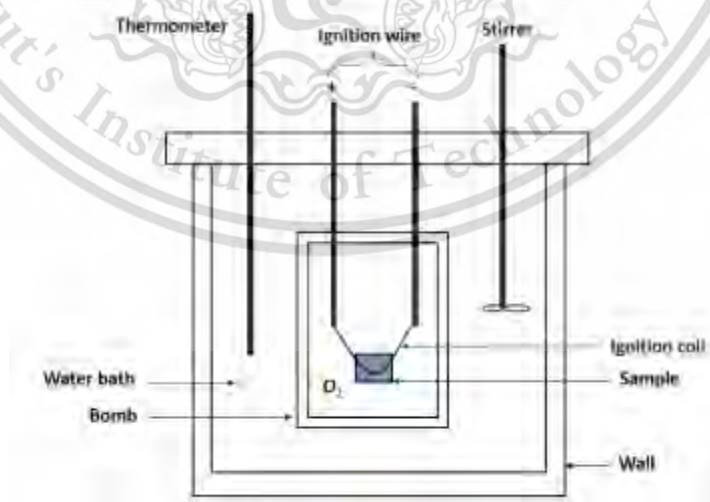


Figure 2.8 Construction of bomb calorimeter

Essentially, the bomb calorimeter comprises a crucible for holding the biomass sample, oxygen, a stainless steel bomb, water, a stirrer, a thermometer, an insulating container, and an ignition circuit linked to the bomb. The bomb is tightly sealed at its top to ensure gas-tight conditions, while the stirrer maintains uniform heat distribution in the water jacket. A thermometer tracks the water temperature changes, and the insulating container prevents heat dissipation. The ignition circuit generates a spark for fuel ignition. When utilizing a stainless steel bomb, the reaction occurs without any observable volume changes.

This device, a constant-volume calorimeter, measures the heat generated during a specific reaction, requiring resilience against high pressure within the chamber. Electrical energy initiates the fuel combustion, heating the surrounding air, which then expands and exits through a tube leading out of the calorimeter. As the air flows through the copper tube, it heats the water surrounding it. The change in water temperature allows for the calculation of the fuel's HHV using the formula:

$$\text{HHV} = m \cdot c_p \cdot \Delta T \quad \text{-----[1.9]}$$

Where HHV represents the biomass's Higher Heating Value, 'm' is the water mass (kg), 'c_p' denotes the specific heat of water (J/kgK), and ΔT signifies the change in water temperature (K).

Bomb calorimeters are generally classified into two categories: Isoperibol and adiabatic calorimeters. Isoperibol calorimeters maintain a constant temperature environment, with temperature changes governed by the thermal resistance between the calorimeter and its surroundings. Adiabatic calorimeters strive to minimize heat exchange between the calorimeter and its surroundings by maintaining a small temperature difference and a large thermal resistance.

In this study, the HHV of the biomass is evaluated using the isoperibol method with an automatic bomb calorimeter (IKA C 200, Staufen, Baden-Württemberg, Germany) [30], as shown in Figure 2.9.



Figure 2.9 Automatic Bomb calorimeter (IKA C 200, Staufen, Baden-Württemberg, Germany)

Note: Refer Chapter 3 /Journal paper titled “Comprehensive assessment of biomass properties for energy usage using near-infrared spectroscopy and spectral multi-preprocessing techniques” for detailed information about HHV and bomb calorimeter.

2.10 Proximate analysis

In the process of biomass combustion, the primary factors influencing the rate of heat release are the parameters of proximate analysis, which encompass moisture content (MC), volatile matter (VM), fixed carbon (FC), and ash content (A) [45, 46]. Among these parameters, MC stands as one of the significant characteristic of biomass. It directly influences the net calorific value and combustion efficiency [47]. Higher MC in the fuel translates to lower heating values and vice versa. The determination of MC typically follows the ASTM E-871-82 (2019) standard [48]. The process involves heating the sample in a muffle furnace at around $105^{\circ}\text{C} \pm 2^{\circ}\text{C}$ for at least 3 hours until its weight stabilizes, reflecting the moisture content in the fuel:

$$\% \text{ MC} = (m_i - m_f) / m_i \text{-----}[1.10]$$

Here, MC represents the moisture content of the biomass sample on a % wet basis, 'm' denotes the sample mass, and the subscripts 'i' and 'f' signify initial and final values, respectively.

VM refers to biomass components that transform into a gaseous form through thermal decomposition. The determination of VM adheres to ASTM E-872-82 (2019) standards [49], entailing heating the biomass sample in a crucible within a muffle furnace at 950°C for 7 minutes. The weight loss post-de-volatilization represents the VM [39]. A high VM content doesn't necessarily ensure a high CV due to constituents derived from non-combustible gases like CO_2 and H_2O .

A, an inorganic residue remaining after the ignition of combustible substances [50, 51], influences the CV by extracting energy required for the thermal breakdown and phase transition of inorganic ash-forming materials [52]. Elevated MC and ash levels in biomass lead to ignition and combustion issues [53], influencing corrosion and slag formation [54]. ASTM E1755-01 (2020) [55] addresses ash content in biomass determination.

FC constitutes the non-volatile fraction of biomass. Lower MC and VM contents contribute to an improved FC in biomass samples [56]. Higher FC indicates a greater carbon content in the biomass, yielding higher energy content and maximal heat during combustion [57]. Therefore, the FC content of biomass positively influences its energy potential and CV. FC is typically derived by subtracting the sum of MC, VM, and ash content from 100% [51, 58].

$$FC (\%) = 100 - (MC + VM + A) \quad \text{-----}[1.11]$$

Assessing the proximate parameters typically demands a substantial investment of time and financial resources, and is susceptible to experimental inaccuracies. Additionally, the process entails analyzing a large number of samples to determine biomass characterization data according to ASTM standards for proximate analysis, rendering it both laborious and destructive. Consequently, numerous researchers have advocated for the utilization of TGA (Thermogravimetric Analysis) for conducting proximate analysis of biomass. Therefore, to evaluate the proximate analysis parameters of biomass, TGA (TG 209 F3 Tarsus, Netzsch, Germany) with a microbalance sensitivity of 0.1 μg resolution was employed under direct combustion in air atmosphere, i.e., O_2 and N_2 in 1:4 ratio, within a temperature range from 35 $^\circ\text{C}$ to 700 $^\circ\text{C}$ [59].

Figure 2.10 shows the TGA (TG 209 F3 Tarsus, Netzsch, Germany) used to evaluate the proximate analysis parameter of biomass.



Figure 2.10 TGA (TG 209 F3 Tarsus, Netzsch, Germany)

Note: Refer Chapter 5 /Journal paper titled “NIR Spectroscopy as an Alternative to Thermogravimetric Analyzer for Biomass Proximate Analysis: Comparison of Chip and Ground Biomass Models” for detailed information about proximate analysis.

2.11 Ultimate analysis

The ultimate analysis provides essential data about the elemental composition of biomass, measuring the weight percentage (wt.%) of carbon (C) hydrogen (H), nitrogen (N), sulfur (S), and oxygen (O) [46]. This analysis quantifies these elements on a dry basis within the biomass, necessitating equipment handled by experts and generally deemed expensive [60]. Specifically, the CHNS/O elemental analyzer measures the wt.% of C, H, N, and S within the biomass sample. To complete the ultimate analysis, the wt.% of O on a dry basis in the biomass sample was calculated as the difference [30, 61].

The calculation for the wt.% of O is as follows:

$$\text{wt.\% O} = 100 - (\text{wt.\% C} + \text{wt.\% H} + \text{wt.\% N} + \text{wt.\% S} + \text{wt.\% A}) \text{ -----[1.12]}$$

However, if the ultimate analysis is carried out without drying the samples at 105 °C prior to CHNS analysis, the oxygen (wt. %) is estimated as:

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

--[1.13]

Here, the determination of proximate parameters (MC and A) within the biomass was obtained using equipment typically found in a standard laboratory, including drying ovens, muffle furnace, or a thermogravimetric analyzer.

Carbon, the primary component of biomass, typically constitutes around 50% depending on the biomass type [62]. Oxygen follows as the second most prevalent element at about 40%, playing a significant role in the combustion process [63]. Hydrogen, around 6% in biomass, serves as an important component of many biofuels which can be utilized for heat and electricity generation. Nitrogen, though a minor component (typically 1-2%), influences combustion efficiency and emission during combustion [64]. Biomass also contains trace amount of S, which, although minimal, can contribute to emissions such as sulfur dioxide [65]. Therefore, the elemental composition of biomass has a profound impact on combustion efficiency and the levels of emission released into the environment. These emissions, in turn, carry significant consequences for both the energy industry and the natural surroundings. Energy release during biomass combustion correlates positively with C and H contents, as they are the primary contributors to its energy value [66]. High C content is desirable for energy production [57], and H high energy content adds value [67]. During combustion, O reacts with C and H, reducing the available energy in biomass. Elevated O and N contents decrease the CV, thereby reducing energy potential [68]. N and S are undesirable elements in biomass due to their contribution to the formation of harmful NO_x and SO_x [69, 70]. To minimize environmental impact and ensure sustainable operation and maintenance of combustion systems, low S content in biomass is preferred [70]. This comprehensive assessment provides a detailed understanding of the elemental makeup of the biomass.

In this study, the wt.% of C, H, N, and S on a dry basis was determined at the Scientific and Technological Research Equipment Center (STREC) at Chulalongkorn University, Bangkok, Thailand, using CHNS/O analyzer (Thermo Scientific TM FLASH 2000, Waltham, MA, USA).

Note: Refer Chapter 3 /Journal paper titled “Comprehensive assessment of biomass properties for energy usage using near-infrared spectroscopy and spectral multi-preprocessing techniques” for detail information about HHV and bomb calorimeter” and to “Effect of combined non-wood and wood spectra of biomass chips on rapid prediction of ultimate analysis parameter using near infrared spectroscopy” for detailed information about ultimate analysis.

2.12 Combustion

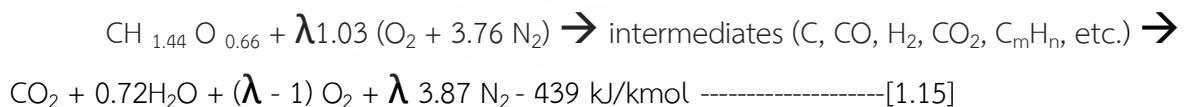
The fulfillment of global primary energy relies mainly on the direct combustion of the biomass [71]. Combustion is an exothermic chemical process of involving oxidation, reduction, and the liberation of heat. Combustion requires an oxidizing agent and O is one of the commonly used oxidizing agent that is naturally present all over the world. Numerous studies have demonstrated that oxygen-enriched atmosphere improves the combustion efficiency. During combustion, when fuel is burned, O in the combustion air chemically combines with H and C in the fuel, forming water and carbon dioxide, respectively. This process releases heat and often light in the form of either a glow or flame [72].

In complete combustion, when biomass are burned, the primary products are typically the most common oxides. For example, C yields carbon dioxide, N yields nitrogen oxide, and S yields sulfur dioxide.



N, being inert, does not actively participate in combustion. However, at high temperatures, some N may convert into nitrogen oxides. The minimum amount of air needed for the complete combustion of a fuel is referred to as stoichiometric or theoretical air. Achieving 100% combustion efficiency in practice is not feasible [73]. If supplied with less than theoretical air, incomplete combustion occurs, resulting in products including carbon monoxide. Therefore, in normal practice, an excess of air, i.e., more than theoretical air, is supplied, leading to O appearing in the products.

The main combustion parameter is the excess air ratio (λ), which describes the ration between the locally available and the stoichiometric amount of combustion air. For typical biomass, the combustion reaction can then be described by the following equation if fuel constituents such as N, potassium, chlorine, etc., are neglected:



Here, $\text{CH}_{1.44} \text{O}_{0.66}$ describes the average composition of typical biomass used for combustion, such as wood, straw, or similar materials [74].

To comprehend biomass combustion, it is crucial to grasp the properties of biomass that dictate its behavior as a fuel for combustion. These properties can be conveniently categorized

into physical, chemical, and thermal properties. Physical properties exhibit significant variability; characteristics like density, porosity, and internal surface area are linked to biomass species, while bulk density, particle size, and shape distribution are associated with fuel preparation methods. Chemical properties for combustion include ultimate analysis, proximate analysis, higher heating value, heating value of the volatiles, and heating value of the char. Thermal properties, such as specific heat, thermal conductivity, and emissivity, vary according to MC, temperature, and degree of thermal degradation. It is noteworthy that the thermal degradation sequence is as follows: hemicellulose > cellulose > lignin [75]. Thermal degradation products of biomass encompass moisture, volatiles, char, and ash. Volatiles further subdivide into gases like light hydrocarbons, carbon monoxide, carbon dioxide, hydrogen, moisture, and tars. Some properties vary with species, location within the biomass, and growth conditions, while others are contingent on the combustion environment [76].

A comprehensive understanding of combustion properties across various biomass types is crucial for selecting appropriate biomass and designing efficient combustion systems. Therefore, parameters like biomass ignition time (t_i) and ignition temperature (T_i), burnout time (t_f) and burnout temperature (T_f), maximum and average combustion rate, among others, play a vital role in evaluating combustion performance indices such as ignition index (D_i), burnout index (D_f), comprehensive combustion performance (S_i), and flammability index (C_i). Accurate assessment of these indices can improve the overall efficiency of biomass combustion systems, mitigate environmental impacts, and propel us toward a sustainable energy future powered by renewable sources [77].

In this study, TGA (TG 209 F3 Tarsus, Netzsch, Germany) has been employed to determine combustion characteristic parameters for evaluating different combustion performance indices. Biomass combustion in TGA primarily involves three stages: (i) water evaporation, (ii) release of volatiles and their combustion, and (iii) char combustion. TGA records the mass loss of biomass over time and temperature. Consequently, the thermogravimetric (TG) curve derived from TGA reveals the biomass sample's mass loss during thermal decomposition and combustion. The DTG curve, obtained from the derivative of the TG curve, provides additional insights into the rate of mass loss at different times and temperatures [15].

2.13 Thermogravimetric Analyzer

The properties of biomass change with variations in ambient conditions such as temperature and relative humidity. Thermogravimetry is a thermal analysis technique wherein the change in the sample's weight is recorded as a function of time or temperature in a defined and controlled environment (including heating rate, gas temperature, flow rate, and crucible type) [78]. A thermogravimetric analyzer (TGA) is sensitive and operates within a range from room temperature to 1000°C, with heating rates varying from 0.001K/min up to 50 K/min. In TGA, a thermobalance with a resolution of 0.1 µg is assembled to measure the instantaneous change in mass, while thermocouples are instrumented to make contact with the crucible made of aluminum oxide (Al₂O₃). A shielded heating element, based on nickel and stainless steel resistance heaters, is used as the heating source, and water is circulated in the system to cool the furnace.

In these techniques, a small sample of the fuel to be tested is placed in a crucible and heated in a specified atmosphere at the desired rate in an electronic microbalance. This provides a continuous record of the weight change of the fuel sample in a thermogravimetric (TG) apparatus. The differential thermogravimetry (DTG) apparatus continuously provides the rate of change in the weight of the fuel sample [46].

N₂ is an inert gas that does not interfere with the sample during thermal treatment. Under an O₂ atmosphere, C starts burning and produces carbon monoxide and carbon dioxide, causing a loss of sample mass. Therefore, TGA instruments use N gas as a protective gas and either O₂ or N₂ as a controlled atmosphere for oxidizing or inert atmospheres, respectively.

In this study, the TGA (TG 209 F3 Tarsus, Netzsch, Germany) is configured with a microbalance sensitivity of 0.1 µg resolution to mimic biomass direct combustion in air, employing O₂ (99.7%) and N₂ (99.99%) in a 1:4 ratio. The biomass samples undergo combustion in a 6.8 mm diameter aluminum oxide (Al₂O₃) crucible within a temperature range spanning from 35°C to 700°C, with a heat flow rate set at 10°C/min. Initially, the samples are isothermally maintained at 35°C for a duration of 10 minutes. N₂ (99.99%) serves as a protective layer in TGA to establish a stable and inert environment, shielding the sample and preventing unwanted reactions with the surrounding air during thermal analysis, thus ensuring accuracy [34]. Proteus 6.0.0 (NETZSCH software, Germany) is employed to generate TG and DTG curves. Utilizing these curves, biomass

proximate parameters (MC, VM, FC, and A) and combustion performance indices D_i , D_f , S_i , and C_i are determined [77].

Figure 2.11 shows the TGA (TG 209 F3 Tarsus, Netzsch, Germany) setup for experimentation to measure the combustion characteristics of biomass.



Figure 2.11 a) TGA (TG 209 F3 Tarsus, Netzsch, Germany), b) TGA water heater circulator, c) Nitrogen (99.9 %) cylinder, and d) oxygen (99.7 %) cylinder

Note: Please refer to Chapter 7 / journal paper titled “Near-infrared Spectroscopy Modelling of Combustion Characteristics in Chip and Ground Biomass from Fast-Growing Trees and Agricultural Residues” for detailed information about the combustion characteristic parameters of biomass.

2.14 Spectroscopy Quantitative Modeling

Regression analysis is the process of developing a quantitative model from existing data to predict a response. It can describe and interpret the relationship between the X variable (spectral data, independent) and the Y variable (reference data, dependent) and predict the Y values of new samples from the values of the x variables. The regression model is generally used for future prediction, ultimately replacing expensive and/or destructive testing in the long term.

The basic NIRS quantitative modelling methodology is illustrated in Figure 2.12 as below:

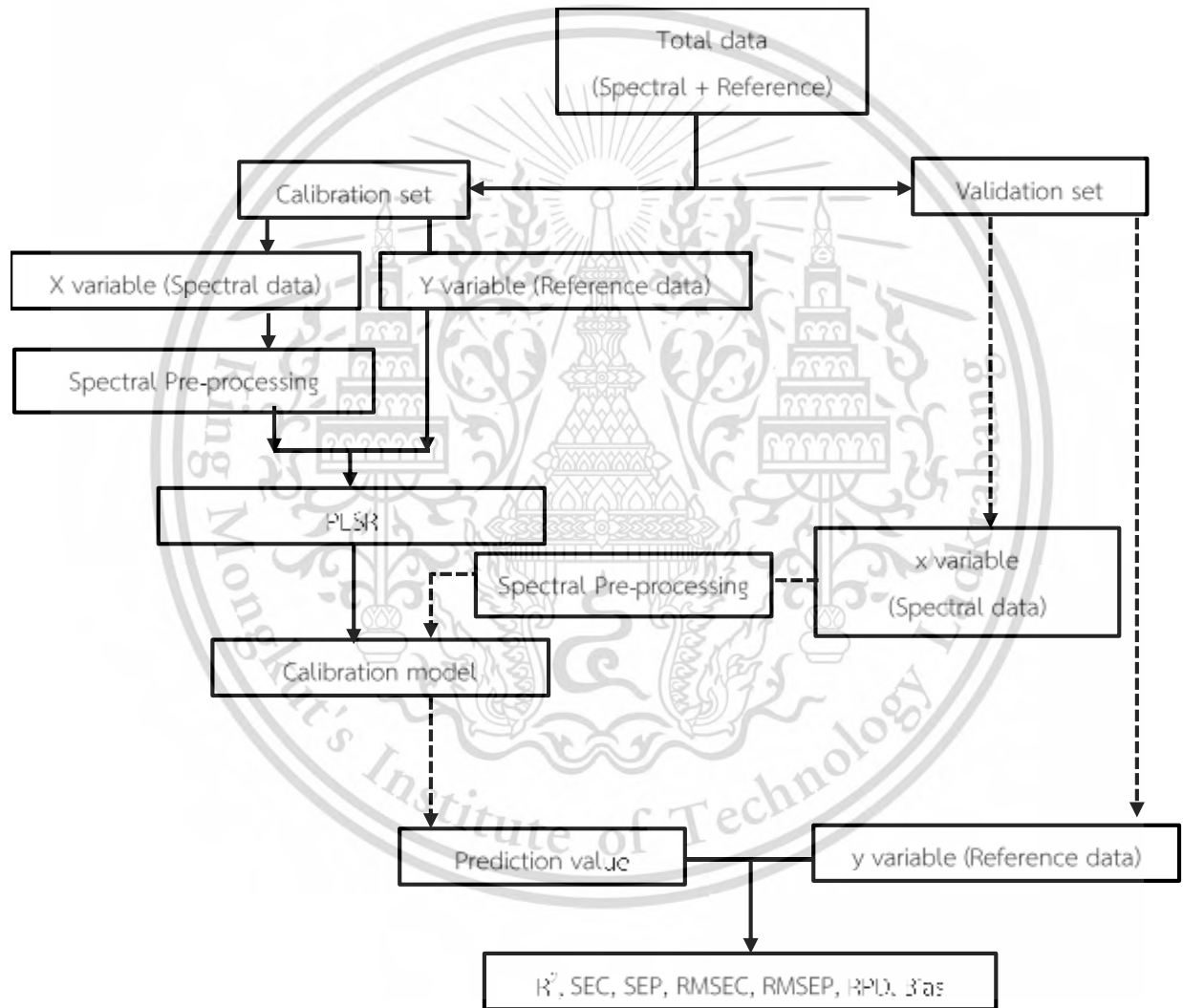


Figure 2.12 Flowchart of NIR quantitative modeling

In quantitative modelling, the following procedures are followed to develop a prediction model using a regression model:

1. Sample collection covering future sample and its appropriate storage

2. Sample scanning and reference data measurement
3. Data separation for calibration set and validation set
4. Spectral data preprocessing
5. Modeling
6. Validation and model performance analysis
7. Model updating

2.14.1 Sample collection covering future sample and its appropriate storage

NIRS requires little to no sample preparation to collect spectral data covering future sample. However, proper sample collection and appropriate storage play vital roles in modeling. Samples should be free from any contaminants and should be representative. After collection, they should be immediately stored in air-sealed aluminum/plastic bags for storage and transportation to the laboratory for experimentation. These bags protect the samples from the exchange of air and moisture [2]. The samples must be stored in a dry, air-conditioned room at an ambient temperature, generally $25\pm 2^\circ\text{C}$ (approximately), until scanned via FT-NIRS and reference data measured using respective instruments, such as HHV measured via bomb calorimeter, proximate analysis data obtained through the TGA, etc.

2.14.2. NIR sample scanning and reference data measurement

Initially, the samples are scanned using an FT-NIR spectrometer to obtain raw spectral data. For biomass, NIR scanning is generally conducted in diffuse reflectance within a selected wavelength range (full wavelength: $12500 - 3600 \text{ cm}^{-1}$) and a resolution of 16 cm^{-1} . The sample scan time and background scan time are set to 32 scans (on average) to collect the spectra. Reference data, such as HHV, proximate analysis parameters, ultimate analysis parameters, and combustion performance indices, are determined immediately after NIR scanning. Samples should be collected from the bottom of the sample scanning container if scanning is done under container through quartz window for better analysis. If scanning was done on top, the top portion material must be taken. Prior to model development, outliers in the reference data are identified using the following equation:

$$\frac{(X_i - \bar{X})}{SD} \geq |\pm 3|, \text{-----} [1.16]$$

Where, X_i is the measured value of sample i , \bar{X} and SD are the average and standard deviation of the measured values of all samples, respectively. If equation 1.16 is satisfied for

sample i , the sample will be considered as an outlier and is advised not to be included in a model for the analysis. Hence, such sample data will be removed from the total data set [30, 79].

Similarly, repeatability (**Rep**) and reproducibility tests must be conducted for both the NIR scanning and reference laboratory. Reference laboratory repeatability is determined by the standard deviation (**SD**) of the difference between duplicates of different samples. It is used to assess the operator's precision and to verify whether the reference method procedure can be consistently replicated. Reference laboratory reproducibility, on the other hand, is determined by the **SD** of the difference between duplicates obtained from blind samples. In this process, one of the duplicates has been previously analyzed, while the other is a blind sample. The maximum coefficient of determination (R_{\max}^2) is calculated as:

$$R_{\max}^2 = \frac{SD_y^2 - \text{Rep}^2}{SD_y^2} \text{-----} [1.17]$$

Where SD_y is the standard deviation of measured value. R_{\max}^2 is used to estimate the possibility of the modeling. If there are no errors in the spectra in modeling, the coefficient of determination R^2 is close to R_{\max}^2 .

Similarly, scanning repeatability is the process used to indicate the precision of the spectrometer or measuring instrument. In the scanning repeatability process, the NIR scan sample is positioned at the same location 'n' times (at least 10 times) to generate a spectrum of absorbance with respect to wavelength. The absorbance value of the scanned sample is recorded at each selected wavelength, and the SD is calculated. The average of all SDs of absorbance values at each wavelength is then calculated. For this calculation, the entire wavelength range can be selected, or wavelengths can be chosen randomly.

Scanning reproducibility is the process used to confirm the homogeneity of the sample. The reproducibility of spectral data is calculated as the SD of absorbance values when the sample is re-loaded and re-scanned 'n' times (at least 10 separate times). In this process, the absorbance values of the re-scanned and re-loaded sample are recorded at each selected wavelength, and the SD is calculated. Subsequently, the average of all SDs of absorbance values at each wavelength is determined.

2.14.3 Data separation for calibration set and validation set

After performing FT-NIR scanning and collecting reference data, the spectral data must be loaded, and corresponding reference values for individual components must be entered. As depicted in Figure 2.12, the total dataset is divided into a calibration set and a validation set. The NIR scanning and reference data for the calibration set are generally denoted by 'X' and 'Y' respectively. Similarly, the NIR scanning and reference data for the validation set are denoted by 'x' and 'y' respectively. Since the model must be developed based on the calibration set, it should cover the widest range of reference data. Therefore, the calibration set should include the highest and lowest reference values. For sufficiently large datasets, it is advisable to divide both the calibration set and validation set into an equal number of spectra. However, if the available data is insufficient, the total dataset is divided into calibration and validation sets using ratios such as 80:20 (in general), 70:30, 60:40, and so on. It is recommended that if the total dataset is limited to 20-50 samples, the model should be validated by cross-validation [80].

2.14.4. Spectral data preprocessing

Once the total dataset is divided into the calibration and validation sets, the spectral data from both sets undergo preprocessing using various techniques. These techniques include but are not limited to, no preprocessing (raw data), first derivatives, second derivatives, normalization, mean centering, baseline offset, multiplicative scatter correction (MSC), standard normal variate (SNV), or combinations thereof, such as first derivative + SNV, first derivatives + MSC, etc. Preprocessing aims to eliminate or reduce noise contributions, address baseline shifts, standardize spectra characteristics, mitigate light scattering effects in samples, and minimize errors in experiments not rectified by regression techniques. Errors in spectra primarily stem from changes in ambient temperature, relative humidity, human errors, sample preparation mistakes, particle size variations, etc. [81].

2.14.4.1 Derivatives

Derivatives are typically employed as spectral preprocessing techniques to address the issues of overlapping peaks and baseline shifts. Starting from the original spectrum, the first derivative is initially applied, followed by the second derivative, and so forth, denoted as A, dA, d²A, and so on. The first derivative is utilized to minimize differences in baseline shifts, while the second derivative is employed to reduce differences in slope and separate overlapped peaks too.

Studies have shown that higher-order derivatives are not recommended as they tend to amplify noise along with the signal. Therefore, before derivative pretreatment the smoothing should be done.

2.14.4.2 Normalization

It is the method to suppress unwanted sources of variability by making a group of spectra to have more feature in common. There are four general types of normalization.

- i. Mean normalization
- ii. Max normalization
- iii. Range normalization
- iv. Vector normalization

The normalization is calculated as below:

- i. Mean normalization

$$X_{\text{mean,nor}} = \frac{X_i}{X_{\text{mean}}} \text{-----}[1.18]$$

Where, $X_{\text{mean,nor}}$ is mean normalization absorbance value of each variable, X_i is absorbance value of each variable, X_{mean} is mean absorption of the whole spectrum.

- ii. Max normalization

$$X_{\text{max,nor}} = \frac{X_i}{X_{\text{max}}} \text{-----}[1.19]$$

Where, $X_{\text{max,nor}}$ is maximum normalization absorbance value of each variable, X_i is absorbance value of each variable, X_{max} is maximum absorption of the spectrum.

- iii. Range normalization

$$X_{\text{range,nor}} = \frac{X_i}{(X_{i,\text{max}} - X_{i,\text{min}})^2} \text{-----} [1.20]$$

Where, $X_{\text{range,nor}}$ is the range normalization absorbance value of each variable, X_i is absorbance value of each variable, $X_{i,\text{max}}$ and $X_{i,\text{min}}$ are maximum and minimum absorption of the spectrum.

- iv. Vector normalization

$$X_{\text{vector,nor}} = \frac{X_i - \bar{X}}{\sqrt{\sum(X_i - \bar{X})^2}} \text{-----}[1.21]$$

Where, $X_{\text{vector,nor}}$ is normalization absorbance value of each variable, X_i is absorbance value of each variable, and \bar{X} is the mean of the spectral data. The norm of each vector normalization spectrum is 1.

2.14.4.3 Mean centering

Mean centering is a common preprocessing operation that occurs before multivariate data analysis. This technique involves shifting the data to the center of the coordinate system, facilitating the interpretation of the model. Mean centering is calculated by mean subtracting the absorbance value of each variable from every data point of the sample. This process ensures that the new mean of each sample is zero. Additionally, mean centering aids in reducing multicollinearity and enhancing the stability of the model.

$$X_{mc,i} = X_i - X_{av} \text{-----}[1.22]$$

Where, $X_{mc,i}$ are the mean centering absorbance value of each spectrum, X_i are absorbance value of each spectrum and X_{av} are mean of spectral data.

2.14.4.4 Baseline offset

Due to physical changes, example the shape and size of particles, baseline shifts are common in NIR spectra. To correct these shifts, the baseline offset preprocessing technique is employed, which adjusts the data to its minimum point. This technique involves subtracting the original spectra from the minimum absorption value. It is calculated as:

$$X_i = X_{org} - X_{min} \text{-----}[1.23]$$

Where, X_i is the new absorbance at sample i , X_{org} is the original absorption and X_{min} is the minimum of absorption in the spectrum.

2.14.4.5 Multiplicative scatter correction (MSC)

Scattering effects are common due to differences in particle size and density, and due to instrument drift, for example by temperature fluctuation or increasing, and they may cause nonlinear behavior in the data, particularly in the NIR. The effects of scattering are typically noticeable when spectra exhibit vertical baseline shifts. To mitigate the multiplicative effect of scattering, MSC adjusts each spectrum to closely align with the mean spectrum. MSC is a method used to correct spectra for spectral noise and background effects that cause baseline shifts and tilting. In MSC, the slope (a) and intercept (b) are first calculated from the regression between each individual spectrum and a reference spectrum.

$$X = aX_{\text{ref}} + b \text{ -----}[1.24]$$

Coefficients **a** and **b** are used to correct each spectrum by using the expression below:

$$X_{\text{corr}} = \frac{X-b}{a} \text{ -----}[1.25]$$

Where, **X** is the original sample spectrum, **X_{ref}** is the reference spectrum (average spectrum) and **X_{corr}** is the corrected sample spectrum.

2.14.4.6. Standard normal variate (SNV)

SNV is also utilized to mitigate the multiplicative effect of scattering, similar to MSC. In this process, the mean and the SD of each spectrum are calculated as follows:

$$X_{\text{SNV},i} = \frac{X_i - X_{\text{avg}}}{\text{SD}} \text{ -----}[1.26]$$

Where, **X_i** represents the absorbance for the spectrum at wavelength **i**, **X_{SNV,i}** denotes the standard normal variate absorbance for the spectrum at wavelength **i**. **X_{avg}** is the mean absorbance of the spectrum, and **SD** is the standard deviation absorbance for spectrum.

2.14.5 Modelling

Numerous techniques exist for optimizing the calibration model to predict reference data in NIR analysis. Simple linear regression, multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression stand out as the most common regression methods in this field. The primary aim of all calibration sets is to ascertain the regression coefficient 'b', also known as the model, derived from independent variables (spectral data - X variables) and dependent variables (reference data - Y variables). Once the regression coefficient 'b' is determined, we can formulate models with and without spectral preprocessing for prediction using the following equation:

$$\text{Validation set: } Y_{\text{pred}} = x \ b \text{ -----}[1.27]$$

$$\text{Calibration set: } Y_{\text{pred}} = X \ b \text{ -----}[1.28]$$

Where **Y_{pred}** = Predicted value, **x** = independent variable from validation set, **X** = independent variable from calibration set, and **b** = regression coefficient.

2.14.5.1 Partial least squares regression (PLSR)

PLSR is a widely used method in chemometrics for multivariate calibration, specifically handling spectroscopic data, and was established to address the problem of making good predictions [82]. PLSR is related to PCR and MLR. PCR captures maximum variance in \mathbf{X} variables, MLR achieves maximum correlation between \mathbf{X} and \mathbf{Y} variables whereas PLSR tries to do both ensuring the maximum correlation between \mathbf{X} variables and \mathbf{Y} variables (reference data) during calibration. PLSR compress data on such a way that the most variance in both \mathbf{X} and \mathbf{Y} is explained. PLSR creates a new variable from the linear combination of the original variables that is called a latent variable (LV) or factors. Factor contains all the relevant information of the investigated system. Considering too few factors leads to insufficient explanation of the changes in the spectral and reference data i.e. under fitting whereas considering the large number of factors accounted for the smallest changes in the data set such as noise leads to overfitting. Therefore, for better quality of the PLS model, the selection of the optimum number of factors (LVs) plays a key role. If several factors lead to comparably good results, it is recommended to select the model with the smallest number of factors.

In PLSR, both the \mathbf{X} and \mathbf{Y} matrixes are decomposed into smaller matrixes as below:

$$\mathbf{X} = \mathbf{T}\mathbf{L}^T + \mathbf{E} \text{-----}[1.29]$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^T + \mathbf{F} \text{-----}[1.30]$$

Where \mathbf{L} and \mathbf{Q} are called X-loading and Y-loading which is used to describe the variable \mathbf{T} in term of matrixes \mathbf{X} and \mathbf{Y} respectively. \mathbf{T} is a score matrix and is defined by \mathbf{X} matrix and weight matrix (\mathbf{W}) and is expressed as:

$$\mathbf{T} = \mathbf{X}\mathbf{W} \text{-----}[1.31]$$

\mathbf{E} and \mathbf{F} of equations 1.29 and 1.30 are the residue matrix of \mathbf{X} and \mathbf{Y} respectively.

As shown in Figure 2.13, the information in \mathbf{X} (x_1, x_2) of two dimensions is reduced from many variables to a few components in the t line, which is used as an independent variable. t is regressed with y (dependent variable) [83] and the regression coefficient matrix (\mathbf{b}) will be calculated. The predicted value is expressed as:

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} \text{-----}[1.32]$$

If data set is of any dimension, \mathbf{Q}^T will be defined as:

$$\mathbf{Q}^T = (\mathbf{T}^T\mathbf{T})^{-1}(\mathbf{T}^T\mathbf{Y}) \text{-----}[1.33]$$

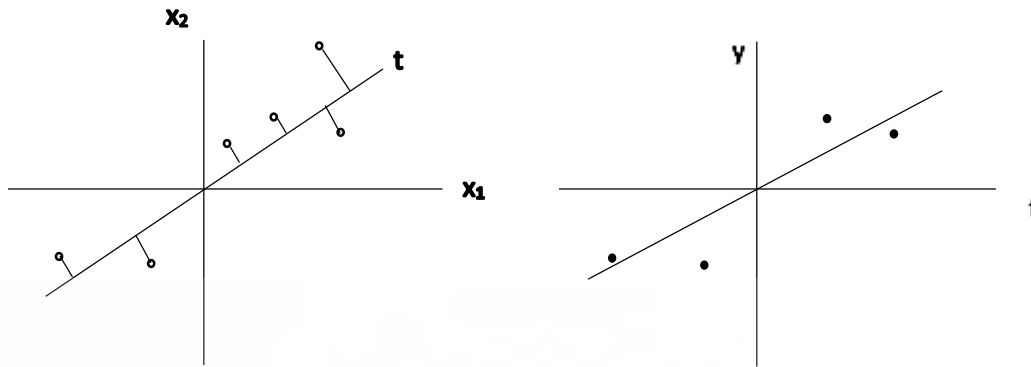


Figure 2.13 Geometrical illustration of the model structure used for the methods of PCR and PLS[83]

Equating equations 1.30 and 1.32, we get

$$X b = T Q^T \quad \text{-----}[1.34]$$

From equation 1.31, substituting the value of T we get,

$$X b = X W Q^T \quad \text{-----}[1.35]$$

Therefore, from equation 1.35 we can write the regression coefficient matrix as:

$$b = W Q^T \quad \text{-----}[1.36]$$

From equation 1.33, we can write equation 1.36 as:

$$b = W (T^T T)^{-1} (T^T Y) \quad \text{-----}[1.37]$$

Therefore, the predicted value matrix is calculated as:

$$\hat{Y} = X b$$

$$\hat{Y} = X W (T^T T)^{-1} (T^T Y) \quad \text{-----}[1.38]$$

2.14.6 Validation, performance analysis and model updating

After the calibration model has been developed using PLSR of spectroscopic data and corresponding reference values, it must be validated. Validation in different chemometric methods permit to identified outliers, the most suitable frequency ranges and also help to determine the optimum number of factors. There are two different types of validation. They are internal validation or cross-validation and external validation or test set validation. If the total number of data sets for model development is limited to 20-50 samples, Conzen [80] recommends cross-validation whereas if the total data set has more than 50 samples, external validation is recommended to perform for performance analysis.

In internal validation, the root mean square error of cross-validation (RMSECV) is calculated for the analysis of the model. RMSECV is a quantitative measure of the mean accuracy of the predictive capability of the chemometric model.

The procedure of a full or leave one out cross-validation is:

1. Remove a sample from the calibration set.
2. Develop a chemometric calibration model with the remaining samples.
3. Analyze the removed test sample and calculate the error from the sample

$$\text{Error}_1 = Y_1^{\text{measured}} - Y_1^{\text{predicted}}$$

4. Return the removed sample to the data set and remove a new sample from the same data set. Calculate a new model and predict a new sample. Calculate the error of this sample.

$$\text{Error}_2 = Y_2^{\text{measured}} - Y_2^{\text{predicted}}$$

5. Repeat step 4 until all the available 'N' samples of the calibration set.
6. Calculate the root mean square error of cross-validation is calculated as:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^N (y_{cv,i} - \hat{y}_{cv,i})^2}{N}} \text{-----}[1.39]$$

Where, $y_{cv,i}$ is the measurement value of the cross-validation test set, $\hat{y}_{cv,i}$ is the prediction value of the cross-validation test and N is the total number of samples. If only one sample is out from the calibration set, it is called full cross-validation or leave one out validation. However, if more than one sample; for example 3 samples were taken out and a model was developed from the remaining samples to predict the value of 3 samples with the same procedure as mentioned above, the process is called leave 3 out cross-validation. The process of taking more than one sample from the calibration set is called segmental cross-validation.

In external validation, the total data set is divided into a calibration set and validation set. Through a calibration set, the calibration model will be created and the model will be validated by the validation set. The performance of the model will be determined in terms of coefficient of determination (R^2), standard error of calibration (SEC), root mean square error of calibration (RMSEC), standard error of prediction (SEP), root mean square error of prediction (RMSEP), Residual prediction deviation (RPD) and bias.

The coefficient of determination is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \text{-----}[1.40]$$

Where y_i is the measured value, \bar{y} is the mean of measured value, \hat{y}_i is the predicted value. We know that error (e), also known as residue, is the difference between the measured value and predicted value. It can be expressed as:

$$e = y_i - \hat{y}_i \quad \text{-----}[1.41]$$

e^2 is called the sum of squared error of residuals (SSR) which is an unexplained variance and can be expressed as:

$$SSR = e^2 = \sum(y_i - \hat{y}_i)^2 \quad \text{-----}[1.42]$$

Also, the sum of squares of the difference between the measured value and its mean value is called total sum squares (TSS). It is called the total variance and is expressed as:

$$TSS = \sum(y_i - \bar{y})^2 \quad \text{-----}[1.43]$$

We know that total variance is the sum of explained variance and unexplained variance. Therefore, we can write:

$$\text{Explain variance} = TSS - SSR \quad \text{-----}[1.44]$$

Also, equation 1.40 can be written as

$$R^2 = 1 - \frac{SSR}{TSS} \quad \text{-----}[1.45]$$

According to Williams et al. (2019) [84] guideline, R^2 can indicate the applicability of the model and is used as follows:

Table 2.2 Guideline for the interpretation of R^2 (Williams et al. (2019))

R^2	Interpretation
up to 0.25	Not usable in NIRS calibration
0.26 - 0.49	Poor calibration, reasons should be researched
0.50 - 0.64	Ok for rough screening
0.66 - 0.81	Ok for rough screening and some other appropriate calibrations
0.83 - 0.90	Usable with caution for most application, including research
0.92 - 0.96	Usable in most applications, including quality assurance
0.98 +	Excellent, usable in any application

Bias is the average error between measured value and predicted value which indicates the accuracy of the quantitative model. It is one of the important parameter to measure the performance of the model.

$$\text{Bias} = \frac{\sum(y_{ip} - \hat{y}_{ip})}{N_p} \quad \text{-----}[1.46]$$

Where y_{ip} is the measured value of the prediction set, \hat{y}_{ip} is the prediction value of the prediction set and N_p is a number of samples in prediction set.

For validation of the model based on the calibration set, root mean square error of calibration (**RMSEC**) is used as one of the performance analysis parameters which gives preciseness of the calibration set.

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^{N_c} (y_{ic} - \hat{y}_{ic})^2}{N_c}} \quad \text{-----}[1.47]$$

where y_{ic} is the measured value of sample i of the calibration value set, \hat{y}_{ic} is the predicted value of the calibration set, and N_c is a number of samples in the calibration set.

Similarly, for validation of the model based on the validation set, the **Root mean square of prediction (RMSEP)** is used and is a qualitative measure for the accuracy of the analysis.

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{N_p} (y_{ip} - \hat{y}_{ip})^2}{N_p}} \quad \text{-----}[1.48]$$

Where y_{ip} is the measured value of the prediction set, \hat{y}_{ip} is the prediction value of the prediction set and N_p is a number of samples in the prediction set.

The standard **error of calibration (SEC)** is the **SD** of the difference between the measured value and the predicted value of the calibration set. It can be calculated as:

$$\text{SEC} = \sqrt{\frac{\sum_{i=1}^{N_c} ((y_{ic} - \hat{y}_{ic}) - \text{Bias}_c)^2}{N_c - 1}} \quad \text{-----}[1.49]$$

For the validation set, **standard error of prediction (SEP)** is used which is the **SD** of the difference between the measured value and the predicted value of the prediction set. It can be calculated as:

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^{N_p} ((y_{ip} - \hat{y}_{ip}) - \text{Bias}_p)^2}{N_p - 1}} \quad \text{-----}[1.50]$$

The ratio of prediction to deviation (RPD) is the ratio of the SD of the measured value of the prediction set to the SEP. RPD is used as a qualitative measure for the assessment of the validation results. RPD can be written as:

$$RPD = \frac{SD}{SEP} \quad \text{-----}[1.51]$$

RPD value should always be high in the performance analysis of the quantitative model. From equation 1.51, it is clear that for a high RPD value, the SD value must be high and SEP should be minimal.

Squaring both sides of equation 1.51- we get,

$$RPD^2 = \frac{SD^2}{SEP^2} \quad \text{-----}[1.52]$$

Substituting the value of SD and SEP in equation 1.52, we get

$$RPD^2 = \frac{\frac{\sum(y_i - \bar{y})^2}{N_p - 1}}{\frac{\sum[(y_i - \hat{y}_i) - \frac{\sum(y_i - \hat{y}_i)}{N_p}]^2}{N_p - 1}} \quad \text{-----}[1.53]$$

If Bias ≈ 0 , solving equation 1.53, we get

$$RPD^2 = \frac{\sum(y_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} \quad \text{-----}[1.54]$$

In equation 1.40, substituting the value of RPD^2 , we get,

$$R^2 = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} = 1 - \frac{1}{RPD^2} \quad \text{-----}[1.56]$$

The correlation coefficient (R) shows the relation between the predicted and measured value and is defined as

$$R = \left[\frac{(SD)^2 - (SEC)^2 \text{ or } (SEP)^2}{(SD)^2} \right]^{1/2} \quad \text{-----}[1.57]$$

Where, SD is the standard deviation of the measured value (y_i) in calibration set or prediction set

According to Zornoza et al. (2008) guideline [85], 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.

Once the model is developed, it should be regularly validated with the unknown samples.



References

1. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Elsevier - Advance Drug Delivery Reviews* **2005**, 57, 1109-1149.
2. 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.
3. Qu, J.-H.; Liu, D.; Cheng, J.-H.; Sun, D.-W.; Ma, J.; Pu, H.; Zeng, X.-A., Applications of near-infrared spectroscopy in food safety evaluation and control: a review of recent research advances. *Critical Reviews in Food Science and Nutrition* **2015**, 55, (13), 1939-1954.
4. Dickens, B.; Dickens, S. H., Estimation of concentration and bonding environment of water dissolved in common solvents using near infrared absorptivity. *Journal of research of the National Institute of Standards and Technology* **1999**, 104, (2), 173.
5. Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy. *Foods* **2017**, 6, (38), 1-11.
6. Jin, X.; Chen, X.; Shi, C.; Li, M.; Guan, Y.; Yu, C. Y.; Yamada, T.; Sacks, E. J.; Peng, J., Determination of hemicellulose, cellulose and lignin content using visible and near infrared spectroscopy in *Miscanthus sinensis*. *Bioresource technology* **2017**, 241, 603-609.
7. Goi, A.; Simoni, M.; Righi, F.; Visentin, G.; De Marchi, M., Application of a handheld near-infrared spectrometer to predict gelatinized starch, fiber fractions, and mineral content of ground and intact extruded dry dog food. *Animals* **2020**, 10, (9), 1660.
8. Chen, P., Use of optical properties of food materials in quality evaluation and materials sorting. *Journal of Food Process Engineering* **1978**, 2, (4), 307-322.
9. Osborne, B. G.; Fearn, T.; Hindle, P. H., *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman scientific and technical: 1993.
10. Yukihiro Ozaki, C. H., Satoru Tsuchikawa, and soren Balling Engelsen, *Near-Infrared Spectroscopy - Theory, Spectral Analysis, Instrumentation, and Applications*. Springer Nature Singapore Pte Ltd. 2021: July 2020.
11. Conzen, J.-P., *Multivariate Calibration - A Practical guide for developing methods in the quantitative analytical chemistry*. Bruker Optik GmbH 2014.
12. *Near Infrared Technology - Getting the best out of light*. AFRICAN SUN MeDIA: 2019.

13. Gorji, R.; Skvaril, J.; Odlare, M., Determining Moisture Content of Basil Using Handheld Near-Infrared Spectroscopy. *Horticulturae* **2024**, 10, (4), 336.
14. Zhao, Y.; Deng, J.; Chen, Q.; Jiang, H., Near-infrared spectroscopy based on colorimetric sensor array coupled with convolutional neural network detecting zearalenone in wheat. *Food Chemistry: X* **2024**, 101322.
15. O'Brien, C.; Falagán, N.; Kourmpetli, S.; Landahl, S.; Terry, L. A.; Alamar, M. C., Non-destructive methods for mango ripening prediction: Visible and near-infrared spectroscopy (visNIRS) and laser Doppler vibrometry (LDV). *Postharvest Biology and Technology* **2024**, 212, 112878.
16. Yang, X.; Zhuang, X.; Shen, R.; Sang, M.; Meng, Z.; Cao, G.; Zang, H.; Nie, L., In situ rapid evaluation method of quality of peach kernels based on near infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2024**, 124108.
17. Phanomsophon, T.; Jaisue, N.; Worphet, A.; Tawinteung, N.; Shrestha, B.; Posom, J.; Khurnpoon, L.; Sirisomboon, P., Rapid measurement of classification levels of primary macronutrients in durian (*Durio zibethinus* Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling. *Measurement* **2022**, 203, 111975.
18. Yan, H.; Dong, C.; Zhang, J.; Hu, X.; Xue, J.; Zhao, Y.; Wang, X., Moisture Prediction of Biomass Fuel Based on Near-Infrared Spectroscopy and Deep Learning Algorithm. *Energy & Fuels* **2024**.
19. 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>.
20. Mancini, M.; Rinnan, Å., Near infrared technique as a tool for the rapid assessment of waste wood quality for energy applications. *Renewable Energy* **2021**, 177, 113-123.
21. Pan, N.; Yu, Z.; Ling, W.; Xu, J.; Liao, Y., Comparative Study on Calibration Models Using NIR Spectroscopy Data. *Journal of Applied Spectroscopy* **2024**, 1-7.
22. Millatina, N. R. N.; Calle, J. L. P.; Barea-Sepulveda, M.; Setyaningsih, W.; Palma, M., Detection and quantification of cocoa powder adulteration using NIR spectroscopy with chemometrics approach. *Food Chemistry* **2024**, 139212.

23. Heise, H. M., Medical applications of NIR spectroscopy. *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications* **2021**, 437-473.
24. Scheeren, T.; Schober, P.; Schwarte, L., Monitoring tissue oxygenation by near infrared spectroscopy (NIRS): background and current applications. *Journal of clinical monitoring and computing* **2012**, *26*, 279-287.
25. P Guo, L.; Q Huang, L.; P Zhang, X.; Bittner, L.; Pezzei, C.; Pallua, J.; Schonbichler, S.; A Huck-Pezzei, V.; K Bonn, G.; W Huck, C., Application of near-infrared spectroscopy (NIRS) as a tool for quality control in traditional Chinese medicine (TCM). *Current Bioactive Compounds* **2011**, *7*, (2), 75-84.
26. Li, K.; Zhang, X.; Zhang, J.; Du, B.; Song, X.; Wang, G.; Li, Q.; Zhang, Y.; Liu, F.; Zhang, Z., Simultaneous Rapid Detection of Multiple Physicochemical Properties of Jet Fuel Using Near-Infrared Spectroscopy. *ACS Omega* **2024**.
27. Biomass Explained. *Independent Statistics & Analysis* June 8, 2021.
28. Determination of Hemicellulose, Cellulose, and Lignin Content in Different Types of Biomasses by Thermogravimetric Analysis and Pseudocomponent Kinetic Model (TGA-PKM Method). *Processes - Production of Biofuels and Numerical Modeling of Chemical Combustion Systems* **2020**, *8*, 1-21.
29. *IEA Bioenergy - Annual Report 2020*; IEA: Ireland, 2021.
30. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B. P., Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* **2023**, *16*, (14), 5351.
31. Characteristics of Cellulose, Hemicellulose and Lignin of MD2 Pineapple Biomass *Chemical Engineering Transactions*, **2018**, *72*, 79-84.
32. The effect of the biomass components lignin, cellulose and hemicellulose on TGA and fixed bed pyrolysis. *Journal of Analytical and Applied Pyrolysis* **May 2013**, *101*, 177-184.
33. Characterization of Canadian biomass for alternative renewable biofuel. *Renewable Energy* **2010**, *35*, 1624-1631.
34. Higher heating values of lignin types from wood and non-wood lignocellulosic biomasses. *Energy Sources* **2017**, 592-598.

35. The effect of the biomass components lignin, cellulose and hemicellulose on TGA and fixed bed pyrolysis. *Journal of Analytical and Applied Pyrolysis* **2013**, 101, 177-184.
36. Combustion characteristics of different biomass materials. *Energy conversion and management* **2003**, 44, 155-162.
37. Review of Physicochemical Properties and Analytical Characterization of Lignocellulosic Biomass. *Renewable and Sustainable Energy Reviews* **2017**, 76, 309-322.
38. Combustion characteristics of different biomass fuels. *Progress in Energy and Combustion Science* **2004**, 30, 219-230.
39. Bioenergy generation from agricultural wastes and enrichment of end products. In *Refining Biomass Residues for Sustainable Energy and Bioproducts*, Academic Press: 2020; pp 337-356.
40. Effect of Particle Size on Mechanical Properties of Pellets Made from Biomass Blends. *Procedia Engineering - Elsevier* **2016**, 148, 93-99.
41. Jetsada, P.; 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.
42. Nieto, P. J. G.; García-Gonzalo, E.; Paredes-Sánchez, B. M.; Paredes-Sánchez, J. P., Forecast of the higher heating value based on proximate analysis by using support vector machines and multilayer perceptron in bioenergy resources. *Fuel* **2022**, 317, 122824.
43. Characterization of north american lignocellulosic biomass and biochars in terms of their candidacy for alternative renewable fuels. *Bioenerg. Res.* **2013**, 6, 663-677.
44. In *Precision test for spectral characteristic of NIR spectroscopy and bomb calorimeter for measuring higher heating value of Leucaena Leucocephala*, MATEC Web of Conferences-ICEAST 2018, 2018; 2018.
45. Aich, S.; Behera, D.; Nandi, B. K.; Bhattacharya, S., Relationship between proximate analysis parameters and combustion behaviour of high ash Indian coal. *International Journal of Coal Science & Technology* **2020**, 7, 766-777.
46. *Biomass Gasification and Pyrolysis - Chapter 2: Biomass Characteristics*. Elsevier Inc. : 2010.
47. Physical properties of solid biomass. *Biomass and Bioenergy* **2011**, 35, 2093-2105.

48. ASTM E871-82 -Standard Test Method for Moisture Analysis of Particulate Wood Fuels; ASTM: West Conshohocken, PA, USA, 2019.
49. ASTM E872-82 Standard Test Method for Volatile Matter in the Analysis of Particulate Wood Fuels; West Conshohocken, PA, USA, 2019.
50. Vassilev, S. V.; Vassileva, C. G.; Song, Y.-C.; Li, W.-Y.; Feng, J., Ash contents and ash-forming elements of biomass and their significance for solid biofuel combustion. *Fuel* **2017**, 208, 377-409.
51. Pazó, J. A.; Granada, E.; Saavedra, Á.; Eguía, P.; Collazo, J., Biomass thermogravimetric analysis: Uncertainty determination methodology and sampling maps generation. *International Journal of Molecular Sciences* **2010**, 11, (7), 2701-2714.
52. Özyuğuran, A.; Yaman, S., Prediction of calorific value of biomass from proximate analysis. *Energy Procedia* **2017**, 107, 130-136.
53. Demirbas, A., Combustion characteristics of different biomass fuels. *Progress in energy and combustion science* **2004**, 30, (2), 219-230.
54. García, R.; Pizarro, C.; Lavín, A. G.; Bueno, J. L., Biomass proximate analysis using thermogravimetry. *Bioresource technology* **2013**, 139, 1-4.
55. ASTM E1755-01 Standard Test Method for Ash in Biomass; West Conshohocken, PA, USA, 2020.
56. Kongto, P.; Palamanit, A.; Ninduangdee, P.; Singh, Y.; Chanakaewsomboon, I.; Hayat, A.; Wae-hayee, M., Intensive exploration of the fuel characteristics of biomass and biochar from oil palm trunk and oil palm fronds for supporting increasing demand of solid biofuels in Thailand. *Energy Reports* **2022**, 8, 5640-5652.
57. Jia, Y.; Li, Z.; Wang, Y.; Wang, X.; Lou, C.; Xiao, B.; Lim, M., Visualization of combustion phases of biomass particles: effects of fuel properties. *ACS omega* **2021**, 6, (42), 27702-27710.
58. Sahito, A. R.; Mahar, R. B.; Siddiqui, Z.; Brohi, K. M., Estimating calorific values of lignocellulosic biomass from volatile and fixed solids. *International Journal of Biomass and Renewables* **2013**, 2, (1), 1-6.

59. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B. P.; Pornchaloempong, P.; Funke, A., NIR Spectroscopy as an Alternative to Thermogravimetric Analyzer for Biomass Proximate Analysis: Comparison of Chip and Ground Biomass Models. *Energies* **2024**, *17*, (4), 800.
60. Ardila, Y. C.; Figueroa, J. E. J.; Maciel, M. R. W., Mathematical models for predicting the higher heating value and ultimate analysis of biomass. *Industrial Crops and Products* **2024**, *208*, 117777.
61. Qian, H.; Guo, X.; Fan, S.; Hagos, K.; Lu, X.; Liu, C.; Huang, D., A simple prediction model for higher heat value of biomass. *Journal of Chemical & Engineering Data* **2016**, *61*, (12), 4039-4045.
62. Seow, Y. X.; Tan, Y. H.; Mubarak, N.; Kansedo, J.; Khalid, M.; Ibrahim, M. L.; Ghasemi, M., A review on biochar production from different biomass wastes by recent carbonization technologies and its sustainable applications. *Journal of Environmental Chemical Engineering* **2022**, *10*, (1), 107017.
63. Żukowski, W.; Jankowski, D.; Wrona, J.; Berkowicz-Plątek, G., Combustion behavior and pollutant emission characteristics of polymers and biomass in a bubbling fluidized bed reactor. *Energy* **2023**, *263*, 125953.
64. Ozgen, S.; Cernuschi, S.; Caserini, S., An overview of nitrogen oxides emissions from biomass combustion for domestic heat production. *Renewable and Sustainable Energy Reviews* **2021**, *135*, 110113.
65. Aleksiejczuk, A.; Teleszewski, T. J., Estimation of Sulfur Dioxide Emissions in an Automatic Boiler with a Retort Burner for Coal and Biomass in a Single-Family House Based on the Measurement of the Heat Consumed. *Environmental Sciences Proceedings* **2022**, *18*, (1), 10.
66. Adamovics, A.; Platace, R.; Gulbe, I.; Ivanovs, S., The content of carbon and hydrogen in grass biomass and its influence on heating value. *Engineering for rural development* **2018**, *17*, (1), 1277-1281.
67. Kalinci, Y.; Hepbasli, A.; Dincer, I., Biomass-based hydrogen production: a review and analysis. *International journal of hydrogen energy* **2009**, *34*, (21), 8799-8817.
68. Silva, D. A. d.; Eloy, E.; Caron, B. O.; Trugilho, P. F., Elemental chemical composition of forest biomass at different ages for energy purposes. *Floresta e Ambiente* **2019**, *26*.

69. Ren, X.; Sun, R.; Meng, X.; Vorobiev, N.; Schiemann, M.; Levendis, Y. A., Carbon, sulfur and nitrogen oxide emissions from combustion of pulverized raw and torrefied biomass. *Fuel* **2017**, 188, 310-323.
70. Vainio, E., Fate of fuel-bound nitrogen and sulfur in biomass-fired Industrial boilers. **2014**.
71. Yi, Q.; Qi, F.; Cheng, G.; Zhang, Y.; Xiao, B.; Hu, Z.; Liu, S.; Cai, H.; Xu, S., Thermogravimetric analysis of co-combustion of biomass and biochar. *Journal of thermal analysis and calorimetry* **2013**, 112, 1475-1479.
72. Magdziarz, A.; Wilk, M.; Straka, R., Combustion process of torrefied wood biomass: a kinetic study. *Journal of Thermal Analysis and Calorimetry* **2017**, 127, 1339-1349.
73. *Handbook of Industrial Hydrocarbon Processes*. CD & W inc. Laramie, WY: United States, 2019.
74. Combustion and co-combustion of biomass: fundamentals, technologies, and primary measures for emission reduction. *Energy & fuels* **2003**, 17, (6), 1510-1521.
75. Combustion characteristics of different biomass fuels. *Progress in energy and combustion science* **2004**, 30, (2), 219-230.
76. Properties of wood for combustion analysis. *Bioresource technology* **1991**, 37, (2), 161-168.
77. Shrestha, B.; Posom, J.; Pornchaloempong, P.; Sirisomboon, P.; Shrestha, B. P.; Ariffin, H., Near-Infrared Spectroscopy Modeling of Combustion Characteristics in Chip and Ground Biomass from Fast-Growing Trees and Agricultural Residue. *Energies* **2024**, 17, (6), 1338.
78. Evaluation of pyrolysis characteristics of milled bamboo using near infrared spectroscopy. *Renewable Energy* **2017**, 103, 653-665.
79. 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.
80. *Multivariate Calibration - A practical guide for developing methods in the quantitative analytical chemistry*. Bruker Optik GmbH: 2014.
81. Non-destructive and rapid measurement of sugar content in growing cane stalks for breeding programmes using visible-near infrared spectroscopy. *Biosystem Engineering* **2020**, 197, 76-90.

82. A Partial Least Squares based algorithm for parsimonious variable selection. *Algorithms for Molecular Biology* **2011**, 6-27.
83. *A user friendly guide to Multivariate calibration and Classification*. NIR publication: Chichester, UK, 2004.
84. Williams, P.; Manley, M.; Antoniszyn, J., *Near infrared technology: getting the best out of light*. African Sun Media: 2019.
85. 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>.



Chapter 3

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

In this study, partial least squares regression (PLSR) models were developed using no preprocessing, traditional preprocessing, multi-preprocessing 5-range, multi-preprocessing 3-range, a genetic algorithm (GA), and a successive projection algorithm (SPA) to assess the higher heating value (HHV) and ultimate analysis of grounded biomass for energy usage by 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 were compared, and the best model was selected based on the coefficient of determination of the validation set, root mean square error of prediction, and the ratio of prediction to deviation values. Based on the model performance results, the proposed HHV model from GA-PLSR and the N models from the multi-preprocessing PLSR 5-range could be used for most applications, including research, whereas the C and H models from GA-PLSR and the O model from the multi-preprocessing PLSR 5 range method 5-range method have fair performance and are 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 these models by 4.1839%, 8.1842%, 3.7587%, and 4.0085%, respectively. Therefore, this method 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 robust application.

Keyword: higher heating value, near-infrared spectroscopy, partial least squares regression, spectral multi-preprocessing method, ultimate analysis



*This chapter constituted the publication article: Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* 2023, 16, 5351. <https://doi.org/10.3390/en16145351>

Abbreviations

%	percentage
C	carbon
CHNS	CHNS Elemental analyzer
GA	genetic algorithm
H	hydrogen
HHV	higher heating value
LVs	latent variable number
Max	maximum
Min	minimum
MP	multi-preprocessing
MSC	multiplicative scatter correction
N	nitrogen
N_T	total number of sample
N_c	number of sample in calibration set
NIRS	near infrared spectroscopy
N_p	number of sample in validation set
O	oxygen
PLSR	partial least squares regression
R^2	coefficient of determination
R^2_c	coefficient of determination of calibration set
R^2_p	coefficient of determination of validation set
RMSEC	root mean square error of calibration set
RMSEP	root mean square error of prediction set
RPD	ratio of prediction to deviation
S	sulfur

SD	standard deviation
SEC	standard error of calibration set
SEL	standard error of laboratory
SEP	standard error of validation set
SNV	standard normal variate
SPA	successive projection algorithm
SW	selected wavenumber
wt.%	weight percentage



3.1. 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]. 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, the 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 3595 to 12,489 cm^{-1} . It predominantly interacts with the bonds of non-symmetrical 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 [3] and elemental composition (C, H, N, O, and S) [16] of grounded bamboo using NIRS. Skvaril et al. [17] reviewed the application of NIRS in biomass energy conversion processes. Zhang et al. [18] studied the fast analysis of HHV and elemental composition of sorghum biomass using NIRS. Xue et al. [19] 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.

NIRS, combined with a broad range of wavelengths and suitable chemometric models, offers extensive applications in various fields, such as food quality control, agriculture, biofuels, and drug analysis [13]. NIRS has been successfully employed for on-line, at-line, off-line, and in-line analysis, using instruments from different NIR ranges. For instance, in-line fiber-optic NIR spectra (300–1160 nm) have been utilized to classify durian pulp samples based on their dry matter content and soluble solids content [20]. FT-NIRS (800–2500 nm) has enabled rapid measurement of macronutrients such as nitrogen, phosphorus, and potassium in durian leaves, aiding in the production of high-quality durian fruits through optimal fertilization practices [21]. FT-NIR (700–2500 nm) has been employed to predict total phenolics and antioxidants in hulled and naked oats of different genotypes [22]. Vis-NIR (570–1031 nm) and Mid-NIR (860–1760 nm) spectroscopy have been utilized for starch content prediction in cassava [23]. The Micro-NIR portable spectrometer (900–1676 nm) has been found to have applications in the classification and quantification of crude oils and fuels [24]. Additionally, a portable NIR analyzer (1300–2600 nm) has been used for rapid confirmation of the presence of illicit drugs, such as cocaine [25].

NIRS provide better spectral reproducibility with a higher signal-to-noise ratio compared to other complementary analytical techniques, such as Raman and IR spectroscopies, making it one of the most important parameters in quantitative calibration [26]. The better penetration depth in samples, minimal or no sample preparation, shorter acquisition times, and wide range of application in diverse fields highlight the multidisciplinary nature of NIRS. In contrast, the presence of a strong water absorption band in the NIR region limits the applicability of NIRS for samples with a high water content. In such cases, Raman spectroscopy can be a suitable alternative as it is relatively unaffected by water interference and can effectively analyze aqueous solutions and biological samples without significant water-related issues [27,28]. However, it is important to note that Raman scattering is inherently a weak phenomenon, often requiring longer acquisition times and being more sensitive to sample fluorescence [26]. In addition, the cost of instrumentation for Raman and IR spectroscopy is higher compared to NIRS. These factors showcase the acceptance of NIRS as a rapid, reliable, and non-destructive method, resulting in energy, environmental, cost, and time savings.

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); a CHNS/O elemental analyzer for wt.% of C, N, H, S, and O; and a thermogravimetric analyzer for wt.% ash content. 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 by employing NIRS.
2. To compare the performance of the PLSR models based on R^2_C , RMSEC, R^2_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.

The research outcomes of this study have practical applications in real life. The developed model offers a rapid, reliable, and non-destructive alternative to traditional laboratory methods for assessing biomass properties. This benefits biomass traders in determining a fair price based on actual energy properties, rather than relying solely on volume or weight. Industries relying on biomass for energy can optimize system efficiency and cost-effectiveness through informed feedstock selection. The model is applicable for process monitoring and quality control in biomass-based energy production facilities. This facilitates real-time adjustments by engineers and operators, ensuring consistent and efficient energy production. Policymakers, energy companies, and researchers can utilize these findings for the proper identification, management, and utilization of bio-resources to meet future energy demands. Moreover, the research outcomes pave the way for NIR-based research in various fields to adopt or enhance similar approaches.

3.2. Materials and Methods

Figure 3.1 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.

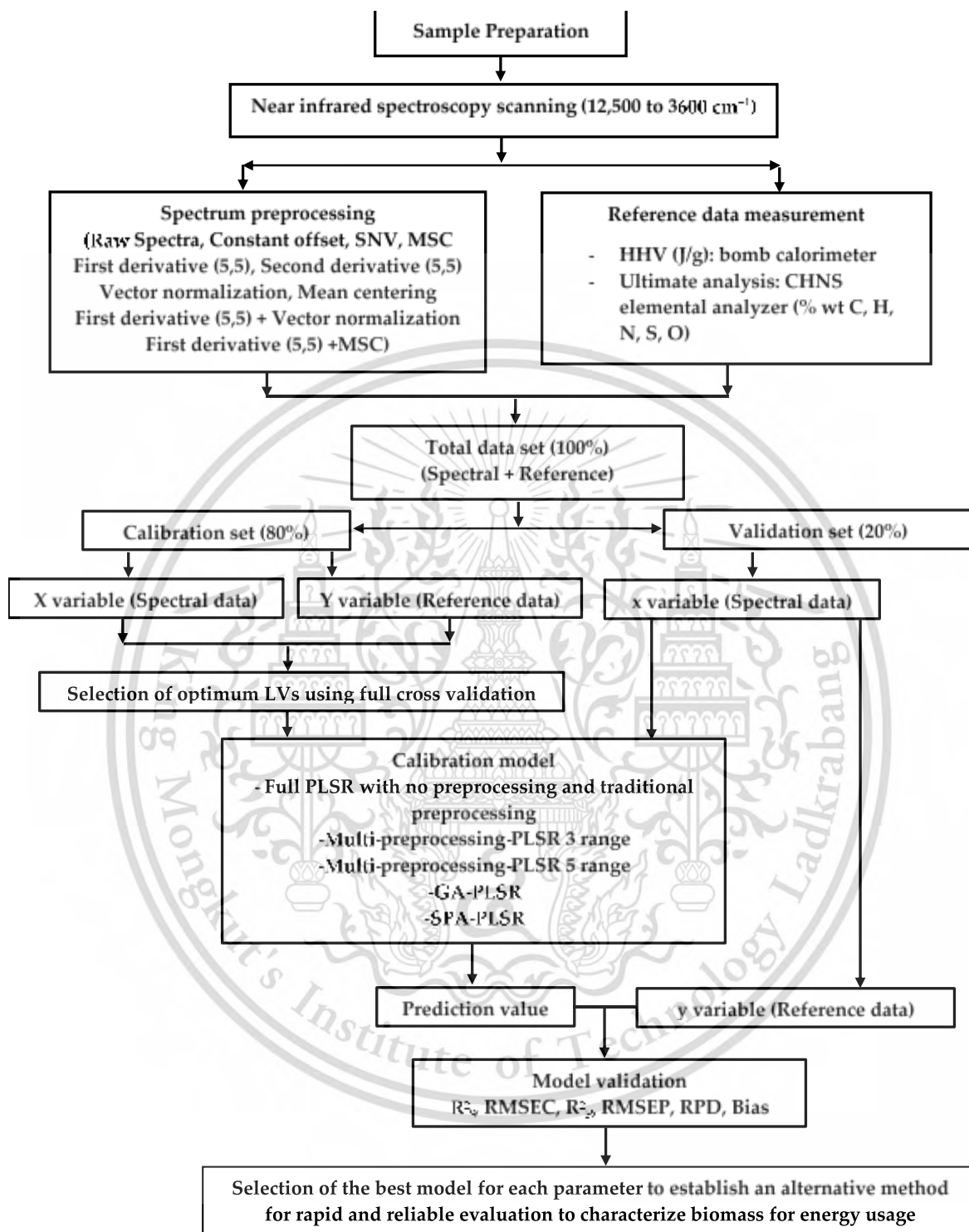


Figure 3.1 Flowchart of the overall research methodology for the evaluation of the HHV and ultimate analysis parameters of grounded biomass for energy usage using NIRS combined with PLSR.

3.2.1. Sample Preparation

The biomass samples were collected from the Terai low flatland and mid-hill regions of Nepal, with altitudes ranging from 86 to 1940 m above sea level. The study included five fast-growing species: (1) *Alnus nepalensis*, (2) *Pinux roxiburghii*, (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 *Pinux roxiburghii* 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 the 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 x 15 mm (refer to Figure 3.2a); 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 the Scientific and Technological Research Equipment Center (STREC) at Chulalongkorn University, Bangkok, Thailand, using the instrument, Mastersizer 3000 (MAL1099267, Hydro MV). Figure 3.3 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 ziplock bags before and during the experiment.



Figure 3.2 Nepal biomass in (a) chips form (>30 mm × 15 mm), (b) grounded form (1.88–3080 μm), (c) FT-NIRS (MPA, Bruker, Ettlingen, Germany) scanning between the wavenumber range 3595 to 12,489 cm⁻¹, and (d) ground sample presentation by transmittance mode.

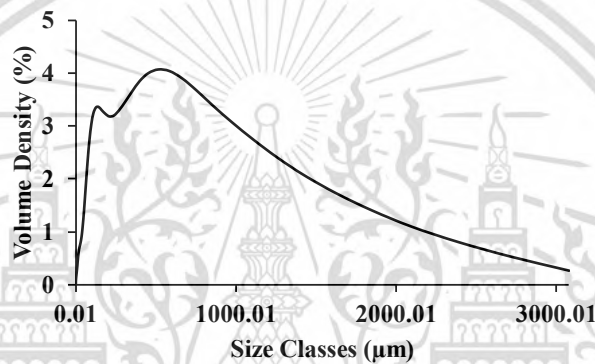


Figure 3.3 Representative particle size distribution of the ground biomass ranging from 0.01 to 3080 μm.

3.2.2. Spectral Data Collection

As shown in Figure 3.2c,d, 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 transmittance mode at the controlled temperature of 25 ± 2 °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 3595 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 3.4a shows the raw spectrum of ten different grounded biomasses within the wavenumber range between 3594.87 to 12,489.48 cm^{-1} , which were used to evaluate the HHV and ultimate analysis parameters.

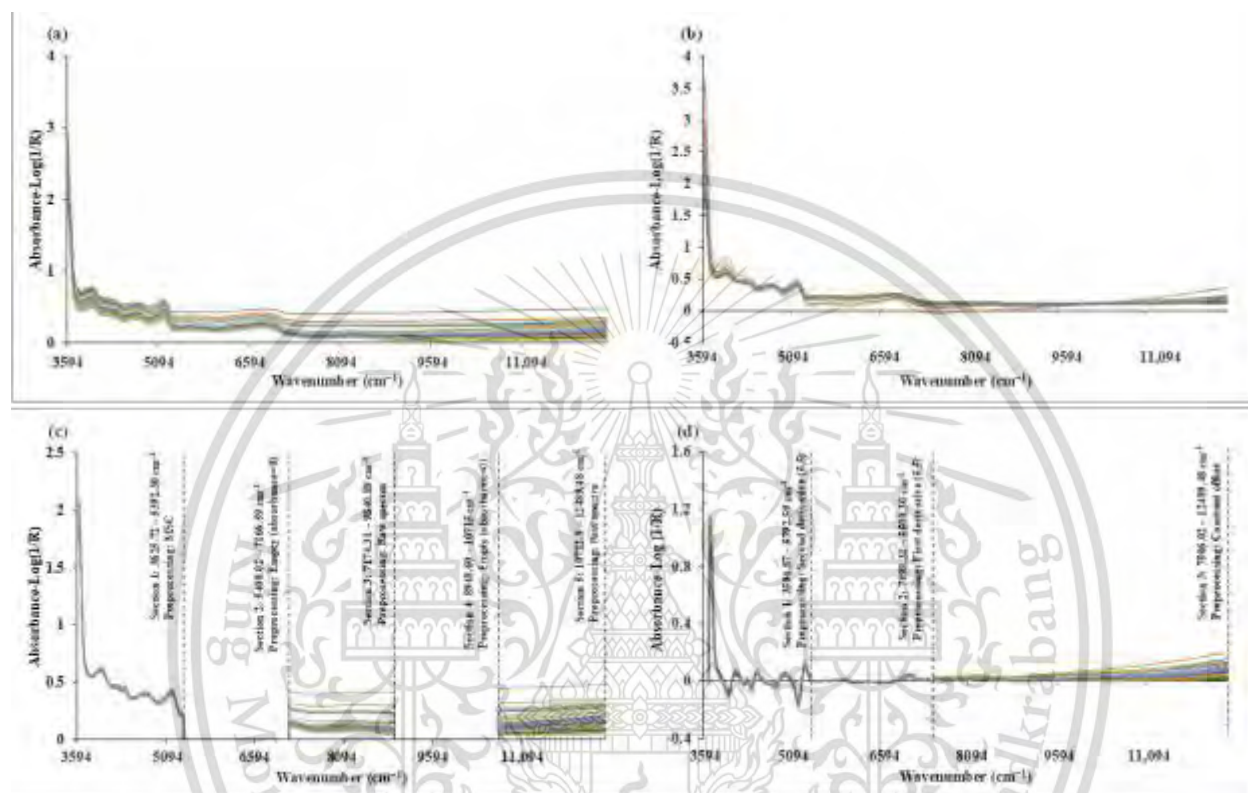


Figure 3.4 (a) Raw spectra of grounded biomass. (b) Preprocessing of raw spectra using traditional approach (MSC). (c) Pretreated spectra by the multi-preprocessing method with a 5-range method in five equal sections: MSC, empty, raw spectra, empty, and raw spectra (from left to right), within the wavenumber range of 3625.72–12,489.48 cm^{-1} . (d) Pretreated spectra by the multi-preprocessing method with a 3-range method in three sections: Second derivative, first derivative, and constant offset (from left to right) within the wavenumber range of 3594.87–12,489.48 cm^{-1} .

3.3. Reference Analysis

Due to the complex nature of NIR absorbance data, it must be correlated with reference values obtained using a standard laboratory method [29]. Thus, the reference data, which include HHV, C, H, N, S, and O, were evaluated after being scanned from a FT-NIR spectrometer.

3.3.1. Higher Heating Value (HHV)

The HHV of the grounded biomass is measured using the isoperibol method with an automatic bomb calorimeter (IKA C 200, Staufen, Baden-Württemberg, 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 L 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 min.

3.3.2. Ultimate Analysis

The ultimate analysis includes quantification of wt.% of C, H, N, S, and O on a dry basis in the ground biomass to determine the major elemental composition. The wt.% of C, H, N, and S in the ground sample were measured using the CHNS/O analyzer (Thermo Scientific™ FLASH 2000, Waltham, Massachusetts, United States). The wt.% of ash content in the ground biomass was measured using the thermogravimetric analyzer (TG 209 F3 Tarsus, Netzsch, Bavaria, Germany). The wt.% of O on a dry basis in the ground biomass sample was calculated as a difference [30] :

$$\text{wt.\% O} = 100 - (\text{wt.\% C} + \text{wt.\% H} + \text{wt.\% N} + \text{wt.\% S} + \text{wt.\% Ash}) \quad (1)$$

3.3.3. Outlier and Standard Error of Laboratory

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:

$$\frac{(X_i - \bar{X})}{SD} \geq \pm 3 \quad (2)$$

If Equation (2) is satisfied for any sample i , the sample is considered as an outlier and is not considered in the total dataset for model development [31].

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:

$$SEL = \sqrt{\frac{\sum_{i=1}^N (y_1 - y_2)^2}{N_T}} \quad (3)$$

3.4. Spectral Preprocessing

Spectral preprocessing is one of the most 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 3.4a). To improve spectral features, it is important to remove noise, address overlapping peaks and baseline shifts, handle collinearity within the spectral data, and enable easy data interpretation for calibration [32,33], and NIR spectral preprocessing is necessary before model development.

To date, models have been developed using a traditional preprocessing approach (refer Figure 3.4b) on the entire available wavelength range for the prediction and evaluation of respective samples. However, there has been a lack of exploration regarding the pretreatment of raw spectra by employing different preprocessing techniques on distinct sections of the entire wavelength range. It is thought that a multi-preprocessing approach, i.e. a unique preprocessing technique that divides the entire spectrum into different sections using different spectral preprocessing methods based on random pairs, will improve the assessment of the biomass properties for energy usage using NIRS. Based on this hypothesis, this study introduced a novel multi-preprocessing approach: the 5-range and 3-range methods (refer to Figure 3.4c,d) as unique components to improve the assessment of biomass properties using NIRS. The research outcomes from the multi-preprocessing technique with PLSR will serve as a pivotal milestone in the research

and development of NIRS. This will benefit NIRS-related research from diverse fields by permitting the upgrading of existing models and their effective utilization in various applications.

Therefore, in this study, the raw spectrum was subjected to two distinct pretreatment approaches. The first approach adhered to the traditional methodology, entailing the application of a single spectral preprocessing method to the entire wavenumber range (3595 to 12,489 cm^{-1}). Meanwhile, the second approach introduced a novel and innovative multi-preprocessing technique, whereby the entire wavenumber range was partitioned into multiple sections and underwent pretreatment using a comprehensive combination of various preprocessing methods. For the traditional approach, ten different types of spectrum pretreatment methods were used for the 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 pretreated with various pretreatment combination sets obtained from seven different preprocessing methods, as indicated by the following markings: 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 (refer Figure 3.4c), the following procedures were adopted:

1. Equally dividing the entire wavenumber range into five sections: 3625.72–5392.30 cm^{-1} , 5400.02–7166.59 cm^{-1} , 7174.31–8940.89 cm^{-1} , 8948.60–10,715 cm^{-1} , and 10,722.9–12,489.48 cm^{-1} . However, since the wavenumber range from 3594.87 to 12,489.48 cm^{-1} is not equally divisible by 5, the last four independent variables were excluded from the total dataset, resulting in 1150 out of 1154 variables being considered for model development.
2. Generating all possible combinations of multi-preprocessing sets from 0 to 6.
3. Selecting the most effective multi-preprocessing combination by evaluating different numbers of random pairs to develop the PLSR-based model.

Similarly, for the multi-preprocessing 3-range method (refer Figure 3.4d), the following procedures were adopted:

1. Dividing the entire wavenumber range into three sections: 3594.87–5492.59 cm^{-1} , 7498.31–5500.30 cm^{-1} , and 7506.02–12,489.48 cm^{-1} .
2. Generating all possible combinations of multi-preprocessing sets from 0 to 6.
3. Selecting the most effective multi-preprocessing combination by evaluating different numbers of random pairs to develop the PLSR-based model.

Figure 3.4c,d shows the spectrum of the grounded biomass obtained from the multi-preprocessing method with the (a) 5-range and (b) 3-range methods, respectively. In Figure 3.4c, the raw spectrum was pretreated with the preprocessing combination set of 3, 0, 1, 0, and 1—i.e., MSC from 3625.72–5392.30 cm^{-1} , empty from 5400.02–7166.59 cm^{-1} , raw spectra from 7174.31–8940.89 cm^{-1} , empty from 8948.60–10,715 cm^{-1} , and raw spectra from 10,722.9–12,489.48 cm^{-1} . With the introduction of empty preprocessing, an improvement in the model indicates that uninformative spectral ranges have been removed from the modeling. The main advantages of introducing empty preprocessing (zero absorbance) include the reduction of potential spectral noise, the simplification of spectral data, improved model interpretability, and decreased computational load, which speeds up both the calibration process and computational time. However, this approach might result in the loss of potentially valuable information. Therefore, the prove of introducing of empty preprocessing is required. Similarly, in Figure 3.4d, the raw spectrum was pretreated with the preprocessing combination set of 4, 4, and 6—i.e., second derivative from 3594.87–5492.59 cm^{-1} , first derivative from 7498.314–5500.30 cm^{-1} , and constant from 7506.02–12,489.48 cm^{-1} . The best combination set for multi-preprocessing is determined by the optimum LVs obtained from full cross-validation.

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

3.5. Model Development

The accuracy of the model is one of the major concerns of NIRS. Accuracy can be improved by using different spectral pretreatments and appropriate data analysis methods. Various research articles related to NIRS modeling have concluded that PLSR is one of the most effective and commonly used quantitative analysis techniques [14,34–36]. Therefore, this study proposes PLSR-based models that can handle highly collinear spectroscopic data [37] 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 obtained after removing outliers was manually divided into an 80% calibration set and a 20% validation set, as illustrated in Figure 3.1. The total data set consists of ten different varieties of biomass comprised of five fast-growing trees and five agricultural residues. Therefore, it is crucial to stratify the total dataset to ensure both the calibration and validation sets encompass representative samples, covering the entire range of variation within the overall sample population. Allocating 80% of the total dataset as the calibration set, which includes the maximum and minimum reference values, ensures proportional representation of all biomass varieties in the model development process. This approach reduces bias, facilitates effective learning of underlying patterns and relationships, and helps prevent issues, such as overfitting or underfitting, to generate a regression model [33]. 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 [38]. 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 the full wavenumber range only, thus avoiding overfitting [39–41]. 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 [42]. 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, thereby maximizing its fitness. The model developed with GA-PLSR shows the lowest prediction error as it maximizes the fitness and co-variance between the spectral and reference data [43,44]. 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_p , RMSEP, RPD, and bias [45]. These parameters can be calculated as follows, where y is the measured value, \hat{y} is the predicted value, i is subscript used to indicate the number of the 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:

$$R^2_c, R^2_p = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4)$$

$$\text{RMSEC, RMSEP} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N_T}} \quad (5)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSEP}} \quad (6)$$

$$\text{Bias} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (7)$$

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 the R^2 and RPD value, were interpreted based on the recommendations of Williams et al. (2019) [46] and Zornoza et al. (2008) [47], respectively.

As per the 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 [46]. 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 [47].

3.6. Results and Discussion

3.6.1. Comparison of Near-Infrared Spectra of Ground Biomass with Pure Cellulose and Hemicellulose

The energy potential and conversion efficiency of fast-growing trees and agricultural residues can be influenced by the composition of lignocellulosic matter [48]. Figure 3.5 compares the near-infrared spectra of pure cellulose and pure lignin with 90 samples of fast-growing trees and 110 samples of agricultural residues, all exhibiting average absorbance values. The figure reveals that the vibration band between approximately 5181–6150 cm^{-1} corresponds to the lignin band (with low absorbance for cellulose), while the range between approximately 6150–6800 cm^{-1} corresponds to the cellulose band (with low absorbance for lignin) [34]. Notably, the spectra of fast-growing trees and agricultural residues exhibit distinct peaks resembling those of both pure cellulose and pure lignin at approximately 4019 cm^{-1} , 4405 cm^{-1} , 4762 cm^{-1} , 5181 cm^{-1} , and 6897 cm^{-1} . This resemblance of distinct peaks provides strong evidence that the ground biomass of fast-growing trees and agricultural residues contains cellulose and lignin.

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 the fundamental O-H stretching band in water and starch [49].

The cellulose and lignin content in biomass is a critical factor in determining its HHV. Biomass with a higher lignin content and lower cellulose content exhibits an improved HHV, and vice versa [50]. In the range between 7598 and 12,500 cm^{-1} , the pure lignin curve exhibits significantly higher absorbance values compare to pure cellulose, fast-growing trees and

agricultural residues, reflecting lignin's stronger natural characteristic of absorptivity behavior in this spectral region. This finding confirms the suitability of the selected fast-growing tree and agricultural residue varieties for various applications that rely on lignocellulose matter. These applications include biomass trading for direct combustion, biomass pellet production, the paper and pulp industry, biomass-based construction and building material, and bioenergy and biofuel production, among others.

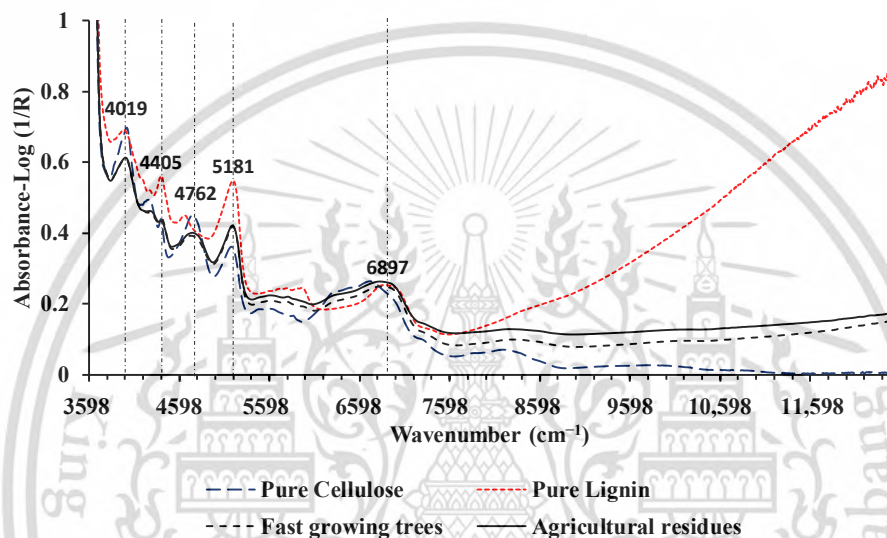


Figure 3.5 Spectra of fast-growing trees and agricultural residues compared to pure cellulose and pure lignin.

3.6.2. Higher Heating Value and Ultimate Analysis in Ground Biomass

Figure 3.6 displays a histogram of the HHV and ultimate analysis values used in the calibration set and validation set for the development of various PLSR-based models. The calibration set is represented by the blue color, while the validation set is represented by the red color. Equation (2) was employed to calculate the outliers, and any data points that satisfied the defined relation were excluded from the model development process.

The normal distribution of all the reference datasets, i.e., HHV (J/g) and wt.% of C, N, H, and O, on a dry basis—used for model development—was analyzed using SPSS 29.0. A histogram analysis revealed that all the datasets exhibit a bell-shaped normal distribution. This suggests that the data points are clustered around the mean value, demonstrating a nearly normal distribution. Additionally, the calculated standard deviations of these datasets (refer to Table 3.1) were found to be low, further indicating that the data points are closely packed around the mean.

Furthermore, a one-sample Kolmogorov–Smirnov test was performed using SPSS 29.0 to calculate the p -values for HHV (J/g) and wt.% of C, N, H, and O, resulting in values of 0.704, 0.060, 0.368, 0.565, and 0.119, respectively. Since all obtained p -values are greater than the significance level of 0.05, the reference data used for modelling is considered to have a uniform distribution. The findings regarding the normal distribution, low standard deviations, and the concentration of data points around the mean value support the validity and reliability of the model developed in this research.

Table 3.2 presents the average HHV, ultimate analysis parameters (C, N, H and O), and ash content of different fast-growing trees and agricultural residues that were included as reference data for developing the model. The HHV is measured in J/g, and the ultimate analysis parameters and ash content are expressed as wt.% on a dry basis. In the case of the biomass samples analyzed using the CHNS/O analyzer (Thermo Scientific FLASH 2000), no sulfur content was detected. This could be attributed to the typically low levels of sulfur present in biomass, which may fall below the lower limit of detection of the analyzer. Therefore, for the purpose of this study, the sulfur content in the ground biomass is assumed to be zero and has not been considered in the model development. The wt.% of O was then calculated using Equation (1). 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 [51]. Table 3.1 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 the measured data of the HHV and elemental composition of ground biomass.

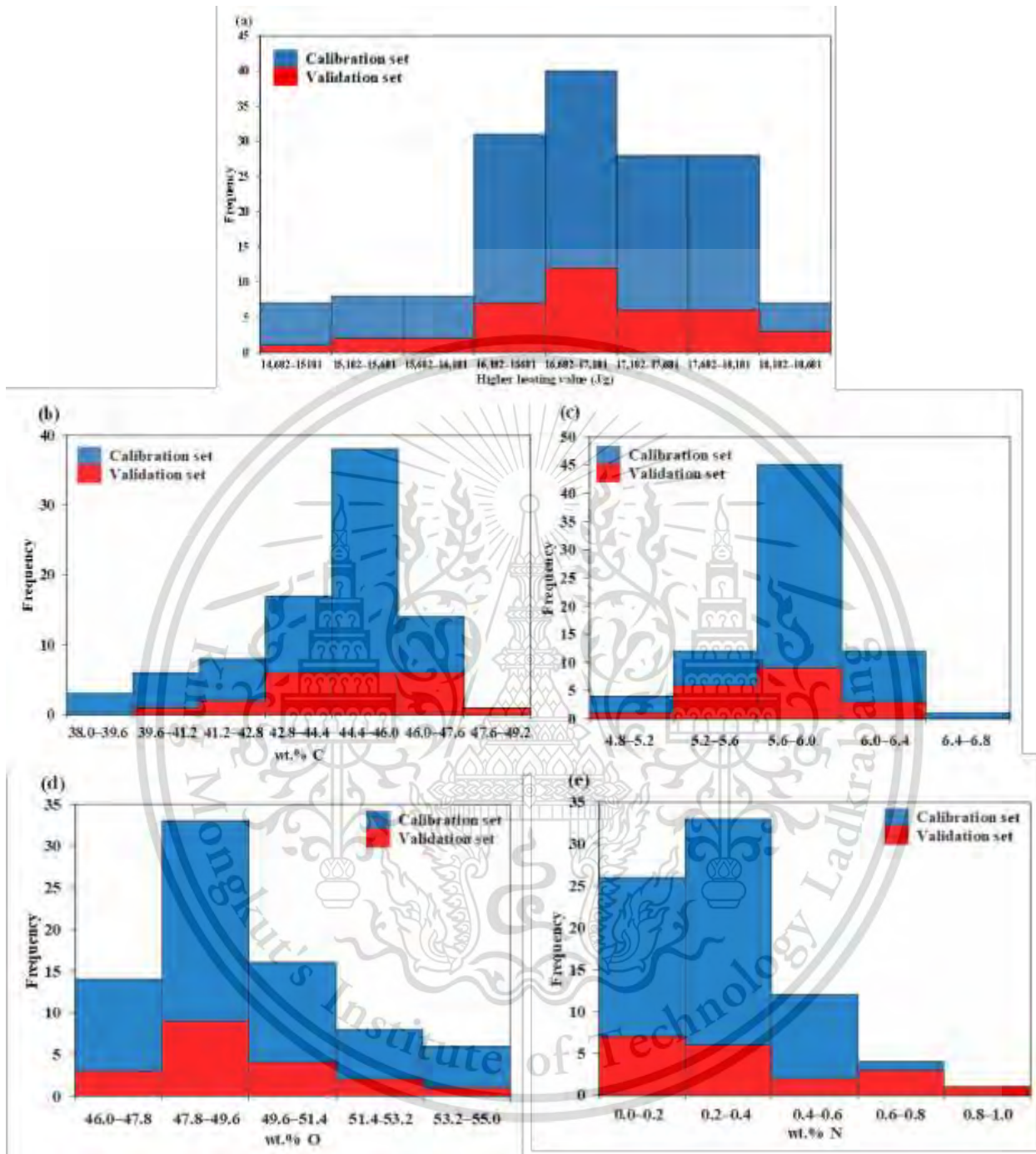


Figure 3.6 Histogram of reference values in calibration and validation sets for the (a) HHV (J/g), (b) wt.% C, (c) wt.% of H, (d) wt.% of O, and (e) wt.% of N.

Table 3.1 shows the statistical summary data for the HHV (J/g) and ultimate analysis parameters, i.e. wt.% of C, N, H, and O on a dry basis used in the calibration set and validation

set for the model development. Similarly, Table 3.2 shows the average reference value of HHV (J/g), ash content (wt.%), and ultimate analysis parameters of fast-growing trees and agricultural residues on a dry basis (wt.%).

The desirable standard error of laboratory (SEL_{desired}) for the bomb calorimeter and CHNS/O elemental analyzer can be calculated using the formula:

$$R^2 \text{ of the best model} = \frac{SD_y^2 - SEL_{\text{desired}}^2}{SD_y^2}$$

Where, R^2 is the coefficient of determination of the best model, and SD_y^2 = standard deviation of the reference value in the calibration set. Therefore, considering the coefficient of determination obtained from the best calibration model for each parameter (refer to Table 3.3) and its respective SD_y^2 , the desirable SEL values for the bomb calorimeter and CHNS/O elemental analyzer were calculated as 252.611 J/g for HHV, 1.3106 wt.% for C, 0.0734 wt.% for N, and 0.1438 wt.% for H, respectively.

Table 3.1 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	18,616	14,682	16,962	848	39	18,553	14,965	17,049	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	99	79	51.1200	37.8200	44.9440	2.5233	20	48.9550	38.8500	44.9505	2.5718

Table 3.2 Average reference value of HHV (J/g), ash content (wt.%), and ultimate analysis parameter of fast-growing trees and agricultural residues on a dry basis (wt.%).

Category	Particular	HHV (J/g)	C (wt.%)	N (wt.%)	H (wt.%)	Ash (wt.%)	O (wt.%)
Fast-growing tree	<i>Alnus nepalensis</i>	17,932	45.9115	0.3115	5.7255	2.3671	45.6844
	<i>Pinus roxiburghii</i>	18,349	46.8367	0.0606	5.8283	2.0900	45.1844
	<i>Bombusa vulagris</i>	17,310	45.6132	0.2327	5.7536	2.8120	45.5884
	<i>Eucalyptus camaldulensis</i>	17,105	44.5536	0.0896	5.6164	3.8158	45.9245
	<i>Bombax ceiba</i>	17,077	44.8557	0.3162	5.8179	5.1271	43.8832
Agricultural residue	<i>Zea mays</i> (cob)	17,297	44.7794	0.2488	5.7619	2.6146	46.5954
	<i>Zea mays</i> (shell)	16,409	45.6518	0.4318	6.2113	3.5826	44.1224
	<i>Zea mays</i> (stover)	16,753	44.3988	0.7069	5.6697	4.0033	45.2212
	<i>Oryza sativa</i>	15,417	40.4261	0.4996	5.3042	13.7073	40.0629
	<i>Saccharum officinarum</i>	17,029	43.6413	0.1047	5.7047	3.0300	47.5194

3.6.2.1. Higher Heating Value

Out of the 200 samples, 4 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 3.3 displays the optimal result of various PLSR-based models using the full wavenumber range (3594.87–12,489.5 cm^{-1}) to evaluate the HHV of the ground biomass from the fast-growing trees and agricultural residues.

Figure 3.7a shows the scatter plot of HHV measured and predicted values from the calibration and validation sets using GA-PLSR. The GA-PLSR with 14 LVs and spectral pretreatment first derivative using 692 important wavenumbers yielded the best performance results, with an R^2_C of 0.9505, RMSEC of 188.0117 J/g, R^2_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 trees and agricultural residues, for model development and had a wider HHV range, resulting in higher R^2_C , R^2_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) [46] and Zornoza et al. (2008) [47], the GA-PLSR model for evaluating HHV is acceptable for most applications with excellent prediction, including quality assurance.



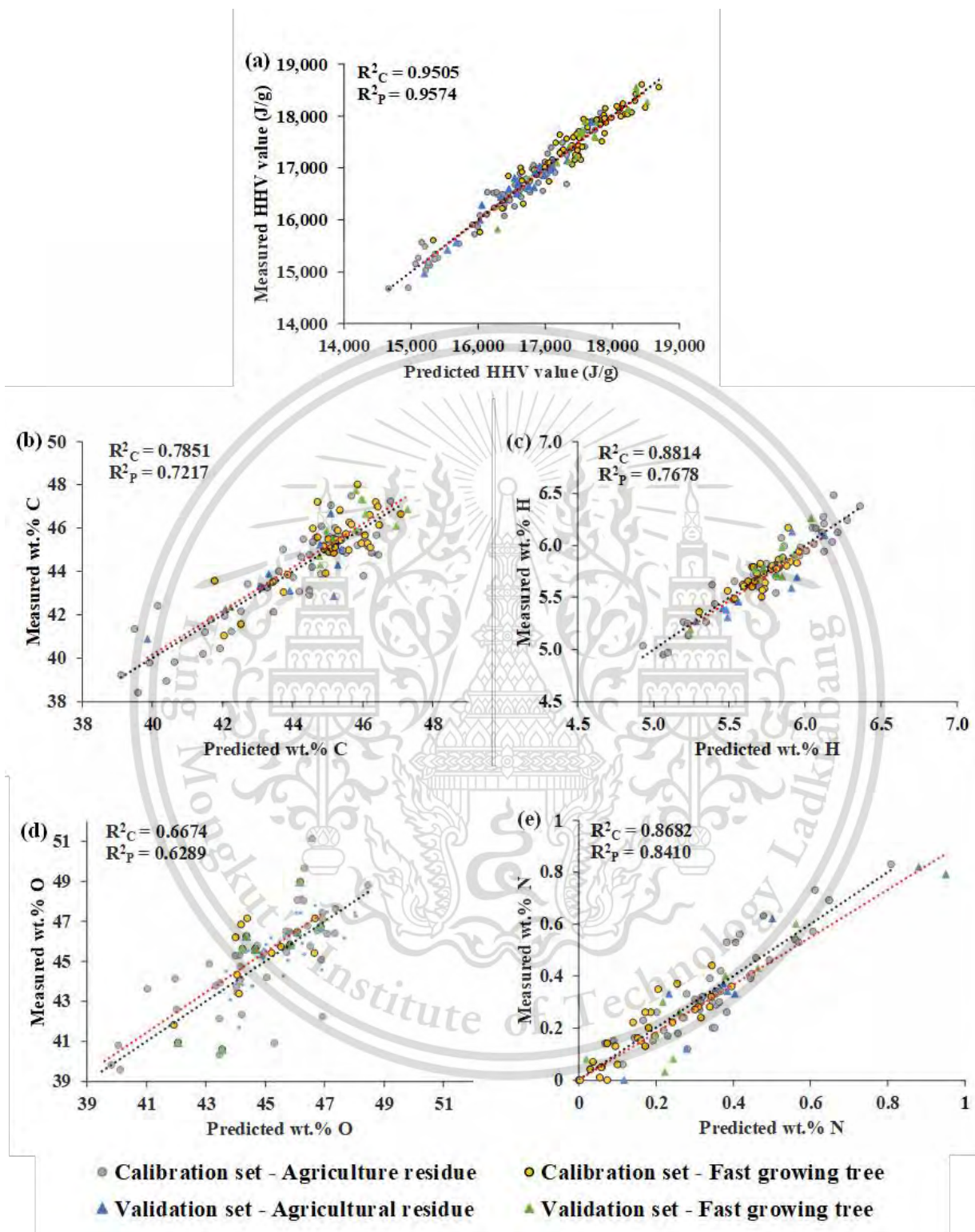


Figure 3.7 Measured versus predicted value in calibration and validation sets for the (a) HHV (J/g), (b) wt.% of C, (c) wt.% of H, (d) wt.% of O, and (e) wt.% of N.

Figure 3.8 shows the average absorbance values obtained after preprocessing with the first derivative, highlighting the 692 selected wavenumbers (marked in red) obtained from GA, which is within the full spectral range of 3594.87–12,489.5 cm^{-1} . The figure highlights important peaks in the following ranges: 4003.73–4111.73 cm^{-1} , 4366.3–4451.16 cm^{-1} , 5091.45–5114.59 cm^{-1} , and 5130.02–5292.02 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 [49]. Lignocellulosic biomass derives its primary energy from cellulose, hemicellulose, and lignin [52,53]. As can be seen in Figure 3.5, 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. [54] and Lestander et al. [55], in which the authors reported that vibration bonds of C-H, C-C, and O-H stretching contribute significantly to the HHV of bamboo and biofuels, respectively. Additionally, Zhang et al. [18] reported that the vibration bond of C-H stretching in an aromatic 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.

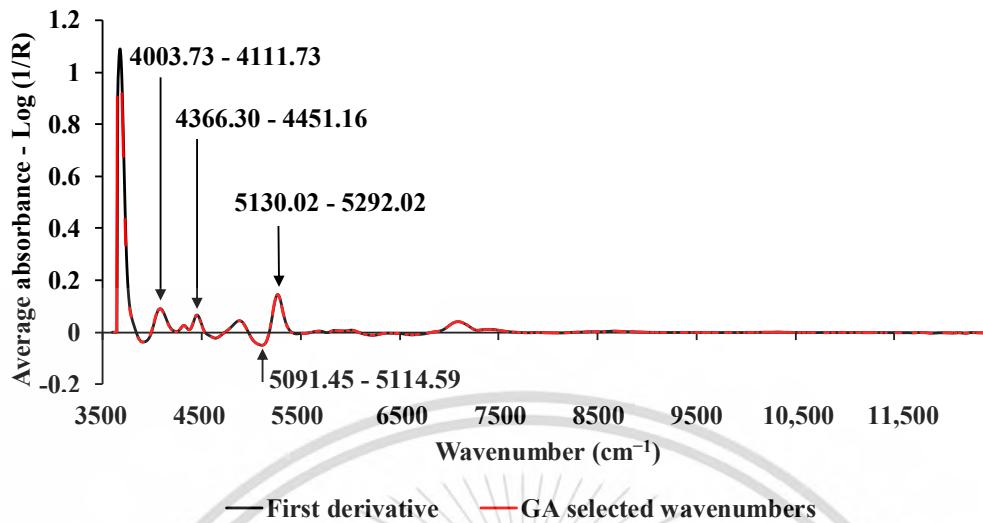


Figure 3.8 The average absorbance value of HHV (J/g) obtained using the first derivative preprocessing with a selection of important wavenumbers obtained from GA, within the full wavenumber range of 3594.87–12,489.5 cm^{-1} .

Table 3.3 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	Preprocessing	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
	Full-PLSR	Raw spectra	13	0.8433	0.8326	0.6129	1.1488	1.67	0.3193
wt.% C	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
	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
wt.% N	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	0.7961	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:	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
Full-PLSR	First derivative (g = 5, s = 5)	12	0.7936	1.1390	0.5972	1.5913	1.5759	-0.0256
SPA-PLSR	First derivative (g = 5, s = 5) (SW=190)	11	0.5938	1.5981	0.6102	1.5654	1.6031	0.0658
wt.% O GA-PLSR	Raw (SW:93)	15	0.6472	1.4892	0.6265	1.5324	1.6602	-0.2593
MP-PLSR: 3-range	Combination set: 4, 6, 4	9	0.6623	1.4570	0.6283	1.5286	1.6591	-0.2222
MP-PLSR: 5-range	Combination set: 3, 2, 4, 6, 0	12	0.6674	1.4461	0.6289	1.5275	1.7147	0.4456

3.6.2.2. Ultimate Analysis

The sulfur content in the ground biomass samples of fast-growing trees and agricultural residues 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[56]. Therefore, PLSR-based models for the wt.% of S were not developed in this study.

3.6.2.2.1 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 3.3 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 3594.87–12,489.48 cm^{-1} . Out of the 120 samples, 11 samples were identified as an outlier and removed from the total dataset for model development. The model developed through GA-PLSR with spectrum preprocessing of first derivative (gap = 5 and segment = 5) and 9 LVs provided better results with an R^2_C , RMSEC, R^2_p , RMSEP, RPD, and bias value of 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 3.7b. According

to the recommendation by Williams et al. (2019) [46], the PLSR model with the 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) [47], the model is acceptable for the prediction of wt.% C content in the grounded biomass.

Figure 3.9 shows the average absorbance values obtained after preprocessing with the first derivative, highlighting the 50 selected wavenumbers (marked in red) obtained from the GA, which is within the full spectral range of $3594.87\text{--}12,489.5\text{ cm}^{-1}$. 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 3.9, 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 ($-\text{CH}_2\text{-OH}$), and the 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 a combination of N-H/C-N/N-H amide II and amide III; spectral structure O-H stretching and C-O stretching; and 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 an O-H aromatic with the spectral structure of an O-H first overtone of the fundamental stretching band, as well as the material type of hydrocarbons [49]. Lignin contains a high carbon content [57]. According to Zhang et al. [19], vibration bands related to C-H stretching, CH_2 , C-H aromatics, O-H stretching, and HOH deformation are essential for predicting the C content of sorghum biomass. Similarly, Posom and Sirisomboon [58] 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 the 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.

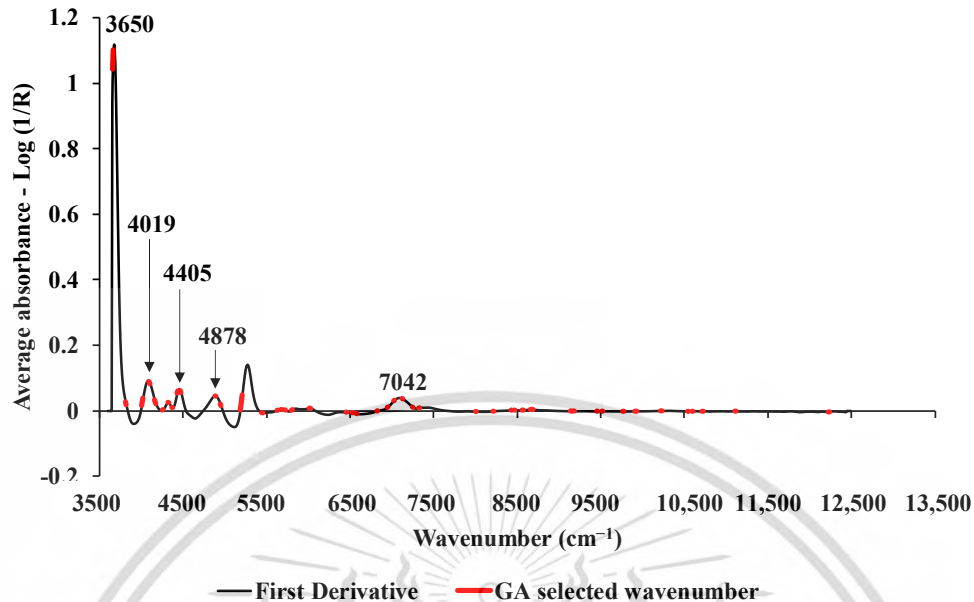


Figure 3.9 The average absorbance value of wt.% of C obtained using the first derivative preprocessing with a selection of important wavenumbers obtained from GA, within the full wavenumber range of 3594.87–12,489.5 cm^{-1} .

3.6.2.2.2 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 optimal results of different PLSR-based models for evaluation of wt.% of H within the full wavenumber range were presented in Table 3.3. 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 3594.87–12,489.48 cm^{-1} and spectral preprocessing from SNV. The best performing model for the evaluation of wt.% of H content in the grounded biomass produced an 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 3.7c 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) [46], 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 a 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 3.10 shows the average absorbance spectrum that was pretreated with the SNV and uses red marks to highlight 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, and 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 a combination of O-H stretching and HOH bending in water. The peak at 6897 cm^{-1} corresponds to the spectral structure of O-H, arising from the first overtone of the fundamental stretching band, with a material-type starch/polymeric alcohol. The peak at 8163 cm^{-1} is associated with the second overtone of the C-H fundamental stretching band and material-type hydrocarbons [49]. The selected peaks mostly fall within a similar range compared to the study conducted by Posom and Sirisomboon [58]. 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.

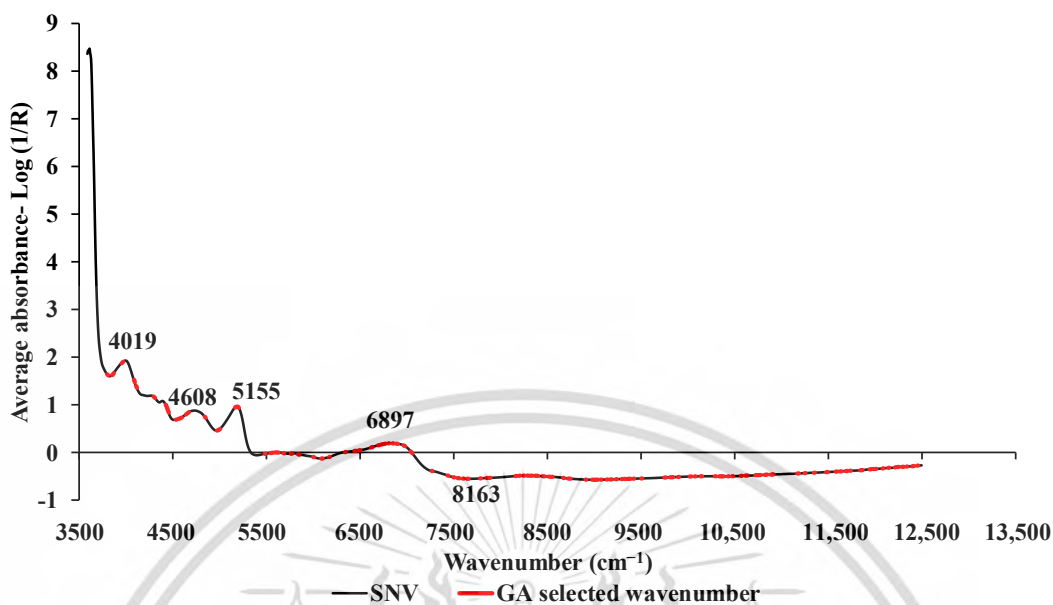


Figure 3.10 The average absorbance value of wt.% of H obtained using SNV preprocessing with a selection of important wavenumbers obtained from GA, within full wavenumber range of 3594.87–12,489.5 cm^{-1} .

3.6.2.2.3 wt.% of O

Based on the assumption that the sulfur content in biomass is zero, as its wt.% is too low to be detected by instruments, the wt.% of O in biomass is calculated using Equation (1). The optimal results of the PLSR-based models for predicting the wt.% of O content in the grounded biomass are shown in Table 3.3. Before modelling, outliers from the reference values were calculated, and 21 out of 120 samples were detected as outliers. Therefore, 99 grounded biomass samples were used for the model development. The best result was obtained from the multi-preprocessing PLSR 5-range method with a spectral preprocessing combination set of 3, 2, 4, 6, and 0, i.e. MSC, SNV, first derivative, constant offset, and empty, respectively, from the range 3625.72–12,489.48 cm^{-1} , which are equally divided into five sections. Figure 3.7d shows the scatter plot for the measured and predicted wt.% of O. With 12 LVs, the best performing model for evaluating the wt.% of O content in the grounded biomass produced an R^2_c of 0.6674, RMSEC of 1.4461 wt.%, R^2_p of 0.6289, RMSEP of 0.1.5275 wt.%, RPD of 1.7147, and a bias of -0.4456 wt.%. Compared with full-PLSR, the multi-preprocessing 5-range method improved the model accuracy by 4.0085%. Based on Williams et al. (2019) [46] and Zornoza et al. (2008) [47], the model with the multi-preprocessing PLSR 5-range method is usable only for rough screening. Therefore, to improve the performance of the model, the inclusion of a larger number of representative

samples spanning a wide range of oxygen contents is recommended. This will enable the model to capture the variability in oxygen levels across different biomass compositions. Additionally, minimizing instrumental errors through proper calibration and maintenance of the CHNS/O analyzer and thermogravimetric analyzer is essential. Exploring alternative methods for measuring the ash content in the biomass could also contribute to improving the accuracy of wt.% of O predictions.

Figure 3.11 shows the regression coefficient plot for wt.% of O content in the grounded biomass, which is obtained from the multi-preprocessing PLSR 5-range method. Significant peaks were observed at wavenumbers 3650, 5155, 5675, 5952, 6330, and 7042 cm^{-1} . The peak at 3650 cm^{-1} corresponds to the O-H functional group typically found in primary alcohols. Similarly, the peaks at 5155 cm^{-1} represent a combination of O-H stretching and HOH bending in water. The negative peak at 5675 cm^{-1} and the positive peak at 5952 cm^{-1} are associated with the spectral structure of the first overtone of the fundamental stretching band of C-H, with hydrocarbons, methylene, and aromatic hydrocarbons as the material types, respectively. The peak at 6330 cm^{-1} corresponds to the functional group of the O-H combination band observed in alcohols, such as R-C-O-H. The peak at 7042 cm^{-1} corresponds to the first overtone of the fundamental stretching band of O-H, which is typically present in hydrocarbons and aromatic compounds [49]. A previous study by Posom and Sirisomboon [58] showed peaks at similar wavenumbers with vibration bands of C-H aromatic, O-H stretching of alcohol, O-H stretching, and HOH bending of water, which supports the findings of this study. Hence, these vibration bands may have a significant influence on the development of the model for the assessment of wt.% of O in grounded biomass.

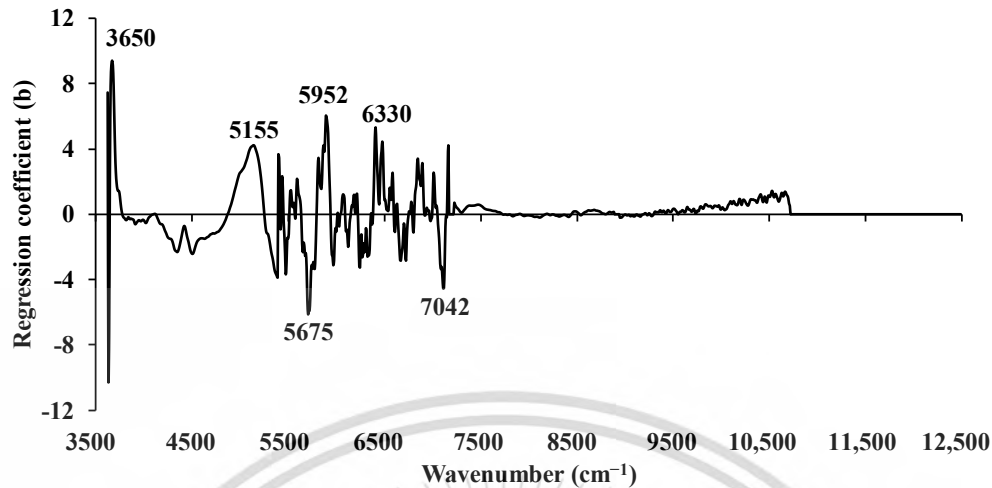


Figure 3.11 The regression coefficient for the wt.% O of grounded biomass using the multi-preprocessing PLSR 5-range method.

3.6.2.2.4 wt.% of N

The SEL of the CHNS elemental analyzer for evaluating the wt.% of N content in grounded biomass was calculated as 0.0761 wt.%. Table 3.3 shows the optimal outcomes of the PLSR-based models for predicting the wt.% of N content in grounded biomass. Out of the 120 samples, 25 samples were identified as an outlier and removed from the total dataset for model development. 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 3625.72–12,489.48 cm^{-1} . Figure 3.7e shows the scatter plot of the 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) [46], the model is suitable for most applications, including research. Based on the recommendation of Zornoza et al. (2008) [47], the prediction of wt.% of N content from the multi-preprocessing PLSR 5-range method with an RPD value of 2.65 is considered good for prediction.

Figure 3.12 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 12,453 cm^{-1} might significantly contribute to the evaluation of wt.% of N content. The negative peak at 4019 cm^{-1} might correspond to a C-H stretching and C-C stretching combination with the material type shown 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 a 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) and 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 the material type shown as starch/polymeric alcohol. The common natures of the peaks were noticed in the range between 11,500 and 12,500 cm^{-1} , for which 12,453 cm^{-1} is described as a reference, which might correspond to the spectral structure of a C-H combination, with the material type being hydrocarbon and aliphatic [49]. The selected regression coefficient peaks show similar peaks compared to the study performed by Posom and Sirisomboon [58], with vibration bands of C-H stretching, C-C stretching, O-H stretching, and HOH 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.

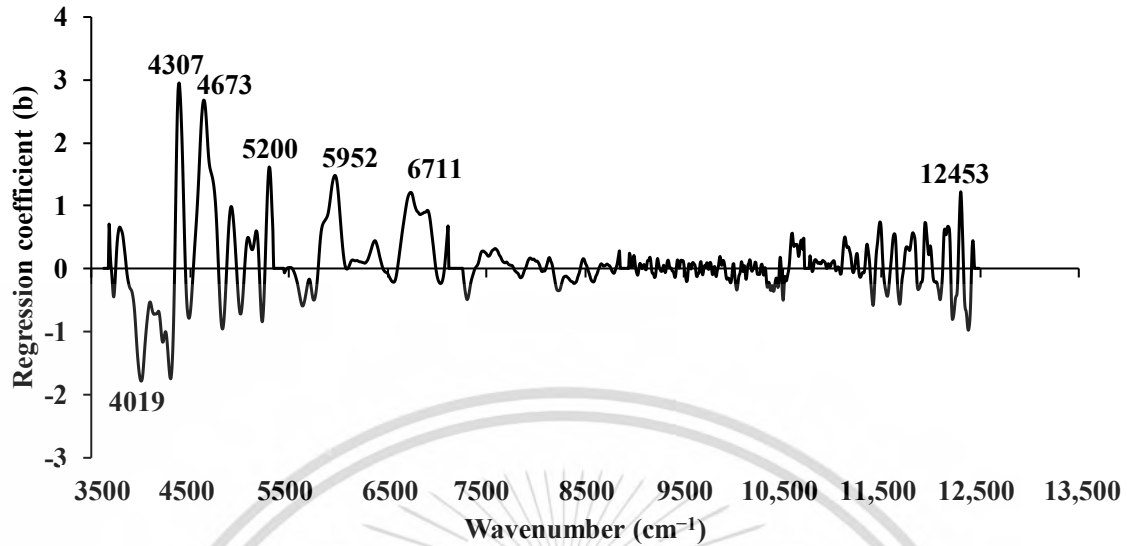


Figure 3.12 The regression coefficient for the wt.% of N of grounded biomass using the multi-preprocessing PLSR 3-range method.

3.7. Comparison with Previous Work

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

In a previous study, Nakawajana et al. [59] evaluated the HHV of grounded cassava rhizome using PLSR and achieved an R^2 of 0.90. Similarly, Nakawajana et al. [34], Posom et al. [3], Zhang et al. [18], and Posom et al. [5] developed PLSR models for rick husk, grounded bamboo, sorghum biomass, and *Leucaena leucocephala* pellets, respectively, with an 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 transfectance 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 [58] 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. [58] showing R^2_p values of 0.80 for C, 0.85 for H, and 0.97 for N for bamboo. Similarly, the models developed by Zhang et al. [18] 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 [60] predicted the ultimate parameters of torrefied 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 a bomb calorimeter and CHNS/O elemental analyzer could have played a catalytic role in achieving a higher model performance when evaluating HHV and N. However, the lower model performance for evaluating C, O, 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, O, and H content in biomass. To enhance the model performance for evaluating C, O, and H content, the number of representative samples with a high concentration of C, O, 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 a 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 transmittance mode rather than diffuse reflectance mode, and by applying a spectral multi-preprocessing technique. To update the model for robust application, the number of ground biomass samples must be increased and validated using unknown samples.

3.8. Conclusions

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 to assess the biomass properties for energy usage. The model with the optimum performance was selected based on the parameters of R^2_C , RMSEP, R^2_P , RMSEP, RPD, and bias. The models for HHV (J/g), and wt.% of N are suitable for most applications, including research, while the models for wt.% of C, wt.% of O, and wt.% of H were 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, thereby considering the wide statistical range of the reference values.

This study showed that the multi-preprocessing 5-range method, a novel 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 a 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 when 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, policymaking think tanks, and energy companies in planning for the proper identification, management, and utilization of bio-resources to meet future energy demand and supply. The research outcomes also generate possibilities for NIR-based research to adopt or apply similar approaches.

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*, 7205–7219.
2. Diez, 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*, 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. *J. Near Infrared Spectrosc.* **2017**, *25*, 301–310. <https://doi.org/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. *J. Mater. Cycles Waste Manag.* **2019**, *21*, 786–800. <https://doi.org/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 Util. Environ. Eff.* **2007**, *30*, 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. *Int. J. Coal Prep. Util.* **2019**, *42*, 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 Bioenergy* **2005**, *28*, 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 Convers. Manag.* **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*, 1793–1797. <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*, 38. <https://doi.org/10.3390/foods6050038>.
12. Roger, J.-M.; Mallet, A.; Marini, F. Preprocessing NIR Spectra for Aquaphotomics. *Molecules* **2022**, *27*, 6795. <https://doi.org/10.3390/molecules27206795>.
13. Maraphum, K.; Saengprachatanarug, K.; Wongpichet, S.; Phuphaphud, A.; Posom, J. Achieving robustness across different ages and cultivars for an NIRS-PLSR model of fresh cassava root starch and dry matter content. *Comput. Electron. Agric.* **2022**, *196*, 106872. <https://doi.org/10.1016/j.compag.2022.106872>.
14. Posom, J.; Phuphaphud, A.; Saengprachatanarug, K.; Maraphum, K.; Saijan, S.; Pongkan, K.; Srimai, K. Real-time measuring energy characteristics of cane bagasse using NIR spectroscopy. *Sens. Bio-Sens. Res.* **2022**, *38*, 100519. <https://doi.org/10.1016/j.sbsr.2022.100519>.
15. Phuphaphud, 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.; Saechua, W. Prediction of Elemental Components of Ground Bamboo Using Micro-NIR Spectrometer. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *301*, 012063.
17. Skvaril, J.; Kyprianidis, K.G.; Dahlquist, E. Applications of near-infrared spectroscopy (NIRS) in biomass energy conversion processes: A review. *Appl. Spectrosc. Rev.* **2017**, *52*, 675–728. <https://doi.org/10.1080/05704928.2017.1289471>.
18. 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>.
19. 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>.

20. Pokhrel, D.R.; Sirisomboon, P.; Khurnpoon, L.; Posom, J.; Saechua, W. Comparing Machine Learning and PLSDA Algorithms for Durian Pulp Classification Using Inline NIR Spectra. *Sensors* **2023**, *23*, 5327.
21. Phanomsophon, T.; Jaisue, N.; Worphet, A.; Tawinteung, N.; Shrestha, B.; Posom, J.; Khurnpoon, L.; Sirisomboon, P. Rapid measurement of classification levels of primary macronutrients in durian (*Durio zibethinus* Murray CV. Mon Thong) leaves using FT-NIR spectrometer and comparing the effect of imbalanced and balanced data for modelling. *Measurement* **2022**, *203*, 111975.
22. Meenu, M.; Cozzolino, D.; Xu, B. Non-destructive prediction of total phenolics and antioxidants in hulled and naked oat genotypes with near-infrared spectroscopy. *J. Food Meas. Charact.* **2023**, 1–12. <https://doi.org/10.1007/s11694-023-02009-0>.
23. Posom, J.; Maraphum, K. Achieving prediction of starch in cassava [*Manihot esculenta* Crantz] by data fusion of VIS-NIR and Mid-NIR spectroscopy via machine learning. *J. Food Compos. Anal.* **2023**, *122*, 105415.
24. Santos, F.D.; Santos, L.P.; Cunha, P.H.; Borghi, F.T.; Romao, W.; de Castro, E.V.; de Oliveira, E.C.; Filgueiras, P.R. Discrimination of oils and fuels using a portable NIR spectrometer. *Fuel* **2021**, *283*, 118854.
25. Kranenburg, R.F.; Ramaker, H.J.; Sap, S.; van Asten, A.C. A calibration friendly approach to identify drugs of abuse mixtures with a portable near-infrared analyzer. *Drug Test. Anal.* **2022**, *14*, 1089–1101.
26. Chung, H.; Ku, M.-S. Comparison of near-infrared, infrared, and Raman spectroscopy for the analysis of heavy petroleum products. *Appl. Spectrosc.* **2000**, *54*, 239–245.
27. Stauffer, M. (Ed). *Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences*; IntechOpen: Rijeka, Croatia, 2016.
28. Durickovic, I. Using Raman spectroscopy for characterization of aqueous media and quantification of species in aqueous solution. In *Applications of Molecular Spectroscopy to Current Research in the Chemical and Biological Sciences*; Stauffer, M., Ed.; IntechOpen: Rijeka, Croatia, 2016; p. 405.
29. Jiao, Y.; Li, Z.; Chen, X.; Fei, S. Preprocessing methods for near-infrared spectrum calibration. *J. Chemom.* **2020**, *34*, e3306. <https://doi.org/10.1002/cem.3306>.

30. Qian, H.; Guo, X.; Fan, S.; Hagos, K.; Lu, X.; Liu, C.; Huang, D. A simple prediction model for higher heat value of biomass. *J. Chem. Eng. Data* **2016**, *61*, 4039–4045.
31. Posom, J.; Maraphum, K.; Phuphaphud, A. Rapid Evaluation of Biomass Properties Used for Energy Purposes Using Near-Infrared Spectroscopy. In *Renewable Energy-Technologies and Applications*; IntechOpen: Bristol, UK, 2020.
32. 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 Anal. Chem.* **2019**, *113*, 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>.
33. 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. *Handb. Vib. Spectrosc.* **2002**, *5*, 3590–3610.
34. 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. *Eng. J.* **2018**, *22*, 45–56. <https://doi.org/10.4186/ej.2018.22.5.45>.
35. 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*, 2001–2012. <https://doi.org/10.1177/0003702817704147>.
36. 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 Chem. X* **2023**, *17*, 100539. <https://doi.org/10.1016/j.fochx.2022.100539>.
37. Shetty, N.; Gislum, R. Quantification of fructan concentration in grasses using NIR spectroscopy and PLSR. *Field Crops Res.* **2011**, *120*, 31–37. <https://doi.org/10.1016/j.fcr.2010.08.008>.
38. Conzen, J. *Multivariate Calibration: A Practical Guide for Developing Methods in the Quantitative Analytical Chemistry*; BrukerOptik GmbH. Ettlingen, Germany, 2006.
39. 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*, 119698. <https://doi.org/10.1016/j.energy.2020.119698>.

40. 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. *Int. J. Anal. Chem.* **2017**, *2017*, 6018769. <https://doi.org/10.1155/2017/6018769>.
41. Zhang, C.; Liu, F.; Kong, W.; Zhang, H.; He, Y. Fast identification of watermelon seed variety using near infrared hyperspectral imaging technology. *Trans. Chin. Soc. Agric. Eng.* **2013**, *29*, 270–277.
42. Liu, D.; Sun, D.-W.; Zeng, X.-A. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food Bioprocess Technol.* **2014**, *7*, 307–323.
43. Santos-Rufo, A.; Mesas-Carrascosa, F.-J.; García-Ferrer, A.; Meroño-Larriva, J.E. Wavelength selection method based on partial least square from hyperspectral unmanned aerial vehicle orthomosaic of irrigated olive orchards. *Remote Sens.* **2020**, *12*, 3426. <https://doi.org/10.3390/rs12203426>.
44. 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. *Eng. Appl. Sci. Res.* **2022**, *49*, 119–126. Available online: <https://ph01.tci-thaijo.org/index.php/easr/article/view/245217> (11 July 2023).
45. 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 Sens.* **2016**, *8*, 755. <https://doi.org/10.3390/rs8090755>.
46. Williams, P.; Manley, M.; Antoniszyn, J. *Near Infrared Technology: Getting the Best Out of Light*; African Sun Media: Stellenbosch, South Africa, 2019.
47. 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 Biol. Biochem.* **2008**, *40*, 1923–1930. <https://doi.org/10.1016/j.soilbio.2008.04.003>.
48. 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. *Appl. Sci.* **2019**, *9*, 2252. <https://doi.org/10.3390/app9112252>.
49. Weyer, L. *Practical Guide to Interpretive Near-Infrared Spectroscopy*; CRC Press: Boca Raton, FL, USA, 2007.
 50. Demirbaş, A. Relationships between lignin contents and heating values of biomass. *Energy Convers. Manag.* **2001**, *42*, 183–188.
 51. Hasan, M.; Haseli, Y.; Karadogan, E. Correlations to predict elemental compositions and heating value of torrefied biomass. *Energies* **2018**, *11*, 2443. <https://doi.org/10.3390/en11092443>.
 52. Zoghalmi, A.; Paës, G. Lignocellulosic biomass: Understanding recalcitrance and predicting hydrolysis. *Front. Chem.* **2019**, *7*, 874. <https://doi.org/10.3389/fchem.2019.00874>.
 53. 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: Amsterdam, The Netherlands, 2018; Volume 3, pp. 161–213.
 54. Sirisomboon, P.; Funke, A.; Posom, J. Improvement of proximate data and calorific value assessment of bamboo through near infrared wood chips acquisition. *Renew. Energy* **2020**, *147*, 1921–1931. <https://doi.org/10.1016/j.renene.2019.09.128>.
 55. 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*, 1182–1189. <https://doi.org/10.1039/B500103J>.
 56. 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>.
 57. 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. *Bioresour. Technol.* **2009**, *100*, 292–298. <https://doi.org/10.1016/j.biortech.2008.06.009>.
 58. 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>.

59. Nakawajana, N.; Posom, J. 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 Conf. Ser. Earth Environ. Sci.* **2019**, *301*, 012032.
60. 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>.



Chapter 4

Evaluating limit of detection and quantification for higher heating value and ultimate analysis of fast-growing trees and agricultural residues biomass using NIRS

Accurate non-destructive assessment of biomass energy properties is essential for optimizing its use as an alternative fuel. In this study, 200 biomass samples were used to determine higher heating value (HHV) and 120 biomass samples for analyzing ultimate analysis parameters using near-infrared spectroscopy within the full wavenumber range of 12489.48 – 3594.87 cm^{-1} . The samples were grounded, and five different types of partial least squares regression (PLSR) models were developed using traditional preprocessing, multi-preprocessing (MP) with 5 range, MP with 3 range, genetic algorithm, and successive projection algorithm. Limit of detection (LOD) and quantification (LOQ) were calculated using the best-performing model among five different PLSR models for HHV in kJ/kg, as well as the weight percentage (wt.%) of carbon (C), oxygen (O), hydrogen (H), and nitrogen (N). The LOD and LOQ for HHV were calculated as 622.42 kJ/kg and 1886.13 kJ/kg, respectively. Additionally, LOD and LOQ for ultimate analysis parameters, including C, O, H, and N were calculated as: 3.24 weight percentage (wt.%) and 9.81 wt.% for C, 2.04 wt.% and 6.18 wt.% for O, 0.35 wt.% and 1.05 wt.% for H, and 0.22 wt.% and 0.68 wt.% for N. The LOD and LOQ values for HHV, C, O, and H were lower than the minimum reference values used for model development, demonstrating the models' high sensitivity and potential to reliably detect and precisely quantify these parameters. However, the LOD and LOQ values exceeded the minimum reference value used during model development for the N, indicating that the selected models have certain limitations in assessing the N content in biomass. The sample range should be expanded for wt.% of N to enhance the model's performance, surpassing the LOD and LOQ values. This will improve the overall sensitivity of the model for reliable detection and quantification of N content in biomass samples.

Keywords: Biomass, Near-infrared spectroscopy, Partial least squares regression, Limit of detection, Limit of quantification



*This chapter constituted the publication article: Shrestha, B.; Shrestha, Z.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Evaluating limit of detection and quantification for higher heating value and ultimate analysis of fast-growing trees and agricultural residues biomass using NIRS. *Engineering and Applied Science Research* 2023, 50 (6), 612-618. <https://ph01.tci-thaijo.org/index.php/easr/article/view/253649>

Abbreviations

%	percentage
A	ash content
C	carbon
FT	fourier transform
GA	genetic algorithm
H	hydrogen
HHV	higher heating value
LOD	limit of detection
LOQ	limit of quantification
LVs	number of latent variables
Mean	average
MSC	multiplicative scatter correction
MP	multi-preprocessing
N	nitrogen
N_c	number of sample in training set
NIRS	near infrared spectroscopy
O	oxygen
PLSR	partial least squares regression
R^2_c	coefficient of determination of training

4.1. Introduction

Biomass, derived from organic matter such as wood, crops, and waste, is a renewable energy source with the potential to mitigate greenhouse gas emissions and enhance energy security [1]. Due to its ability to be replenished relatively quickly, biomass is considered a sustainable energy source [2]. In recent years, biomass has gained significant attention as a renewable energy option, offering extensive applications in cooking, heating, electricity generation, transportation, agriculture waste management, and industrial processes [3, 4]. However, biomass can be challenging to handle due to its energy properties, including HHV, proximate analysis parameters (moisture content, volatile matter, fixed carbon, ash content), and ultimate analysis parameters C, H, N, S and O [5], which can vary based on factors such as biomass type, harvesting conditions, storage conditions, and during transportation [6]. The fluctuating quality of biomass feedstocks can significantly affect their suitability for efficient energy conversion [7].

The HHV is a crucial characteristic of biomass fuel as it is mainly used to calculate combustion efficiency and the amount of energy that can be generated by the fuel [8]. Carbon, the primary component of biomass, serves as the main energy source during combustion. The carbon content can vary depending on the biomass type, but it typically ranges around 50% [9]. Oxygen, the second most prevalent element in biomass, plays a significant role in the combustion process [10]. The oxygen content in biomass can vary, but it is typically around 40%. Hydrogen, the third most prevalent element in biomass, is an important component of many biofuels. In addition to its clean-burning properties, hydrogen can be utilized for heat and electricity generation. Evaluating the feasibility of hydrogen production from biomass requires determining the hydrogen content within the biomass [11]. The amount of hydrogen in biomass varies depending on the type but is typically around 6%. Nitrogen, despite being a minor component in biomass, has an impact on both combustion efficiency and emissions during combustion. The nitrogen content in biomass can vary, typically ranging between 1% and 2% [12]. Biomass also contains sulfur (S) in quantities less than 1%. Even in small amounts, sulfur can contribute to emissions such as sulfur dioxide [13]. Therefore, the fast and precise prediction of biomass energy properties is crucial for the effective and efficient utilization of biomass resources.

NIRS, a fast, reliable, and non-invasive method, has been widely utilized for quality control purposes in various biomass applications. It enables the rapid and accurate assessment of

important parameters such as HHV, proximate and ultimate analysis parameters [14]. This technique offers improved prediction accuracy models, thereby facilitating efficient and precise characterization of biomass properties.

In our prior research, we conducted a thorough evaluation of ground biomass characteristics for energy applications. We employed NIRS and spectral multi-preprocessing methods to enhance the performance of a PLSR model, specifically focusing on predicting HHV and parameters related to ultimate analysis [15]. However, to comprehensively evaluate the sensitivity of the selected NIR-based models, it is essential to assess two key parameters: the LOD and the LOQ [16]. In NIRS, the calculation of LOD and LOQ is influenced by factors such as the type of analyte, the instrument used, and the employed method [17]. These parameters are important in NIR modeling as they determine the minimum detectable and quantifiable concentration of the substance being measured, ensuring reliable predictions. LOD and LOQ play a vital role in regression analysis by ensuring the validity of analysis findings [18]. They enhance the model's reliability by minimizing false positives and negatives, enabling effective comparison of results across various analyses. Additionally, LOD and LOQ ensure data quality by establishing limits for allowable substance concentrations in samples. This further improves the usefulness of these measures in selecting the optimal training range.

To our knowledge, no research has been reported regarding the calculation of LOD and LOQ for HHV and the ultimate analysis parameter of ground biomass. Therefore, the main objectives of this research were to assess the LOD and LOQ based on the performance of the best training model for HHV and the ultimate analysis parameters to improve the model sensitivity.

4.2. Materials and methods

4.2.1 Samples

The commonly used biomass samples from fast-growing trees (5) and agricultural residues (5) were collected from various locations in Nepal. The samples included: 1) *Alnus nepalensis*, 2) *Pinus roxburghii*, 3) *Bombusa vulagris*, 4) *Bombax ceiba*, 5) *Eucalyptus camaldulensis*, 6) *Zea mays* (cob), 7) *Zea mays* (Shell), 8) *Zea mays* (stover), 9) *Oryza sativa*, and 10) *Saccharum officinarum*. In this study, 200 biomass samples were used to develop a PLSR-based models for HHV, while 120 biomass samples were used for the ultimate analysis parameters [15]. Each sample varieties were grounded and scanned using FT-NIRS, and its HHV was measured using bomb calorimeter. Similarly, the ultimate analysis parameters, wt.% of C, H, N, and S, were measured using the CHNS/O analyzer. The wt.% of O was calculate as a difference from wt.% of C, H, N, S and A [19]. Here, the wt.% of A is measured employing a thermogravimetric analyzer (TG 209 F3 Tarsus, Netzsch, Bavaria, Germany).

4.2.2 Instrument

FT-NIRS (MPA, Bruker, Ettlingen, Germany) was used to log the spectral absorbance data of ground biomass samples in transreflectance mode within a wavenumber range of 12489.48 – 3594.87 cm^{-1} [15]. Figure 4.1 illustrates the representative particle size distribution of the ground biomass, covering a range from 0.01 to 3080 μm . The ground biomass was placed up to a height of 10 mm inside a glass vial with a diameter of 20 mm and a height of 48 mm for scanning the samples in a controlled laboratory environment at 25 ± 2 °C. Although the amount of ground biomass sample used for scanning in the tranflectance mode is small, the variation of the constituent is less, and light is diffuse in the homogenous ground biomass, allowing for better penetration and diffusion to obtain high-quality spectra. Each sample underwent two scans, and the resulting average value was used for the model development.

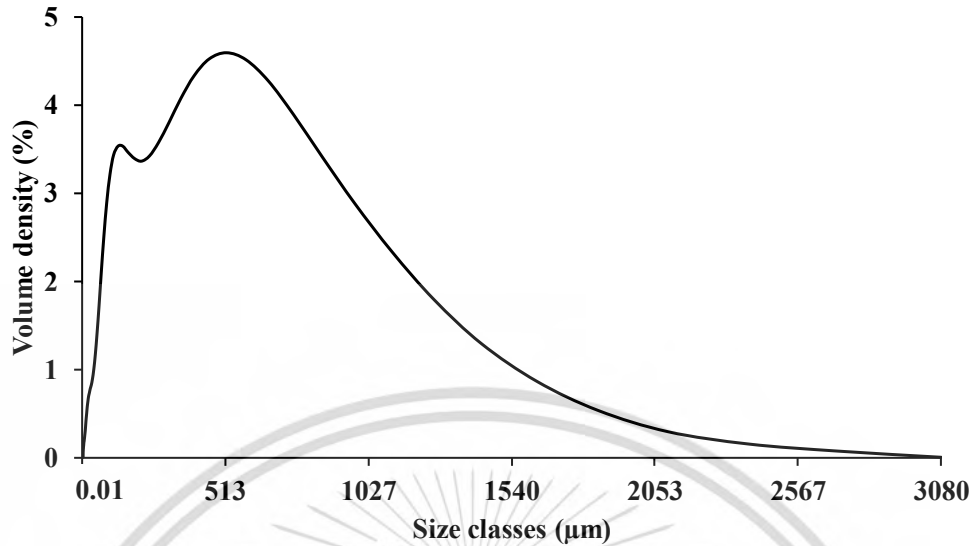


Figure 4.1 Representative particle size distribution of the ground biomass ranging from 0.01 to 3080 μm .

A bomb calorimeter (IKA C 200, Germany) was used to measure the HHV, and the CHNS/O elemental analyzer (Thermo Scientific™ FLASH 2000) was used to measure wt.% of the ultimate analysis parameters. The HHV was measured in kJ/kg, while the ultimate analysis parameters were measured as wt.%. The reference parameters for each biomass sample were measured twice, and their average were regarded as the reference values for the model development. The CHNS/O elemental analyzer did not detect the S content in the grounded biomass samples. Therefore, the sulfur content has not been taken into account for the calculation of LOD and LOQ in this study.

4.2.3. Spectral preprocessing and Model development

After logging spectral data (independent variable) and their respective reference data (dependent variable), the total data set for each evaluating parameter is prepared to develop a training model. First, outliers were calculated from the entire dataset as follows:

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

Where X_i represents the measured value of sample i , while \bar{X} and SD denote the average and standard deviation of the measured values of all samples [20].

If equation (i) is satisfied, a sample is considered an outlier and is removed from the dataset. After the removal of outliers, the total dataset is manually divided into a training set (80%) and a

testing set (20%) to develop a model. The training set includes samples incorporating both the highest and lowest reference values to facilitate model development.

PLSR, recognized as a leading regression technique for NIR data analysis, is highly effective in developing calibration models due to its ability to handle the data's multi-collinearity and high dimensionality [21-25]. Therefore, this study developed five different types of PLSR-based models, which are as follows:

1. Full wavenumber-PLSR
2. MP-PLSR – 3 range method
3. MP-PLSR – 5 range method
4. GA-PLSR
5. SPA-PLSR

Full wavenumber PLSR is a conventional approach for developing PLSR models. It involves preprocessing the spectra using a single preprocessing technique applied across the entire wavelength range. Before model development, the spectral data were pretreated using following preprocessing techniques, and individual models were developed for each technique: 1) sd_1 , 2) sd_2 , 3) constant offset, 4) SNV, 5) MSC, 6) vector normalization, 7) min-max normalization, 8) mean centering, 9) sd_1 + vector normalization, and 10) sd_1 + MSC. The training model is then developed using raw spectra and each preprocessing technique and is validated using a testing set. In contrast, the MP technique, a new approach [15], divides the full wavelength range into distinct sections. For example, in the five-range method, the full wavenumber range is divided into five segments, and in the three-range method, it is divided into three segments. Each section is subjected to preprocessing using a combination set of 5 for the 5 range method and 3 for the 3 range method from 7 different preprocessing methods (codes: 0 = empty (all the absorbance values = 0), 1 = raw spectra, 2 = SNV, 3= MSC, 4= sd_1 , 5= sd_2 , and 6 = constant offset). For example, the combination set 3, 0, 1, 0, and 1 which corresponds to MSC, empty, raw, empty, raw for 5 range method, signifies segmenting the entire spectral range into 5 equal segmentations and preprocessing them as follows: MSC from 3625.72–5392.30 cm^{-1} , empty from 5400.02–7166.59 cm^{-1} , raw spectra from 7174.31–8940.89 cm^{-1} , empty from 8948.60–10,715 cm^{-1} , and raw spectra 10,722.9–12,489.48 cm^{-1} . Similarly, for 3 range method, if the combination set is 4, 4, and 4 (i.e., sd_1 , sd_1 , and sd_1), it signifies that the spectral spectrum is segmented into three segments

and subjected to preprocessing using the sd_1 for the segments $3594.87\text{--}5492.59\text{ cm}^{-1}$, $7498.31\text{--}5500.30\text{ cm}^{-1}$, and $7506.02\text{--}12,489.48\text{ cm}^{-1}$ [15]. Based on this combination set of preprocessing techniques, PLSR model were developed. This approach aims to enhance model performance by incorporating various preprocessing techniques and their combinations across different wavelength regions. By leveraging this diverse preprocessing approach, the MP-PLSR model strives to create a more robust model. GA and SPA are two different optimization techniques that can effectively capture important wavenumbers while avoiding collinearity issues. GA-PLSR leverages the strengths of both GAs (ensuring maximum fitness) and PLSR (maximizing covariance between absorbance values and the target of interest) [26, 27]. On the other hand, SPA-PLSR carefully selects feature variables to mitigate redundancy and address collinearity issues. These optimization techniques are employed to select the important wavenumbers that are then used to develop a PLSR model [15].

Figure 4.2 shows the average raw spectrum of fast-growing trees and agricultural residues of Nepal's ground biomass logged from FT-NIRS within the range of $12489.48\text{--}3594.87\text{ cm}^{-1}$.

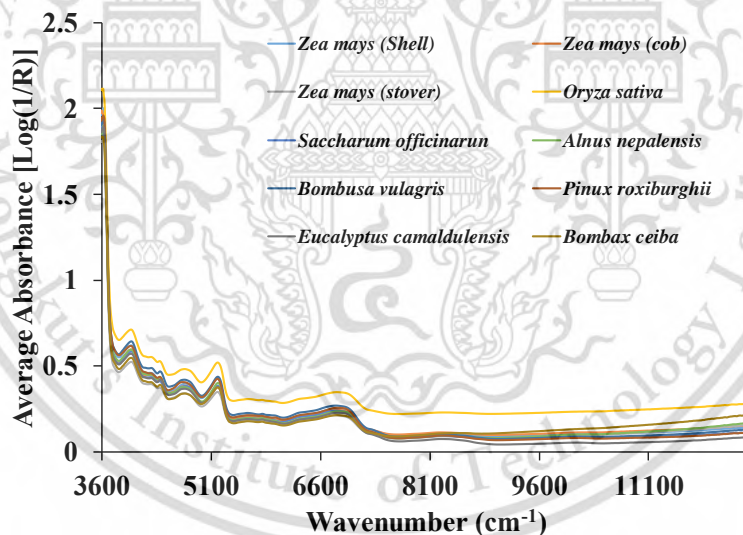


Figure 4.2 The average raw spectra of the fast-growing trees and agricultural residues of Nepal grounded biomass spanning the range of $12489.48\text{--}3594.87\text{ cm}^{-1}$ [15].

The accuracy of each model was assessed with the following specific parameters: 1) R^2_C , 2) R^2_P , 3) RMSEC, 4) RMSEP, 5) RPD, and 6) bias. These parameters were employed to evaluated and compare the models' performance. There are several factors that need to be considered to optimize the model's performance. All the instruments used for measuring the reference data should be regularly calibrated to minimize instrumental errors. The laboratory environment,

mainly the temperature, should be kept constant throughout the NIR scanning of the biomass. Care should be taken to prevent leakage of NIR radiation as it leads to data loss. The sample should be as homogeneous as possible, and contamination during sample preparation should be avoided to reduce analysis errors as well. Additionally, it is important to properly identify outliers in the data to ensure accurate calibration. Outliers can significantly impact the performance of the model and should be carefully addressed and managed during data analysis.

4.2.4. Limit of detection and limit of quantification

The LOD refers to the lowest concentration of an analyte in a test sample set that can be reliably distinguished from background noise, but not necessarily quantified accurately. Similarly, the LOQ represents the lowest concentration of the analyte that can be both reliably detected and quantified with an acceptable level of accuracy and precision [28]. In this study, LOD and LOQ are calculated based on the standard deviation of the response to slope [29], as follows:

$$\text{LOD} = 3.3 \frac{\sigma_C}{S_C} \quad \text{----- (ii)}$$

$$\text{LOQ} = 10 \frac{\sigma_C}{S_C} \quad \text{----- (iii)}$$

Where, σ_C is the residual standard deviation i.e., the error obtained from measured and predicted values of the training set and S_C is the slope of the regression line [30].

4.3. Results and Discussion

In our previous study, we developed five different types of PLSR models for ground biomass and compared their performance to select the best model for a comprehensive assessment of HHV and ultimate analysis parameters [15]. In this study, LOD and LOQ values were calculated from the best performance model, based on the standard deviation of the response to slope from the training set.

Table 4.1 displays the analysis results of the reference parameters of the grounded biomass [15], providing an assessment of the LOD and LOQ for each parameter. The LOD and LOQ values for HHV from GA-PLSR were calculated as 622.42 kJ/kg and 1886.13 kJ/kg, respectively. Additionally, LOD and LOQ for ultimate analysis parameters, including C from GA-PLSR, O from the MP-PLSR- 5 range method, H from GA-PLSR, and N from the MP-PLSR-5 range method were

calculated as follows: 3.24 wt.% and 9.81 wt.% for C, 2.04 wt.% and 6.18 wt.% for O, 0.35 wt.% and 1.05 wt.% for H, and 0.22 wt.% and 0.68 wt.% for N. These values indicate the lowest concentration at which the reference parameters of the biomass can be reliably detected and precisely quantified. It is evident that the LOD and LOQ for all reference parameters, except the wt.% of N, are lower than the minimum value used for developing the models. This suggests that the selected model i.e., GA-PLSR for HHV, C, and H and the MP-PLSR- 5 range method for O has the potential to reliably detect and precisely quantify these parameters based on their corresponding LOD and LOQ values, indicating high sensitivity. However, the model predicting the N content, i.e., the MP-PLSR-5 range method in ground biomass, has limitations due to higher LOD (0.22) and LOQ (0.68) values compared to the minimum reference values used for model development. Hence, to ensure a reliable assessment of wt.% of N, it is essential to incorporate an adequate number of representative samples with predicted wt.% of N values surpassing the LOD and LOQ values. Additionally, considering alternative modeling methods for evaluating N content in the ground biomass would enhance the model sensitivity.

The S_c of the regression line from the training set was found to be 1 for all the measuring parameters. This suggests the accuracy of the prediction of those constituents. A S_c value close to one indicate high accuracy in prediction, while a low σ_c of residual indicate the high precision of the model. A lower value of σ_c and a higher value of S_c result in the smaller values of LOD and LOQ, indicating higher sensitivity of the model.

The coefficient of determination values obtained from the analysis demonstrate a strong correlation between the independent variables and their respective parameters. Specifically, GA-PLSR yielded an R^2_c of 0.95 for HHV, 0.79 for C, and 0.88 for H in the training set. The 5 range-PLSR model resulted in an R^2_c of 0.89 for O and 0.86 for N in the training set. These findings highlight the high degree of relationships between the variables and their corresponding parameters.

Figure 4.3 illustrates the comparison between the measured versus predicted values in the training set for various parameters: a) HHV (kJ/kg), b) wt.% of C, c) wt.% of O, d) wt.% of H, and e) wt.% of N.

Table 4.1 Analysis result of the HHV and ultimate analysis parameters of the ground biomass for assessment of LOD and LOQ

Reference value	HHV	Carbon	Oxygen	Hydrogen	Nitrogen
Unit	kJ/kg	wt.%	wt.%	wt.%	wt.%
Reference value range	14,682 - 18,616	38.4 - 48.0	46.26 - 54.36	4.95 - 6.48	0 - 0.83
Algorithm	GA-PLSR	GA-PLSR	MP-PLSR-5 range	GA-PLSR	MP-PLSR-5 range
Preprocessing technique	sd ₁	sd ₁	5, 0, 1, 0, 1	SNV	4, 4, 4
N _c	157	87	77	74	76
Mean	17005	44.62	44.95	5.72	0.30
SD	842	2.00	2.55	0.30	0.22
LVs	14	9	14	14	10
R ² _C	0.95	0.79	0.67	0.88	0.86
R ² _P	0.96	0.72	0.63	0.77	0.84
RMSEC	188.01	0.98	1.45	0.10	0.07
RMSEP	170.33	0.97	1.53	0.14	0.10
RPD	4.89	1.93	1.71	2.14	2.65
bias	-21.96	0.19	0.45	-0.0	-0.03
σ _C	188.61	0.98	0.62	0.10	0.07
S _C	1.00	1.00	1.00	1.00	1.00
LOD	622.42	3.24	2.04	0.35	0.22
LOQ	1886.13	9.81	6.18	1.05	0.68

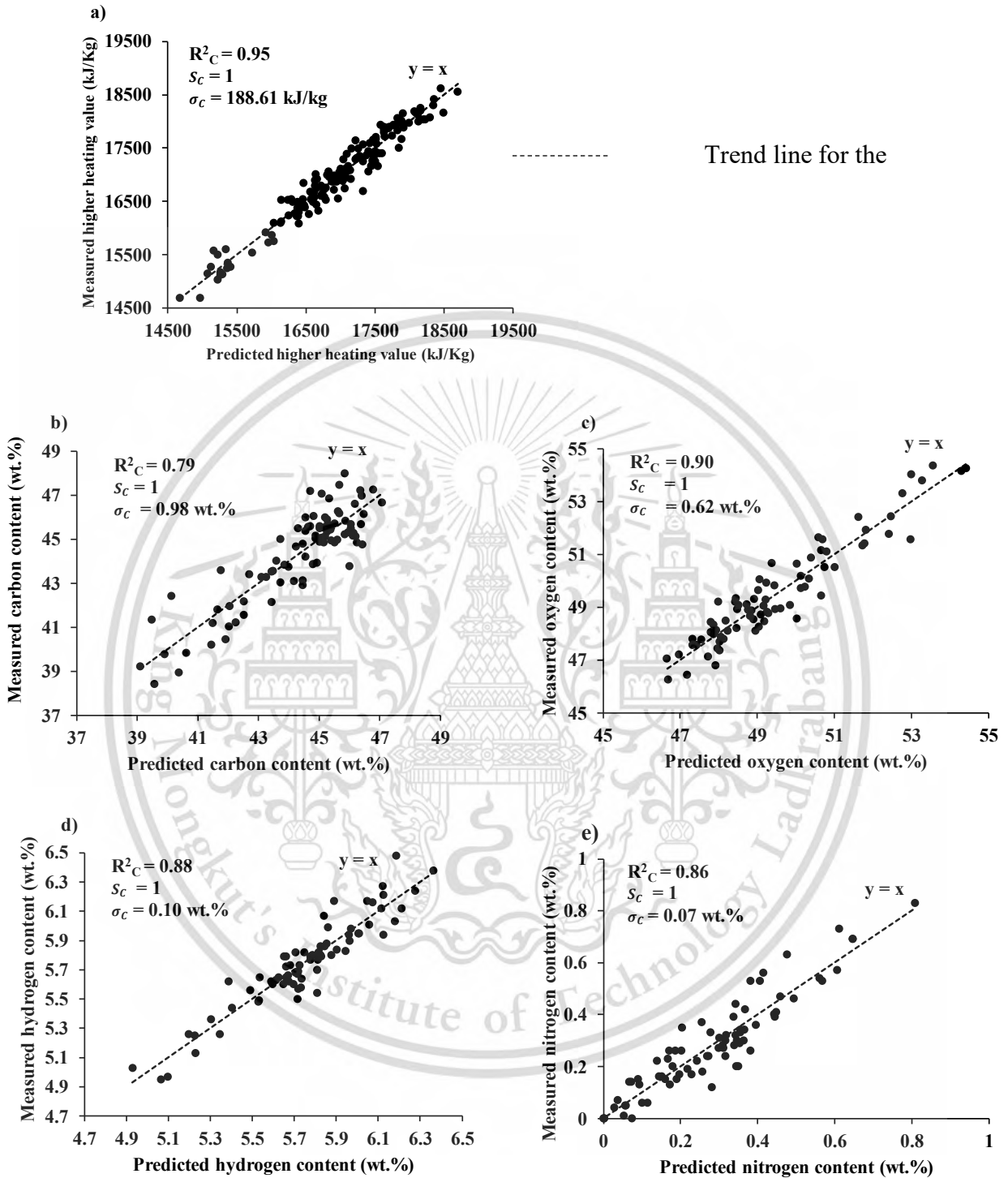
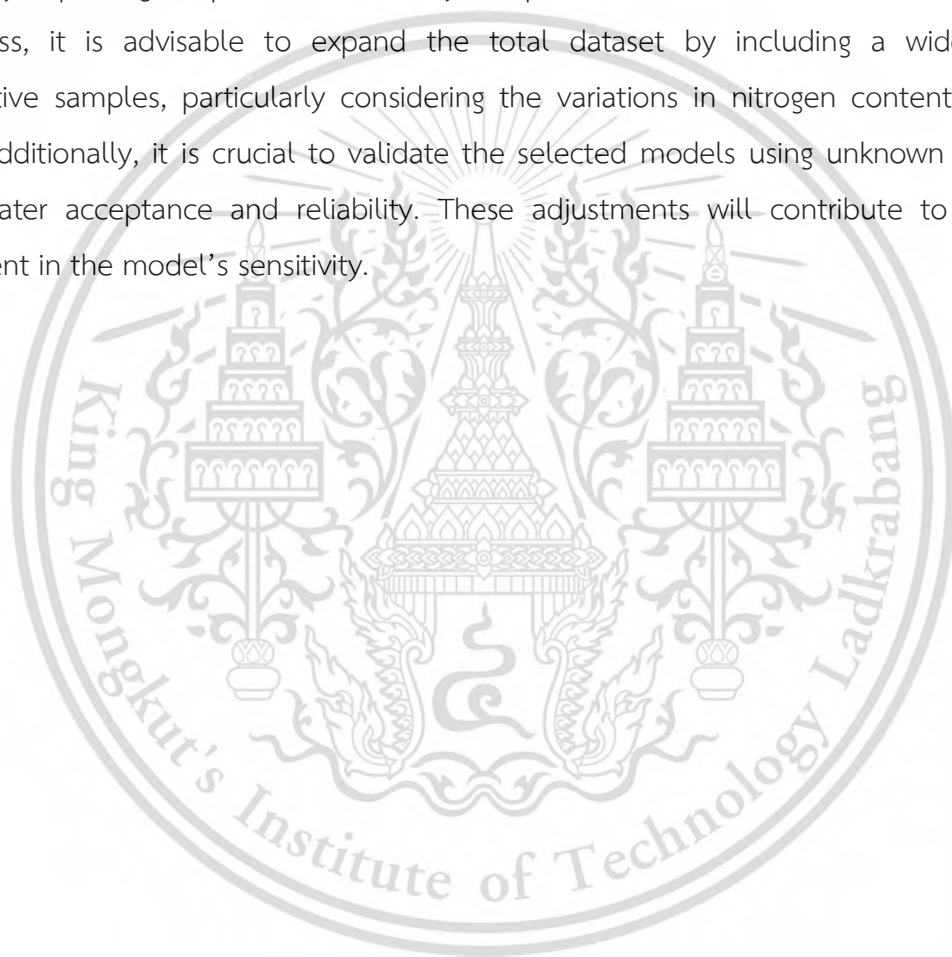


Figure 4.3 Scatter plot of measured vs predicted values from the training set for: a) HHV, b) wt.% of carbon, c) wt.% of oxygen, d) wt.% of hydrogen, and e) wt.% of nitrogen.

4.4. Conclusions

The results of this experiment show that the model has high sensitivity in detecting and quantifying parameters such as HHV in kJ/kg, wt.% of C, H, and O in grounded biomass. These results also establish a minimum threshold for detectability and quantifiability of these reference parameters. However, the LOD and LOQ for wt.% of N are higher than the minimum range value used during model development, indicating the necessity for modification to enhance its sensitivity by improving the prediction accuracy and precision. Therefore, to enhance the model's effectiveness, it is advisable to expand the total dataset by including a wide range of representative samples, particularly considering the variations in nitrogen content within the biomass. Additionally, it is crucial to validate the selected models using unknown samples to ensure greater acceptance and reliability. These adjustments will contribute to an overall improvement in the model's sensitivity.



References

1. EIA Biomass explained. <https://www.eia.gov/energyexplained/biomass/> (2021),
2. Rahman, A.; Farrok, O.; Haque, M. M., Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic. *Renewable and Sustainable Energy Reviews* **2022**, 161, 112279.
3. Srivastava, R. K.; Shetti, N. P.; Reddy, K. R.; Nadagouda, M. N.; Badawi, M.; Bonilla-Petriciolet, A.; Aminabhavi, T. M., Valorization of biowastes for clean energy production, environmental depollution and soil fertility. *Journal of Environmental Management* **2023**, 332, 117410.
4. Deora, P. S.; Verma, Y.; Muhal, R. A.; Goswami, C.; Singh, T., Biofuels: An alternative to conventional fuel and energy source. *Materials Today: Proceedings* **2022**, 48, 1178-1184.
5. Xing, J.; Luo, K.; Wang, H.; Gao, Z.; Fan, J., A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy* **2019**, 188, 116077.
6. Bakker, R. R.; Elbersen, H. In *Managing ash content and quality in herbaceous biomass: an analysis from plant to product*, 14th European Biomass Conference, 2005; 2005; p 21.
7. Obi, O. F.; Pecenka, R.; Clifford, M. J., A review of biomass briquette binders and quality parameters. *Energies* **2022**, 15, (7), 2426.
8. Nieto, P. J. G.; García-Gonzalo, E.; Paredes-Sánchez, B. M.; Paredes-Sánchez, J. P., Forecast of the higher heating value based on proximate analysis by using support vector machines and multilayer perceptron in bioenergy resources. *Fuel* **2022**, 317, 122824.
9. Seow, Y. X.; Tan, Y. H.; Mubarak, N.; Kandedo, J.; Khalid, M.; Ibrahim, M. L.; Ghasemi, M., A review on biochar production from different biomass wastes by recent carbonization technologies and its sustainable applications. *Journal of Environmental Chemical Engineering* **2022**, 10, (1), 107017.
10. Żukowski, W.; Jankowski, D.; Wrona, J.; Berkowicz-Plątek, G., Combustion behavior and pollutant emission characteristics of polymers and biomass in a bubbling fluidized bed reactor. *Energy* **2023**, 263, 125953.
11. Jara-Cobos, L.; Abril-González, M.; Pinos-Vélez, V., Production of Hydrogen from Lignocellulosic Biomass: A Review of Technologies. *Catalysts* **2023**, 13, (4), 766.

12. Ozgen, S.; Cernuschi, S.; Caserini, S., An overview of nitrogen oxides emissions from biomass combustion for domestic heat production. *Renewable and Sustainable Energy Reviews* **2021**, 135, 110113.
13. Aleksiejczuk, A.; Teleszewski, T. J., Estimation of Sulfur Dioxide Emissions in an Automatic Boiler with a Retort Burner for Coal and Biomass in a Single-Family House Based on the Measurement of the Heat Consumed. *Environmental Sciences Proceedings* **2022**, 18, (1), 10.
14. Posom, J.; Maraphum, K.; Phuphaphud, A., Rapid evaluation of biomass properties used for energy purposes using near-infrared spectroscopy. In *Renewable Energy-Technologies and Applications*, 2020; p 181:197.
15. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B. P., Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* **2023**, 16, (14), 5351.
16. Ytsma, C. R.; Dyar, M. D., Calculations of and effects on quantitative limits for multivariate analyses of geological materials with laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2022**, 191, 106395. <https://doi.org/10.1016/j.sab.2022.106395>.
17. Ershadi, S.; Shayanfar, A., Are LOD and LOQ Reliable Parameters for Sensitivity Evaluation of Spectroscopic Methods? *Journal of AOAC INTERNATIONAL* **2019**, 101, (4), 1212-1213. 10.5740/jaoacint.17-0363.
18. Shabir, G. A.; John Lough, W.; Arain, S. A.; Bradshaw, T. K., Evaluation and application of best practice in analytical method validation. *Journal of liquid chromatography & related technologies* **2007**, 30, (3), 311-333.
19. 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.
20. 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, 786-800.

21. Mishra, P.; Nikzad-Langerodi, R., Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit. *Infrared Physics & Technology* **2020**, 111, 103547.
22. Killner, M. H.; Rohwedder, J. J.; Pasquini, C., A PLS regression model using NIR spectroscopy for on-line monitoring of the biodiesel production reaction. *Fuel* **2011**, 90, (11), 3268-3273.
23. Mehmood, T.; Martens, H.; Sæbø, S.; Warringer, J.; Snipen, L., A Partial Least Squares based algorithm for parsimonious variable selection. *Algorithms for Molecular Biology* **2011**, 6-27.
24. Posom, J.; Phuphaphud, 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>.
25. 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>.
26. Maraphum, K.; Saengprachatanarug, K.; Wongpichet, S.; Phuphaphud, 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>.
27. Saenphon, C.; Ditcharoen, S.; Malai, C.; Saengprachatanarug, K.; Wongpichet, S.; Sirisomboon, P.; Saechua, W.; Khurnpoon, L.; Phuphaphud, A.; Maraphum, K., Total soluble solids, dry matter content prediction and maturity stage classification of durian fruit using long-wavelength NIR reflectance. *Journal of Food Composition and Analysis* **2023**, 105667.
28. Armbruster, D. A.; Pry, T., Limit of blank, limit of detection and limit of quantitation. *The clinical biochemist reviews* **2008**, 29, (Suppl 1), S49.
29. Guideline, I. H. T., Validation of analytical procedures: text and methodology. *Q2 (R1)* **2005**, 1, (20), 05.

30. Clua-Palau, G.; Jo, E.; Nikolic, S.; Coello, J.; Maspoch, S., Finding a reliable limit of detection in the NIR determination of residual moisture in a freeze-dried drug product. *Journal of Pharmaceutical and Biomedical Analysis* **2020**, *183*, 113163.



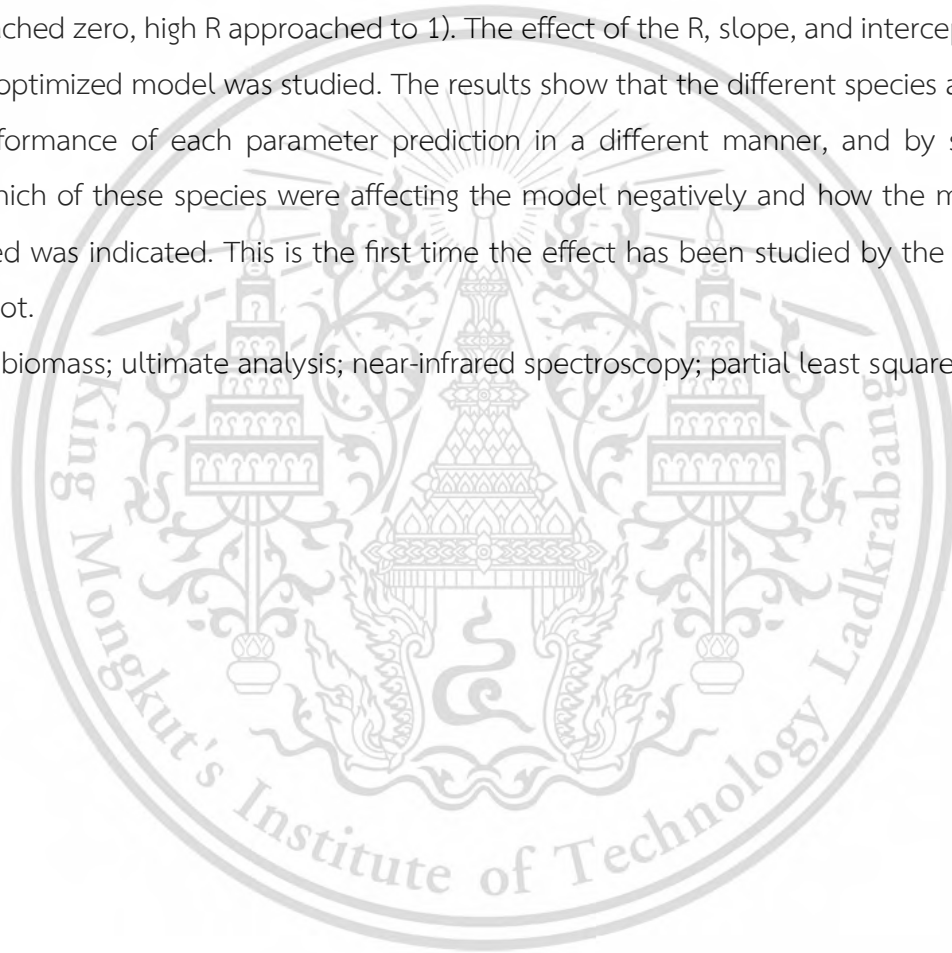
Chapter 5

Effect of combined non-wood and wood spectra of biomass chips on rapid prediction of ultimate analysis parameters using near infrared spectroscopy

The ultimate analysis parameters, including carbon (C), hydrogen (H), nitrogen (N), and oxygen (O) content in biomass, were rarely found to be predicted by non-destructive tests to date. In this research, we developed partial least squares regression (PLSR) models to predict the ultimate analysis parameters of chip biomass using near-infrared (NIR) raw spectra of non-wood and wood samples from fast-growing tree and agricultural residue and nine different traditional spectral preprocessing techniques. These techniques include first derivative (sd1), second derivative (sd2), constant offset, standard normal variate (SNV), multiplicative scatter correction (MSC), vector normalization, min-max normalization, mean centering, sd1 + vector normalization, and sd1 + MSC. Additionally, we employed a genetic algorithm (GA), successive projection algorithm (SPA), multi-preprocessing (MP) 5-range, and MP 3-range to develop a PLSR model for rapid prediction. A dataset consisting of 120 chip biomass samples was utilized for model development in which the samples were non-wood samples of 65–67% and wood samples of 33–35%, and the model performance was evaluated and compared. The selection of the optimum performing model was mainly based on criteria such as the coefficient of determination in the prediction set (R^2_p), root mean square error of the prediction set (RMSEP), and the ratio of prediction to deviation (RPD). The optimal model for weight percentage (wt.%) of C was obtained using GA-PLSR, yielding R^2_p , RMSEP, and RPD values of 0.6954, 1.1252 wt.%, and 1.8, respectively. Similarly, for wt.% of O, the most effective model was obtained using the multi-preprocessing PLSR-5 range method with R^2_p of 0.7150, RMSEP of 1.3088 wt.%, and RPD of 1.9. For wt.% of N, the optimal model was obtained using the MP PLSR-3 range method, resulting in R^2_p , RMSEP, and RPD values of 0.6073, 0.1008 wt.%, and 1.6, respectively. However, wt.% of the H model provided R^2_p , RMSEP, and RPD values of 0.5162, 0.2322 wt.%, and 1.5, respectively. Notably, the limit of quantification (LOQ) values for C, H, and O were lower than the minimum reference values used during model development, indicating a high level of sensitivity. However, the LOQ for N exceeded the minimum reference

value, implying the samples to be predicted by the model must be in the range of reference range in the calibration set. By scatter plot analysis, the effect of combined non-wood and wood spectra of biomass chips on rapid prediction of ultimate analysis parameters using NIR spectroscopy was investigated. To include different species in a model, the species have to be not only in the different values of the constituents to make a wider range for a robust model, but also must provide their trend line characteristics in the scatter plot, i.e., correlation coefficient (R), slope, and intercept (same slope and slope approached to 1, and intercept is same (no gap) and approached zero, high R approached to 1). The effect of the R, slope, and intercept to obtain the better-optimized model was studied. The results show that the different species affected the model performance of each parameter prediction in a different manner, and by scatter plot analysis, which of these species were affecting the model negatively and how the model could be improved was indicated. This is the first time the effect has been studied by the principle of a scatter plot.

Keywords: biomass; ultimate analysis; near-infrared spectroscopy; partial least squares regression



*This chapter constituted the publication article: Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Funke, A. Effect of Combined Non-Wood and Wood Spectra of Biomass Chips on Rapid Prediction of Ultimate Analysis Parameters Using Near Infrared Spectroscopy. *Energies* 2024, 17, 439. <https://doi.org/10.3390/en17020439>

Abbreviations

%	percentage	R	Correlation coefficient
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
LVs	latent variable number	RMSEP	root mean square error of prediction set
LOQ	Limit of quantification	RPD	ratio of prediction to deviation
Max	maximum	S	sulfur
Min	minimum	SD	standard deviation
MP	multi-preprocessing	sd1	first derivative
MSC	multiplicative scatter correction	sd2	second derivative
N	nitrogen	SEC	standard error of calibration set
NT	total number of samples	SEP	standard error of validation set
N _c	number of samples in calibration set	SNV	standard normal variate
NIRS	near infrared spectroscopy	SPA	successive projection algorithm
N _p	number of samples in validation set	SW	selected wavenumber
O	oxygen	TGA	thermogravimetric analysis
PLSR	partial least squares regression	wt.%	weight percentage

5.1 Introduction

The world is undergoing a significant transition away from fossil fuels, embracing modern renewable energy technologies to meet its escalating energy needs and demands. Bioenergy, derived from sources such as woody biomass, agricultural residues, and organic materials and waste, is pivotal in this paradigm shift, constituting the largest share (two-thirds) of global renewable energy utilization [1]. It is anticipated that bioenergy will continue to have a decisive share in future net zero emission scenarios and that its contribution to energy supply will further increase. This transition underscores the growing significance of biomass energy within the global energy landscape. However, it is worth noting that billions of people still rely on the inefficient use of traditional biomass for cooking and heating [1]. The combustion of biomass produces air pollutants similar to those emitted by fossil fuels, with the exception of sulfur oxides [2]. Furthermore, research has shown that the health impacts attributed to emissions from biomass and wood combustion can be more harmful than those from fossil fuels [3]. These emissions primarily result from incomplete biomass combustion and the release of solid particulate matter.

The adoption of woody biomass and non-wood biomass, such as agricultural residues, coupled with efficient combustion energy technologies, holds the potential to substantially reduce harmful emissions into the atmosphere while increasing its contribution to energy supply, making it a viable alternative to fossil fuels. Due to efficiency increase as compared to traditional biomass use, it is an important cornerstone of future scenarios. Despite significant investments in the research and development of biomass energy technologies, a knowledge gap persists, particularly concerning efficient, low-cost determination of biomass properties, including its elemental compositions (carbon (C), hydrogen (H), nitrogen (N), oxygen (O), sulfur (S), and others). During inefficient and incomplete combustion, harmful pollutants such as carbon monoxide, sulfur oxides (SO_x), nitrogen oxides (NO_x), along with particulate matter (PM_{2.5} and PM₁₀) are continuously released into the environment as smoke, posing significant health risks through indoor and outdoor exposure, with women and children being the most vulnerable [4–6].

The elemental composition of biomass has a profound impact on combustion efficiency and the emission levels released into the environment. These emissions, in turn, carry significant consequences for both the energy industry and the natural surroundings. Energy release during biomass combustion correlates positively with carbon and hydrogen contents, as they are the

primary contributors to its energy value [7]. High carbon content is desirable for energy production [8], and hydrogen's high energy content makes it valuable [9]. During combustion, oxygen reacts with carbon and hydrogen, reducing the available energy in biomass. Elevated oxygen and nitrogen contents decrease the calorific value, thereby reducing energy potential [10]. Nitrogen and sulfur are undesirable elements in biomass due to their contribution to the formation of harmful NO_x and sulfur dioxide [11,12]. To minimize environmental impact and ensure sustainable operation and maintenance of combustion systems, low sulfur content in biomass is preferred [12]. Hence, it is crucial to rapidly, accurately, and non-invasively assess the elemental composition of biomass, including C, N, O, H, and S. This assessment is essential for understanding biomass elemental composition and the potential emissions risks during energy production.

In our previous research [13], an investigation was conducted into the application of NIR spectroscopy (NIRS) for the comprehensive analysis of the ultimate analysis parameters of ground biomass intended for energy utilization. The study concludes that NIRS offers a reliable and non-destructive alternative method for rapidly assessing the elemental composition of ground biomass for energy-related purposes. Despite the valuable findings from previous research, these findings primarily served academic and research institutions. However, biomass normally is made into pellet form for export and to increase energy density where grinding is necessary before making pellets. Woodchips are especially useful, as they are easy to use, and sometimes, ground wood is not suitable for power operations due to the high cost and length of time necessary for sample preparation; therefore, it is a popular source of energy for power plants because of low preparation costs. Meanwhile, woodchip quality could be more effectively examined to achieve higher levels of plant efficiency [14]. Hence, this study aims to improve the applicability of NIR spectroscopy to assess the ultimate analysis parameters of chipped biomass, i.e., biomass with particle sizes commonly found in industrial applications. In consequence, this research outcome may directly benefit traders and energy companies, facilitating the utilization of research outcomes without the need for extensive biomass preparation such as grinding.

The data structure of samples used for model development in this present work was in two forms, i.e., non-wood and wood samples. As reported, the non-wood and wood species were different in their lignocellulosic constituents. Non-wood material of agricultural waste compost of lignin, holocellulose, α -cellulose, pentosan, and ash [15]. For example, agricultural residues, such

as hemp and sugarcane bagasse, contained higher concentrations of cellulose and lower levels of recalcitrant lignin when compared to the average woody biomass [16,17]. However, Hawanis et al. [18] reported that non-wood contained lower cellulose and lignin while wood contained higher [19–21]. Therefore, incorporating a wider range of ultimate analysis parameters (C, H, N, O, and S) as reference values will enhance the model robustness for prediction. Previous studies have strongly correlated ultimate analysis parameters to higher heating values in biomass [22]. Hence, by predicting the ultimate analysis parameters and leveraging these correlations, the fuel heating value can be characterized. This study specifically investigated the effect of combined non-wood and wood spectra from biomass chips on rapidly predicting ultimate analysis parameters using NIR spectroscopy (NIRS).

The volume of available published studies is limited in which wood and non-wood biomasses are characterized concurrently. Generally, only one specific species of biomass was used for prediction modeling, and the determination of ultimate analysis constituents by NIRS was rarely reported. Only two reports were found, including Posom and Sirisomboon [23], who optimized the PLS models using NIR spectra of 80 bamboo chip samples for evaluation of C, H, N, S, and O content. The models showed the coefficient of determination of prediction set (R^2_p) and the ratio of prediction to deviation (RPD) of 0.803 and 2.31 for C; 0.856 and 2.65 for H; 0.973 and 6.6 for N; 0.785 and 2.19 for S; and 0.522 and 1.46 for O, respectively. Similarly, the models developed by Zhang et al. [24] used 100 accessions of sorghum biomass with R^2_p 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.

There were two reports found in the available database that developed a model for two similar species to evaluate ultimate analysis parameters, C, H, N, O, and S. A total of 222 rice straw and wheat straw, collected from 24 provinces of China, were used for NIRS calibration and validation in this study where R^2_p and standard error of predictions (SEP) of independent validation were, respectively, 0.97 and 0.37% for C, 0.77 and 0.17% for H, and 0.87 and 0.10% for N [25]. Saha et al. [26] developed models by using 276 wood chip ground samples of pine trees of two species (Loblolly (*Pinus taeda*) and slash (*Pinus elliottii*)), where the biomass spectra ranged from 400 to 2498 nm at 2 nm intervals. The samples were a mix of bark, branch, needle, wood, or whole tree biomass. The prediction results show for C (sample number (n) = 43; coefficient of R^2_p

= 0.90; RPD = 3.14; ratio of prediction to interquartile (RPIQ) = 3.23); for N (n = 44; $R^2_p = 0.95$; RPD = 4.33; RPIQ = 5.96); and for S (n = 42; $R^2_p = 0.93$; RPD = 3.67; RPIQ = 3.24).

There were two reports of our group contributed to the research results of NIR prediction models for ultimate analysis parameters of the non-wood and wood samples, including Pitak et al. [27] who developed the PLS regression using the spectra obtained by line-scan NIR hyperspectral imager in which the most effective model for the prediction of C, H, and N content of 160 non-wood and wood biomass pellets including filter cake (15 pellets), *Leucaena leucocephala* (10 pellets), bamboo (15 pellets), cassava rhizome (15 pellets), bagasse (15 pellets), sugarcane leaves (15 pellets), straw (15 pellets), rice husk (15 pellets), eucalyptus bark (15 pellets), napier grass (15 pellets), and corn cob (15 pellets) developed using iGA wavelength selection and standard normal variate (SNV) spectral pretreatment and provided the highest accuracy with R^2_{pp} and SEP of 0.83 and 1.33% for C; 0.84 and 0.17% for H and 0.90 and 0.098% for N; respectively. The second report was contributed by Shrestha et al. [13], where the ground non-wood and wood samples spectra, which were 110 samples of agricultural residues and 90 samples of fast-growing trees, were used to develop the PLSR models combined with multi-preprocessing methods for ultimate analysis showed R^2_p and RPD for C of 0.7217 and 1.9, for N of 0.8410 and 2.7, for H of 0.7678 and 2.1, and for O of 0.6289 and 1.7, respectively.

The main objectives of this research include:

1. Develop PLSR models using NIR raw spectra, traditional preprocessing, MP 5-range, MP 3-range, GA, and SPA for assessing chip biomass properties for energy usage by employing NIRS while the spectra of the biomass were from non-wood (agricultural residue and bamboo) and wood (fast growing trees) samples.
2. Compare the performance of the PLSR models based on R^2_C , RMSEP, R^2_p , RMSEP, RPD, and bias.
3. Study the effect of combined non-wood and wood species in model development on model performance by scatter plot analysis.
4. Select the better performing PLSR-based model for each ultimate analysis parameter, compared with the performance of the ground biomass for rapidly assessing biomass properties for energy usage.

- Determine the limit of quantification (LOQ) value of the proposed model calibration set for each ultimate analysis parameter in chip biomass.

5.2 Materials and Methods

Figure 5.1 shows the overall research methodology for rapid prediction of ultimate analysis parameters of chip biomass by NIRS using PLSR.

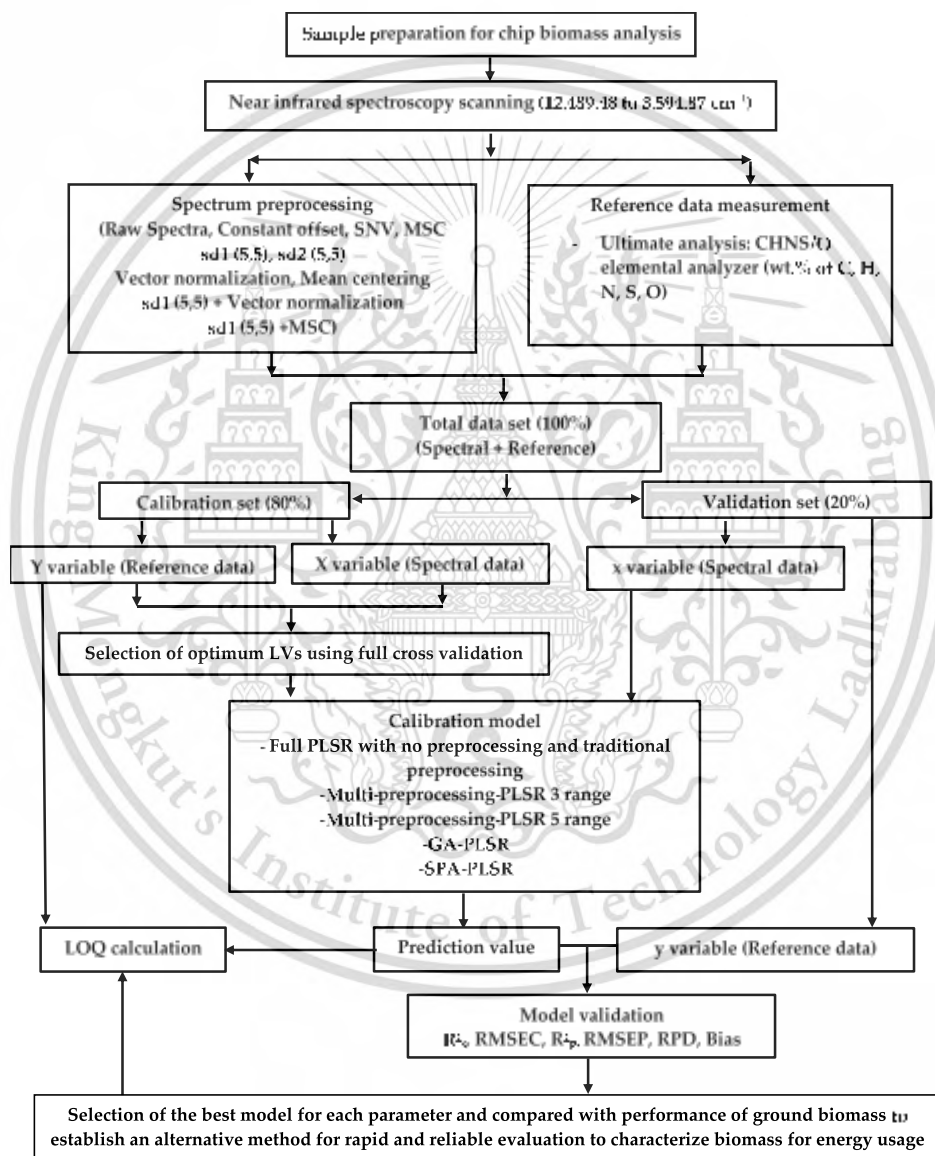


Figure 5.1 Flowchart of the overall research methodology for the rapid prediction of the ultimate analysis parameters of chip biomass for energy usage by NIRS using PLSR.

5.2.1 Sample Preparation

A total of 120 samples were collected from ten different biomass varieties, which included wood samples and non-wood samples from various geographical locations in Nepal. Wood samples included four fast-growing species: (1) *Alnus nepalensis*, (2) *Pinus roxburghii*, (3) *Bombax ceiba*, and (4) *Eucalyptus camaldulensis*. Non-wood samples were five agricultural residues: (1) *Zea mays* (cob), (2) *Zea mays* (shell), (3) *Zea mays* (stover), (4) *Oryza sativa*, and (5) *Saccharum officinarum*, and one fast-growing tree (6) *Bombusa vulagris*. The biomass samples, except *oryza sativa*, were manually chipped into smaller pieces, approximately sized 30 mm × 15 mm, for NIR scanning and for the reference measurement of ultimate analysis parameters [13].

5.2.2 Spectral Data Collection

All chip biomass samples were scanned using an FT-NIR spectrometer (MPA, Bruker, Ettlingen, Germany) in diffuse reflectance with sphere macro sample rotating mode, covering the wavelength range from 3594.87 to 12489.48 cm^{-1} , with a resolution of 16 cm^{-1} . The scanning process consisted of 32 scans (on average) for both sample and background scans to collect the raw spectra. These raw spectra were acquired in a controlled laboratory environment with air conditioning maintaining a room temperature of 25 ± 2 °C.

To compensate for the ambient influence and instrument drift on the measurement setup, background scanning was regularly performed on a gold plate as a reference for every new sample. Each biomass sample was scanned twice without changing its position, and the average of its absorbance values was calculated. All the spectra were logged as $\log(1/R)$ versus wavenumber (cm^{-1}), where R is the diffuse reflectance from the biomass sample.

Each sample was then subjected to a reference measurement of C, H, N, and S by a CHNS/O analyzer. This analyzer employs the flash dynamic combustion method, inducing complete combustion of the biomass sample within a high-temperature reactor (about 1800 °C), allowing for an accurate and precise determination of the ultimate analysis parameters.

5.2.3 Reference Analysis

The wt.% of C, H, N, and S on a dry basis in the chip biomass were determined at the Scientific and Technological Research Equipment Center (STREC) at Chulalongkorn University,

Bangkok, Thailand, using CHNS/O analyzer (Thermo Scientific TM FLASH 2000, Waltham, MA, USA) [13]. The wt.% of O on a dry basis is calculated as:

$$\text{wt.\% O} = 100 - \text{wt.\% C} - \text{wt.\% H} - \text{wt.\% N} - \text{wt.\% S} - \text{wt.\% ash} \quad (1)$$

Here, wt.% ash is determined using a thermogravimetric analyzer (TG 209 F3 Tarsus, Netzsch, Bavaria, Germany) by combusting biomass within the temperature range between 35 to 700 °C.

5.2.4 Outlier and Standard Error of Laboratory

Outliers on the reference data were identified and removed using the following equation:

$$\frac{(X_i - \bar{X})}{SD} \geq |\pm 3| \quad (2)$$

where, X_i is the measured value of sample i , \bar{X} is the average, and SD is the standard deviation of the measured values of all samples [13,28].

5.2.5. Spectral Preprocessing and Model Development

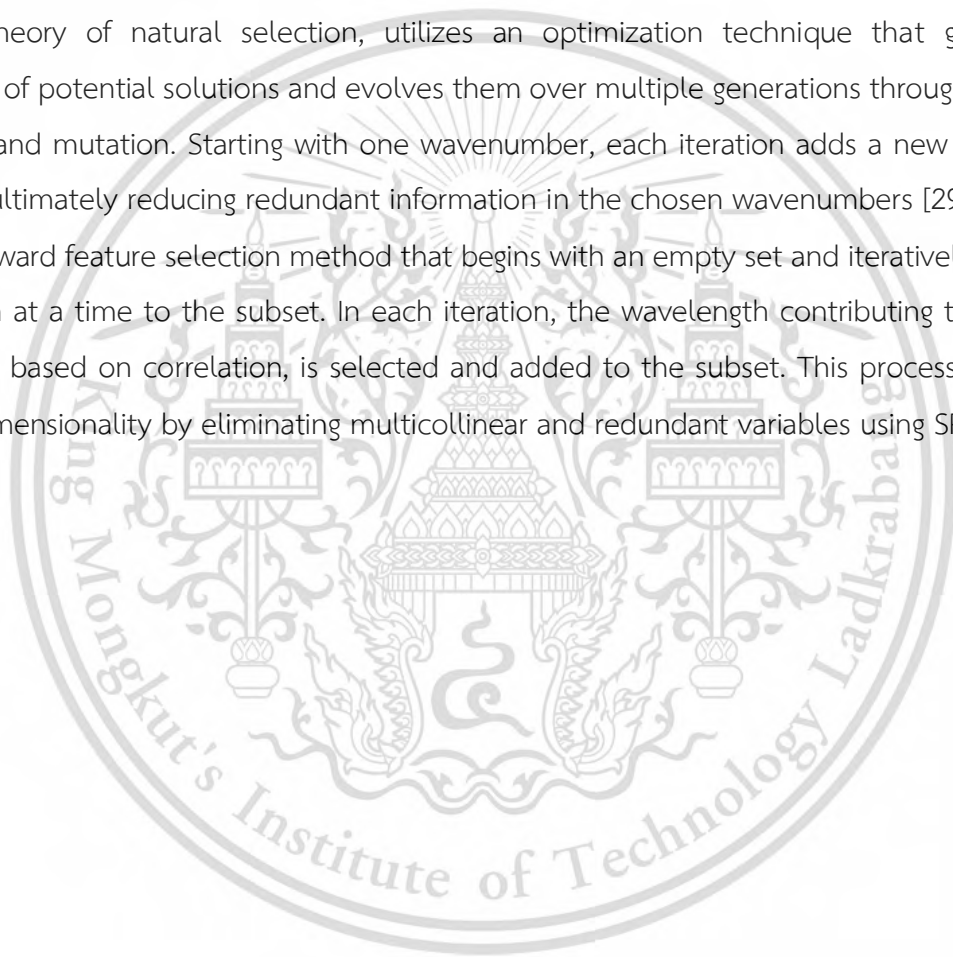
As shown in Figure 5.1, this study incorporates nine different types of spectral preprocessing applied to the raw spectra. These methods include constant offset, SNV, MSC, sd1, sd2, vector normalization, mean centering, sd1 + vector normalization, and sd1 + MSC.

Five different types of PLSR-based regression models, namely Full-PLSR, MP PLSR-5 range, MP PLSR-3 range, GA-PLSR, and SPA-PLSR, were developed to compare and select the best-performing model for each ultimate analysis parameter to establish a reliable and non-destructive alternative method for rapidly assessing biomass properties for energy usage [13].

The primary objective of the MP method is to optimize model performance by applying various preprocessing techniques to different divided sections within the entire wavenumber range. A built-in code in MATLAB R2020b was utilized to obtain a combination set of different preprocessing techniques based on the desired number of random pairs. The optimal combination set for each selected number of random pairs is determined through a cross-validation procedure using PLSR on reference and spectroscopic data. Using the selected combination set of preprocessing techniques, the PLSR model was developed. Here, we generate a combination set of preprocessing techniques using seven different options: 0 = empty (all

absorbance values = 0), 1 = raw spectra, 2 = SNV, 3 = MSC, 4 = sd1, 5 = sd2, and 6 = constant offset. In the MP approach, two methods were adopted: in the MP PLSR-5 range method, the spectral range is divided into five equal sections, while in the MP PLSR-3 range method, it is divided into three sections. The best MP combination set for model development is then determined [13].

Both GA and SPA were employed to select concise and influential wavenumbers, aiming to prevent overfitting and result in an improved prediction model [27]. GA, inspired by Charles Darwin's theory of natural selection, utilizes an optimization technique that generates a population of potential solutions and evolves them over multiple generations through selection, crossover, and mutation. Starting with one wavenumber, each iteration adds a new one to the selection, ultimately reducing redundant information in the chosen wavenumbers [29]. Similarly, SPA is a forward feature selection method that begins with an empty set and iteratively adds one wavelength at a time to the subset. In each iteration, the wavelength contributing the most to the model, based on correlation, is selected and added to the subset. This process effectively reduces dimensionality by eliminating multicollinear and redundant variables using SPA [30–32].



5.2.6 Limit of Quantification (LOQ)

Based on the SD of the response to slope method from the calibration model, LOQ, which represents the lowest concentration of the analyte that can be detected and quantified with an acceptable level of accuracy and precision [28,33], is calculated as follows:

$$\text{LOQ} = 10 \frac{\sigma_c}{S_c} \quad (3)$$

where, σ_c is the residual standard deviation, i.e., the precision obtained from measured and predicted values of the calibration set, and S_c is the slope of the model regression line.

5.3 Results and Discussion

Table 4.1 shows the number of non-wood samples and wood samples in the calibration set and validation set. The wood sample number is about 33–35% of the total sample number; hence, the non-wood sample number is 65–67%. Out of 120 samples, the number of outlier samples can be evaluated by the data in Table 5.1.

Table 5.2 presents statistical data for the ultimate analysis parameters of chip biomass obtained using CHNS/O elemental analyzer (Thermo Scientific™ FLASH 2000). This data was used in both the calibration and prediction sets for model development. S content in the chip biomass was not detected, possibly due to its very low content falling below the detection threshold. Therefore, a PLSR-based model for S content in the chip biomass was not developed in this study. The wt.% of O is calculated using Equation (1).

Table 5.1 The number of non-wood samples and wood samples in calibration set and validation set.

Parameter	Total Sample	Calibration Set			Validation Set		
		Wood	Non-Wood	Total	Wood	Non-Wood	Total
wt.% C	111	31	58	89	8	14	22
wt.% H	119	32	63	95	8	16	24
wt.% N	116	31	62	93	9	14	23
wt.% O	102	28	54	82	8	12	20

Table 5.2 The statistical data of the ultimate analysis parameters of the chip biomass obtained using CHNS/O elemental analyzer used in PLSR model development.

Parameter	N_T	Calibration Set					Validation Set				
		N_C	Max	Min	Mean	SD	N_p	Max	Min	Mean	SD
C (wt.%)	111	89	48.7500	38.9300	44.6330	2.1380	22	47.2800	49.7550	44.4439	2.0878
H (wt.%)	119	95	6.6200	4.9100	5.7620	0.3485	24	6.5700	4.9500	5.6490	0.3411
O (wt.%)	102	82	51.1200	37.3600	44.6322	2.8521	20	48.8000	38.8500	45.1159	2.5149
N (wt.%)	116	93	0.9100	0.0000	0.2987	0.2250	23	0.6200	0.0000	0.2714	0.1645

Table 5.3 shows the results of the PLSR-based model for ultimate analysis (wt.%) of chip biomass, where the bolded model shows the best performance. However, it is essential to consider the recommendation provided by Williams et al. [34], where with an R^2_p value between 0.66–0.81, the model can be used for rough screening and other suitable calibration purposes. Therefore, C, O, and N models were. For the H model, according to Williams guideline [34], a model with an R^2_p value between 0.50–0.64 is only suitable for very rough screening. Likewise, every model of biomass chips for ultimate analysis parameters was in alignment with the recommendation from Zornoza et al. [35], in which any model with an RPD value below 2 was deemed insufficient for any application.

Table 5.3 Results of the PLSR-based model for ultimate analysis (wt.%) of chip biomass, bolded model showing the best performance.

Parameter	Algorithm	Preprocessing	LVs	Calibration Set		Prediction Set			
				R ² _C	RMSEC	R ² _P	RMSEP	RPD	Bias
wt.% C	Full-PLSR	sd2 (g = 5, s = 5)	10	0.8215	0.8982	0.6489	1.2081	1.7	0.0854
	GA-PLSR	sd2 (SW: 306)	9	0.8078	0.9320	0.6954	1.1252	1.8	0.0053
	SPA-PLSR	sd2 (SW: 634)	10	0.8030	0.9435	0.6520	1.2028	1.7	0.1036
	MP-PLSR: 3 range	Combination set: 4,2,4	9	0.7132	1.1386	0.5514	1.3655	1.5	-0.1433
	MP-PLSR: 5 range	Combination set: 4,1,4,3,1	13	0.8628	0.7875	0.5467	1.3727	1.5	-0.1226
wt.% H	Full-PLSR	sd1 (g = 5, s = 5)	6	0.5086	0.2429	0.4996	0.2361	1.5	-0.0660
	GA-PLSR	Vector normalization (SW: 67)	11	0.5456	0.2336	0.5162	0.2322	1.5	-0.0781
	SPA-PLSR	sd2 (SW: 22)	15	0.5172	0.2408	0.4478	0.2481	1.4	-0.0586
	MP-PLSR: 3 range	Combination set: 5,5,0	7	0.5179	0.2406	0.4711	0.2428	1.4	-0.0644
	MP-PLSR: 5 range	Combination set: 5,4,4,0,4	8	0.5964	0.2201	0.4877	0.2389	1.4	-0.0625
wt.% O	Full-PLSR	sd2 (g = 5, s = 5)	8	0.6243	1.7376	0.6362	1.4788	1.7	0.0814
	GA-PLSR	Mean Centering (SW: 1025)	11	0.6347	1.7134	0.6064	1.5381	1.6	0.2414
	SPA-PLSR	Min-max normalization (SW:354)	11	0.5800	1.8370	0.5815	1.5860	1.6	0.3466
	MP-PLSR: 3 range	Combination set: 4,5,0	11	0.6572	1.6597	0.6153	1.5207	1.6	0.1064
	MP-PLSR: 5 range	Combination set: 2,5,2,1,5	15	0.8097	1.2366	0.7150	1.3088	1.9	0.0733
wt.% N	Full-PLSR	MSC	10	0.7232	0.1177	0.5865	0.1035	1.6	-0.0065
	GA-PLSR	SNV (SW: 39)	10	0.5916	0.1429	0.5625	0.1064	1.5	-0.0132
	SPA-PLSR	Min-max normalization (SW:413)	7	0.6396	0.1343	0.5869	0.1034	1.6	-0.0190
	MP-PLSR: 3 range	Combination set: 4,0,0	15	0.8656	0.0820	0.6073	0.1008	1.6	0.0191
	MP-PLSR: 5 range	Combination set:1,4,4,1,0	7	0.6436	0.1335	0.5700	0.1055	1.5	0.0143

5.3.1 wt.% of C

Table 5.3 presents the results of the PLSR-based model within the full wavenumber range of 3594.87–12489.48 cm⁻¹ for the wt.% C of chip biomass, with the best-performing model highlighted in bold.

The model, developed using GA-PLSR with spectrum preprocessing involving the sd2, a gap, and segments of five each, along with nine LVs, provided better results. It achieved R^2_C , RMSEC, R^2_p , RMSEP, RPD, and bias values of 0.8078, 0.9320 wt.%, 0.6954, 1.1252 wt.%, 1.8, and 0.0053 wt.%, respectively. By determining RMSEP, these results represent a 6.8566% improvement in the model performance compared to Full-PLSR. Utilizing Equation (3), the LOQ value was calculated as 9.3724 wt.% for C. Notably, the LOQ value is lower than the minimum wt.% C value used during model development, indicating that the model exhibits high sensitivity and can quantify wt.% C starting from 9.3724 wt.%.

Figure 5.2a shows a scatter plot comparing the predicted and measured wt.% of C, which was obtained using GA-PLSR. The trend line for the prediction set and calibration set overlap, indicating the same slope. The slope shows the rate of change of Y (measured value) as a function of the rate of change of X (predicted values) [34] or vice versa, hence indicating that predicted values of both sets of data have changed with the same rate and this characteristic is same for the models for O and N shown in Figure 5.2c,d.

Figure 5.3 displays the average sd2 absorbance values obtained after preprocessing, highlighting 306 selected wavenumbers (marked in red) identified through GA. These wavenumbers fall within the full spectral range of 3594.87–12489.48 cm^{-1} . Peaks were observed at 3722, 4091, 5181, and 5285 cm^{-1} , all of which might have the potential to enhance the model performance. The wavenumbers 3722 cm^{-1} and 4091 cm^{-1} are associated with the C–H aromatic functional group, specifically the C–H aryl material type [36]. The peak at 5181 cm^{-1} corresponds to a combination of O–H stretching and HOH bending, indicative of polysaccharides [36]. Similarly, the peak at 5285 cm^{-1} is associated with the functional group of O–H hydrogen bonding between water and exposed polyvinyl alcohol OH groups [36].

Previous studies by Zhang et al. [24] and Posom and Sirisomboon [23] have demonstrated that vibrational bands related to C–H aromatic, C–H stretching, N–H stretching, N–H deformation, O–H stretching, HOH bending, O–H hydrogen bonding, and similar factors play a crucial role in predicting the wt.% of C in various biomass varieties. These findings align with the vibration bands observed in our study, providing support for our results and suggesting that these selected peaks likely have a significant influence on the model performance.

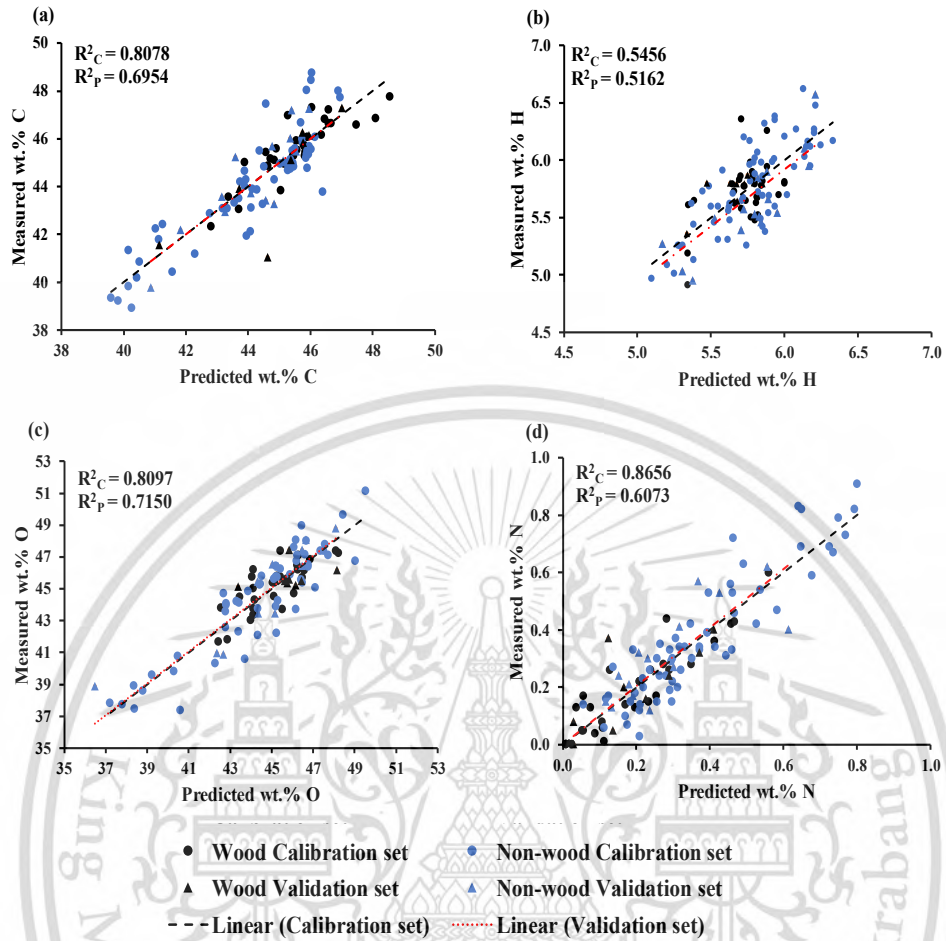


Figure 5.2 Measured versus predicted value in calibration and prediction sets for (a) wt.% of C, (b) wt.% of H, (c) wt.% of O, and (d) wt.% of N.

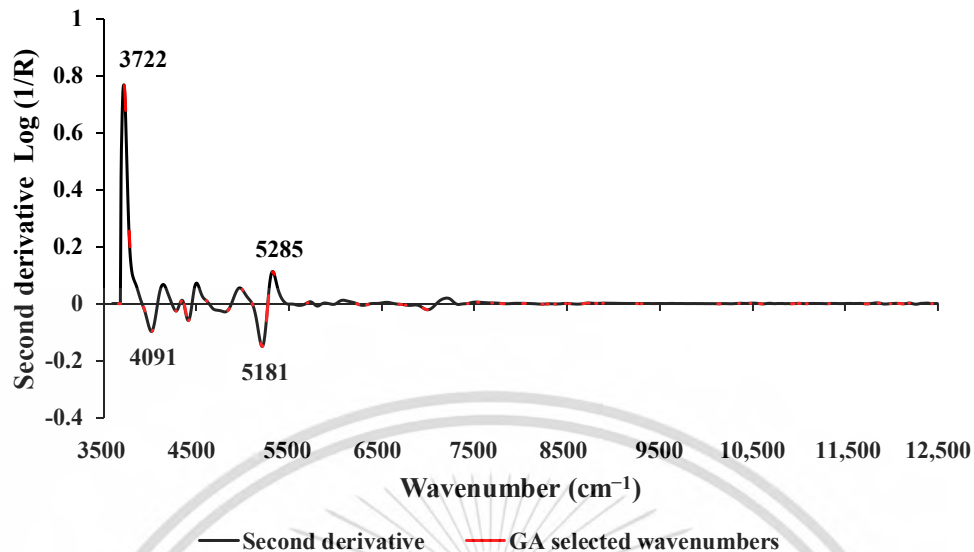


Figure 5.3 The second derivative absorbance value of studied biomass obtained using the sd2 preprocessing with a selection of important wavenumber obtained from GA for prediction of wt.% of C, within the full wavenumber range of 3594.87–12489.48 cm^{-1} .

5.3.2 wt.% of H

The model developed using GA-PLSR with vector normalization as preprocessing showed the best performance with 11 LVs (Table 5.3). It selected 67 important wavenumbers using GA. The model performance, in terms of R^2_C , RMSEC, R^2_P , RMSEP, RPD, and bias values, was 0.5456, 0.02336 wt.%, 0.5162, 0.2322 wt.%, 1.5, and -0.0781 wt.%, respectively. Compared with Full-PLSR, the GA improved the PLSR model accuracy by 1.6743%. The LOQ value was calculated as 2.3484 wt.%, which is lower than the minimum reference value used for the model development. This suggests that the selected model is sensitive and can sensitively quantify H from 2.3484 wt.%.

Figure 5.2b displays a scatter plot comparing the predicted and measured wt.% of H, which was obtained using GA-PLSR. It is clear that the trend line for the prediction set exhibits an offset in relation to the trend line of the calibration set and the 45-degree line. This offset raises concerns about the model constant bias along the range of the data, indicating the overestimating model.

Figure 5.4 displays the average absorbance values within the range of 3594.87–12489.48 cm^{-1} . These values were obtained after preprocessing using vector normalization and highlight 67 selected wavenumbers, marked in red, which were identified using GA. Significant peaks were observed at the wavenumbers 4019, 4850, 5155, and 9852 cm^{-1} , respectively, and these may

have an influence on the model performance. The peak at 4019 cm^{-1} is associated with the spectra-structure combination of C–H stretching and C–C stretching, with the material type being cellulose [36]. The peak at 4850 cm^{-1} corresponds to the functional group of N–H combination bands found in secondary amides within proteins [36]. The peak at 5155 cm^{-1} is related to the combination of O–H stretching and HOH bending, with the material type being water [36]. Finally, the peak at 9852 cm^{-1} is associated with the second overtone of the fundamental stretching band of N–H asymmetric stretching, and the material type is aromatic amine [36].

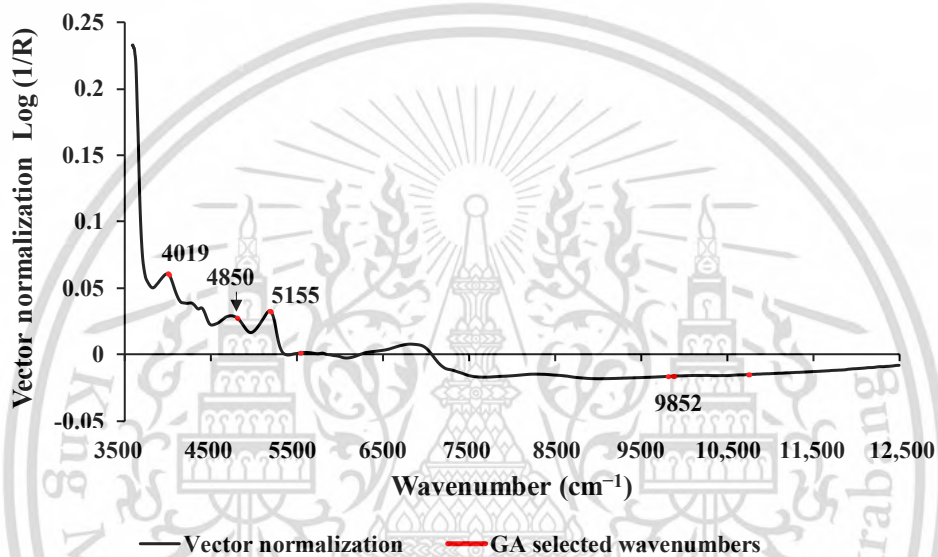


Figure 5.4 The vector normalization absorbance value of studied biomass obtained using the vector normalization preprocessing with a selection of important wavenumber obtained from GA for prediction of wt.% of H, within the full wavenumber range of $3594.87\text{--}12489.48\text{ cm}^{-1}$.

In comparison to previous studies conducted by Shrestha et al. [13], Zhang et al. [24], and Posom and Sirisomboon [23] that focused on measuring the wt.% of H in biomass using NIRS, our study discovered similar peaks within the range of $4000\text{--}9900\text{ cm}^{-1}$ and vibration bands such as O–H stretching, HOH bending, C–H stretching, and C–C stretching. Therefore, our study findings align with these earlier studies on this specific aspect. However, when evaluating the overall performance of various PLSR-based models, this study suggests that the wt.% of H was not sufficiently explained by the vibration of those mentioned bonds.

5.3.3 wt.% of O

Assuming that the S content in chip biomass is negligible, as its wt.% is too low to be detected by the instrument, we calculated the wt.% of O in the chip biomass for 120 samples using Equation (1). The wt.% of ash content for each biomass was determined using a TGA. Table 5.3 presents the optimal results from five different types of PLSR-based models. The most effective model was developed using the MP PLSR 5-range method, incorporating a spectral preprocessing combination set of 2, 5, 2, 1, and 5, which corresponded to the following ranges: 3625.72–5392.30 cm^{-1} with SNV, 5400.02–7166.59 cm^{-1} with the sd2, 7174.31–8940.89 cm^{-1} with SNV, 8948.60–10,715 cm^{-1} with raw spectra, and 10722.9–12489.48 cm^{-1} with the sd2. This model employed 15 LVs. Figure 5.2c illustrates the scatter plot comparing measured versus predicted wt.% of O obtained from the MP PLSR 5-range method. This method yielded R^2_c of 0.8097, RMSEC of 1.2366 wt.%, R^2_p of 0.7150, RMSEP of 1.3088 wt.%, RPD of 1.9, and a bias of 0.0733 wt.%. Compared with Full-PLSR method performance, the MP PLSR 5-range method significantly improved the model accuracy by 11.4913%. The LOQ value for wt.% of O was calculated as 12.4424 wt.%, which is lower than the minimum wt.% of O used during model development. This indicates that the model is highly sensitive and can quantify O content in chip biomass from 12.4424 wt.%.

Figure 5.5 displays the regression coefficient plot for wt.% of O content in chip biomass obtained from the MP PLSR 5-range method. Several notable peaks were observed at 3650, 4405, 8163, and 8621 cm^{-1} , each potentially exerting a significant influence on the model performance. Specifically, the peak at 3650 cm^{-1} corresponds to the O–H functional group found in the primary alcohols, characterized by the fundamental stretching vibrational absorption band of O–H [36]. The peak at 4405 cm^{-1} represents the combination of O–H stretching and C–O stretching, with cellulose as the material type [36]. The peaks at 8163 cm^{-1} and 8621 cm^{-1} are associated with the second overtone of the fundamental stretching band of C–H and the fourth overtone of the fundamental stretching band of C=O, respectively, which are typically found in hydrocarbons and aliphatic compounds [36].

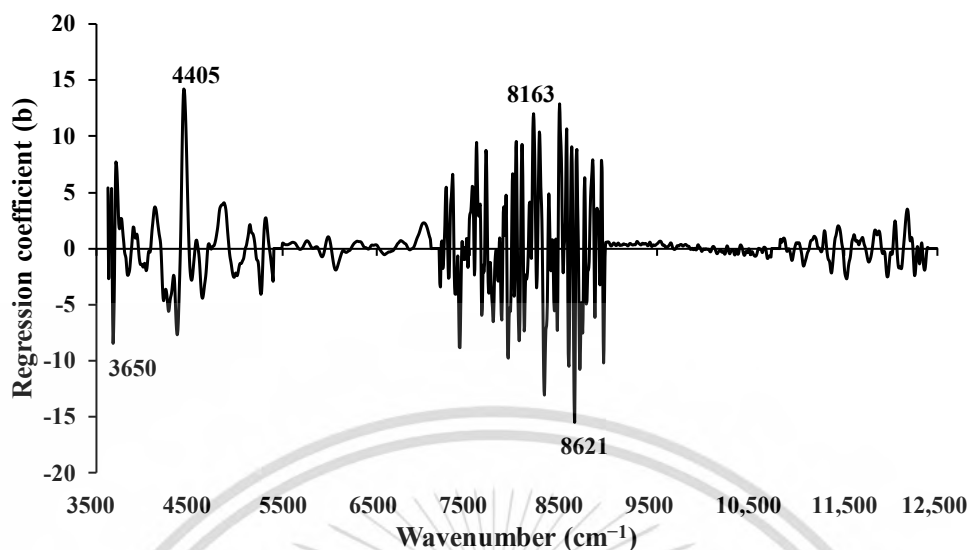


Figure 5.5 The regression coefficient for the wt.% O of chip biomass using the MP PLSR 5-range method.

When compared with previous studies on wt.% of O in biomass, such as those by Shrestha et al. [13], Zhang et al. [24], and Posom and Sirisomboon [23], this study reveals some contradictory peaks. However, the vibrational bands, such as O–H from primary alcohol, C=O stretching, and C–H stretching, among others, were similar. These findings supports the research result of this study, suggesting that the significant peaks observed in this study have an impact on the development of the model for assessing wt.% of O in chip biomass.

5.3.4 wt.% of N

The best model for rapid prediction of wt.% of N was obtained using the MP PLSR 3-range method with a spectral preprocessing combination set of 4, 0, and 0 (Table 5.3). This set corresponds to the sd1 from 3594.87 to 5492.59 cm^{-1} and zero absorbance from 7498.314 to 12489.48 cm^{-1} . Figure 5.2d illustrates the scatter plot of measured versus predicted wt.% of N content in the chip biomass, obtained from the MP PLSR 3-range method with 15 LVs. The best-performing model achieved an R^2_C of 0.8656, RMSEC of 0.0820 wt.%, R^2_P of 0.6073, RMSEP of 0.1008 wt.%, RPD of 1.6, and a bias of 0.0191 wt.%. These results indicate that within the range 3594.87–5492.59 cm^{-1} (refer Figure 5.6), by effectively correcting baseline shifts and assigning zero absorbance value within the remaining wavenumber range, the model performance is enhanced. Compared with Full-PLSR using RMSEP value, the MP PLSR 3-range method improved the model performance by 2.5473%. However, based on R^2_C and R^2_P values, the selected model indicates overfitting. This suggests that our model fits the training data too closely, and too much less accurate in prediction the validation

set. This was discussing in Section 5 Comparison of Model Performance between Using Chipped and Ground Biomass Spectra by refer to Cawley and Talbot [37].

Figure 5.6 illustrates the regression coefficient plot for the wt.% of N in chip biomass, obtained using the multi-preprocessing PLSR 3-range method. Significant peaks that could potentially influence the model performance were observed within the wavenumber range of 3594.87–5492.59 cm^{-1} only. These significant peaks were noticed at wavenumbers 3693, 4019, 4365, 4505, 4701, and 5285 cm^{-1} . Specifically, the peak at 3693 cm^{-1} is associated with the function group of C–H aromatic C–H bands, characterized by the material type C–H aryl. At 4019 cm^{-1} , the peak represents functional groups with a combination of C–H stretching and C–C stretching from cellulose [36]. The peak at 4365 cm^{-1} corresponds to CONH_2 , specifically due to C=O bonded to the N–H of the peptide link termed the α -helix structure [36]. The peak at 4505 cm^{-1} is associated with the N–H combination band [36]. Similarly, the peak at 4701 cm^{-1} corresponds to the function group of N–H/C=O combination from polyamide II [36]. Lastly, the peaks at 5285 cm^{-1} are associated with O–H hydrogen bonding between water and exposed polyvinyl alcohol OH [36]. These peaks are crucial in understanding the composition of the chip biomass and are important for model development and analysis. Furthermore, in the range of 7498.314–12,489.48 cm^{-1} , the regression coefficient value equals zero. This indicates an insufficient linear relationship between the dependent (spectral information) and independent (reference value) variables in this range, and it does not significantly contribute to the predictive model for the prediction of wt.% of N.

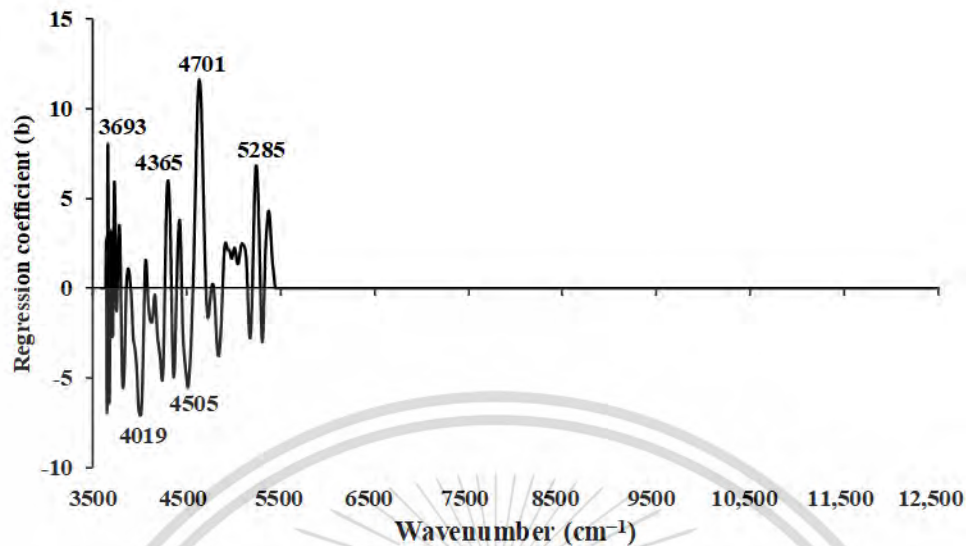


Figure 5.6 The regression coefficient for the wt.% N of chip biomass using the MP PLSR 3-range method.

The previous study conducted by Posom and Sirisomboon [23], which aimed to evaluate the wt.% of N in bamboo, also revealed significant peaks within the range of 4424 to 6920 cm^{-1} . Similarly, Shrestha et al. [13] conducted a study on wt.% of N in ground biomass from the same source and exhibited important peaks within a similar range, specifically within 4019 to 6711 cm^{-1} . This finding aligns with the results of our study, providing additional support for our research. It is noteworthy that in both studies, common vibrational bands, such as N–H stretching, C=O stretching, C–H stretching, C–C stretching, aromatic C–H, and O–H bonds between water and alcohol, among others, were identified. This consistency in vibration bonds reinforces our study findings and suggests that these specific peaks likely play a crucial role in influencing the model performance.

5.3.5 Effect of Non-Wood and Wood Samples on Model Performance

Table 5.4 shows the reference values of wt.% of C, H, N, and O of non-wood and wood samples in calibration and validation sets. From Figure 5.2 and Table 5.4, it is obvious that the range of every element content is wider after the two sets were combined for modeling. Therefore, the models can now be regarded as more robust models than only one set was used. From Figure 5.2a,c, the range of wt.% of C and O of wood samples was narrower than those of the non-wood samples which were extended more to the lower wt%. Figure 5.2d illustrates the opposite way, where the value range of N of wood samples was lower and narrower than those of the non-wood samples. Therefore, models for wt% of C, O, and N had better performance than that of the H

model. The wood sample reference values of H were grouped together and more or less had the same range as the range of non-wood samples. (Figure 2b).

Table 5.4 The range of wt.% of C, H, N, and O of non-wood and wood samples in calibration and validation sets.

Parameter	Calibration Set		Validation Set	
	Wood	Non-Wood	Wood	Non-Wood
wt.% C	47.77–42.33	48.75–39.93	47.28–41.02	47.24–39.76
wt.% H	6.36–4.91	6.62–4.97	6.57–4.95	5.87–5.36
wt.% N	0.60–0.00	0.91–0.00	0.40–0.00	0.62–0.12
wt.% O	47.40–41.68	51.12–37.36	47.43–45.14	48.80–38.85

The literature shows that the one species model of non-wood, which were bamboo wood chips [23] and sorghum [24] for evaluation of ultimate analysis parameters, C, H, N, O, and S had better performance than our combined non-wood and wood models as the results described in the introduction of this manuscript. Similarly, the two similar species of rice straw and wheat straw model [25] and the pine tree of two species (Loblolly (*Pinus taeda*) and slash (*Pinus elliottii*)) model [26] indicated better prediction performance, though they were homogeneous ground samples which might make their model performance better than the chip ones due to less scattering problem. Shrestha et al. [13] worked with ground samples of the same batch of non-wood and wood samples. Spectra from this experiment showed better R^2_p and RPD for C, N, H, and O, which is claimed to be due to the same merit of homogeneous samples.

Using larger biomass particle sizes, Pitak et al. [27] combined the non-wood and wood biomass pellet NIR spectra obtained by averaging every pixel spectrum of the pellets from a hyperspectral image (HSI). This approach provided better performance in predicting elements from the ultimate analysis than our model, i.e., in-detail data collection by the HSI leads to significant improvements.

Figure 5.7 shows the scatter plots of the highest performance models in this study in predicting the C, H, O, and N content of the wood and non-wood samples, which is the same as Figure 5.2, but the difference is Figure 5.7 shows the simple regression lines of each group of non-

wood and wood samples both for calibration set and prediction set. For better vision, Table 5.5 shows the numeric data of R^2 , slope, and intercept calculated from the scatter plots of wood and non-wood calibration and prediction sets. Williams et al. explained that the slope of the trend line plotted between Y (measured value) and X (NIR predicted value) indicated the rate of change of Y as a function of the rate of change of X [34]. The intercept of different species illustrated the same trend as slope interpretation, especially when the slope is more than 1, the intercept was with a minus sign, and if less than 1, the intercept was with a plus sign. While the slope was 1, the intercept was low, close to zero, and when the slope was more or less than 1, the intercept was high, far from zero.

The perfect relationship between the reference values and the predicted values is when the correlation coefficient (R) and slope are equal to 1 and the intercept is equal to zero [34]. From Table 5.5, for the C model, the non-wood samples contributed slightly more merit on calibration model performance than wood samples for more R the slope was closer to 1, and the intercept was closer to zero. But the prediction set of non-wood provided a steeper slope and intercepted far more from zero.

By the same way of interpretation, the model for H obtained more merit from non-wood samples, while for the wood samples, the R of the trend line was very low, the slope was far from 1, and the intercept was slightly far from zero. The incongruous trend lines of both sets makes the overall performance of the model worse as shown in Table 5.3.

For the N model, the wood and non-wood calibration set samples more or less had the same trend line characteristics, which supplement the good calibration model performance, though the prediction sample set of both biomass species trend line characteristics shows less R and slope far from 1 led to overfit calibration models of both biomass groups (Table 5.5).

For the O model, the non-wood group had better trend line characteristics and contributed good merit to the model, while the poorer trend line characteristics of the wood group made the overall model inferior but by a small portion because the number of samples in the non-wood group was much more (Table 5.5). By the strong merit of the non-wood group, the overall model performance for O prediction was fairly acceptable (Table 5.3).

Table 5.5 The trend line characteristics of the wood and non-wood species in scatter plots of the best models for C, H, N, and O.

Element	Wood						Non-Wood					
	R^2_C	R^2_P	Slope _C	Slope _P	Intercept _C	Intercept _P	R^2_C	R^2_P	Slope _C	Slope _P	Intercept _C	Intercept _P
C	0.7243	0.6456	0.8353	1.0139	7.5532	-0.8994	0.7962	0.7681	1.0243	1.2109	-1.0960	-9.1465
H	0.2683	0.5028	0.7876	0.7066	1.2085	1.7444	0.6111	0.7185	1.0342	1.1318	-0.1925	-0.9224
N	0.8335	0.5486	0.8915	0.7670	0.0197	0.0502	0.8454	0.6289	1.0368	0.8541	-0.0139	0.0708
O	0.6187	0.0992	0.8272	0.1840	7.8316	37.2740	0.8311	0.8063	1.0209	0.9519	-0.9462	2.3866

R^2_C : Coefficient of determination in the calibration set, R^2_P : Coefficient of determination in the validation set, Slope_C: Slope of trendline in the calibration set, Slope_P: Slope of trendline in the validation set, Intercept_C: Intercept in the calibration set, Intercept_P: Intercept in the validation set.

Tables 5.6–5.9 show the trend line characteristics, including R^2 , slope, and intercept of each specific plant of wood and non-wood samples used in the optimized models for evaluation of C, N, H, and O, respectively. It was observed that most of the R^2_P of every plant was equaled to 1 for the samples of those plants in the optimized model, with only two samples connected to a straight line. Therefore, we ignored interpreting of the trend line characteristics of the prediction set, and only the R^2_C , slope, and intercept of the calibration set will be interpreted. As indicated by Williams et al. [34], when the R approached 1 and the slope approached 1 and the intercept approached zero, the model approached excellence. Therefore, to include different species in a model, the species have to be not only in the different values of the constituents to make a wider range for a robust model, but also they must provide the characteristic of the same rate of change of NIR predicted values with the measured values (same slope and slope should approach 1, and intercept is same (no gap) and approached zero). As expected, the trend of R^2 , slope, and intercept of different species were not the same for their different characteristics. However, in some species whose characteristics were similar, the trends were common supported the each other but might positively or negatively to the prediction performance of the model.

From Tables 5.6–5.9, as expected, the intercept of different species illustrated the same trend as slope interpretation, especially when, by the fact, the slope is more than 1 the intercept was with minus sign, and if less than 1 the intercept was with plus sign. While the slope was 1, the intercept was low, closer to zero, and when the slope was more or less than 1, the intercept was high, far from zero.

Therefore, the following were the effects of specific species on the performance of the optimized models interpreted by scatter plot analysis using the R^2 and slope of the trend line of the specific plant in the model developed.

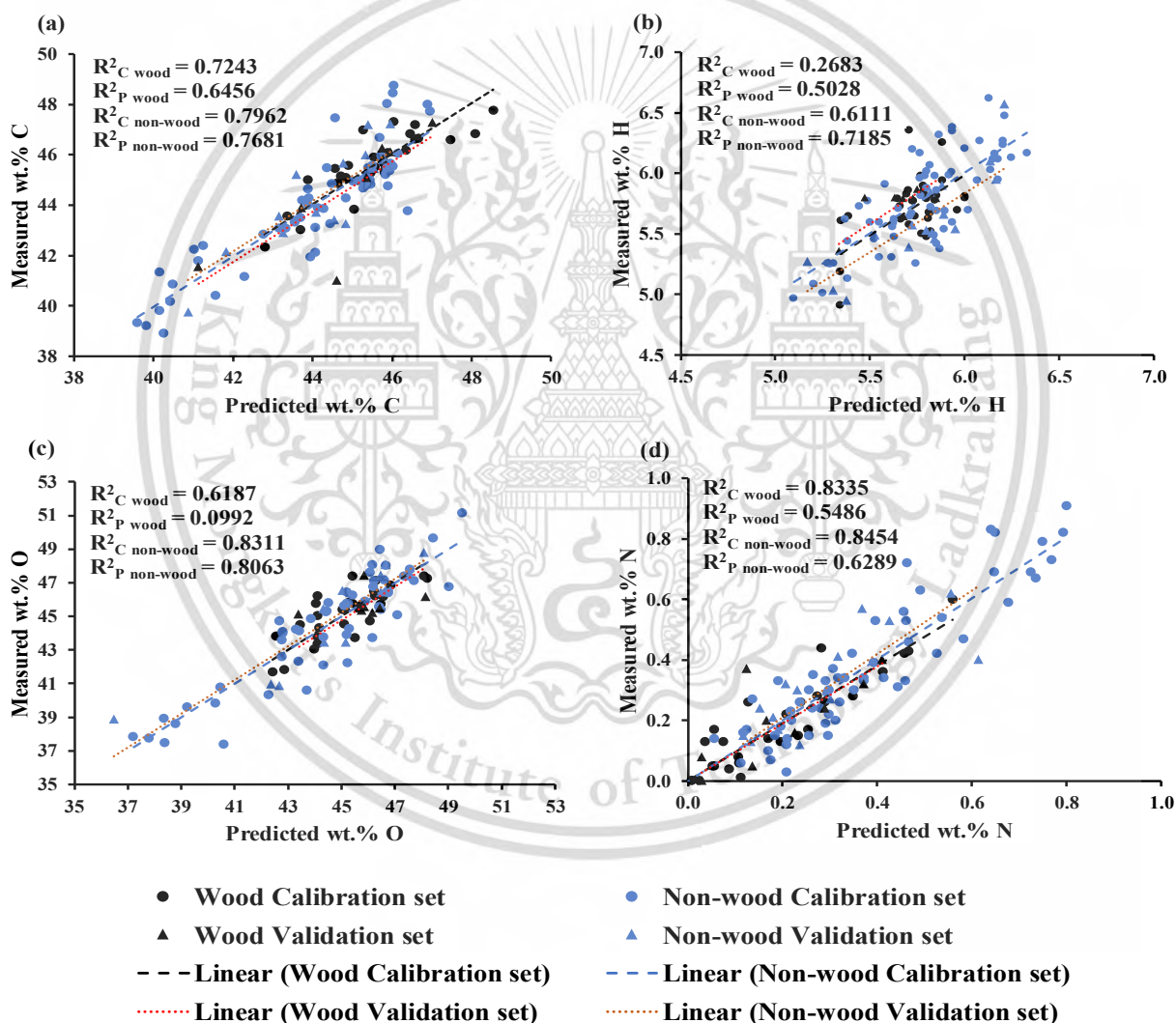


Figure 5.7 The scatter plots of optimized model for wt.% of (a) C, (b) H, (c) O, and (d) N where the simple regression lines of non-wood group and wood group illustrated both in calibration set and validation set.

For C (Table 5.6), by R^2_C interpretation, most non-wood species (agricultural waste) except bagasse and bamboo show unacceptable trend lines compared to wood species samples except pines. Therefore, including the mentioned non-wood species caused a poor effect on the C model. By interpretation of slope, there were three groups of slope (by value round up), i.e., 1 including Eucalyptus, Alnus and Bombax in wood species and corn cob, corn shell, rice husk, and bamboo in non-wood species, less than 1 including pine in wood specie, and more than 1 including corn stover and bagasse indicating unequal slope of different species in the same optimized model show the effect of specific species on model performance. These can be summarized that for the model to be better, pine and corn stover should not be included in modeling for C prediction.

By the same way of interpretation, from Table 5.7, the optimized model for N, pine, and bagasse should not be included; from Table 5.8, for H, pine, Alnus, corn shell, and bagasse should not be included; and from Table 5.9, for O, pine should not be included for better performance of the models. These were due to the poor R and slope of the eliminated species, which were not in accordance with the other species.

These results show that the different species affected the model performance of each parameter prediction in a different manner, and by scatter plot analysis, which of these species were affecting the model negatively and how to improve the model performance were indicated.

Table 5.6 The trend line characteristics of specific biomass species for Carbon evaluation optimized model.

		Carbon (wt.%)					
Particular	Biomass Species	R^2_C	R^2_P	Slope _C	Slope _P	Intercept _C	Intercept _P
Wood	Euca	0.6779	1.0000	0.9808	5.4617	0.8006	-202.6600
	Pine	0.2502	1.0000	0.2264	1.0848	36.2520	-3.7219
	Alnu	0.7491	1.0000	0.7254	-16.8990	12.7000	819.4200
	Bombax	0.8110	1.0000	1.1270	0.9097	-5.3606	4.1430
Non-Wood	<i>Zea mays</i> -Cob	0.2480	0.9542	0.6228	1.8112	16.7390	-35.8510
	<i>Zea mays</i> -Stover	0.6332	1.0000	1.7168	0.2151	-32.1370	33.6140
	<i>Zea mays</i> -Shell	0.3300	0.4618	0.8945	0.2524	5.0232	34.2500
	Ricehusk	0.3770	1.0000	0.9257	2.5087	2.9918	-62.7580
	Bagass	1.0000	1.0000	2.6090	-0.1076	-70.2900	48.2050
	Bamboo	0.9313	1.0000	1.3789	7.6002	-17.0530	-297.8600

Table 5.7 The trend line characteristics of specific biomass species for Nitrogen evaluation optimized model.

		Nitrogen (wt.%)					
Particular	Biomass Species	R^2_c	R^2_p	Slope _c	Slope _p	Intercept _c	Intercept _p
Wood	Euca	0.5701	1.0000	0.7531	0.4663	0.0233	-0.0135
	Pine	0.2317	1.0000	0.2828	0.8790	0.0283	0.0543
	Alnu	0.5878	0.9633	0.5742	1.2687	0.1426	-0.1337
	Bombax	0.9410	1.0000	1.1614	-2.0520	-0.0748	0.6245
Non-Wood	<i>Zea mays</i> -Cob	0.6807	0.5554	0.8615	1.1809	0.0443	-0.0372
	<i>Zea mays</i> -Stover	0.6200	1.0000	0.9025	0.2654	0.0472	0.4721
	<i>Zea mays</i> -Shell	0.8641	0.6536	1.1203	1.0135	-0.0629	0.0569
	Ricehusk	0.8848	1.0000	1.1485	0.2615	-0.0518	0.2394
	Bagass	0.4801	1.0000	0.2992	-1.7907	0.0333	0.5128
	Bamboo	0.8200	1.0000	1.4186	1.6937	-0.1260	-0.0966

Table 5.8 The trend line characteristics of specific biomass species for Hydrogen evaluation optimized model.

		Hydrogen (wt.%)					
Particular	Biomass Species	R^2_C	R^2_P	Slope _C	Slope _P	Intercept _C	Intercept _P
Wood	Euca	0.7289	1.0000	1.5193	0.8197	2.9877	0.9851
	Pine	0.0462	N/A	0.4235	-	3.3450	5.7900
	Alnu	0.0701	1.0000	-0.9476	-0.0456	11.1870	6.0566
	Bombax	0.1629	1.0000	0.5887	0.2547	2.5182	4.4059
Non-Wood	<i>Zea mays</i> -Cob	0.2752	1.0000	1.4447	-0.7296	-2.6372	9.7617
	<i>Zea mays</i> -Stover	0.1173	0.7335	1.2590	1.2413	-1.5538	-1.7143
	<i>Zea mays</i> -Shell	0.0404	0.6033	0.3791	6.5956	3.8515	-34.5000
	Ricehusk	0.7273	0.9896	1.5136	-1.5656	-2.7759	13.3580
	Bagass	0.0067	1.0000	-0.1394	-4.9031	6.4990	34.7330
	Bamboo	0.4456	0.7685	0.9438	1.0741	0.4841	-0.4794

Table 5.9 The trend line characteristics of specific biomass species for Oxygen evaluation optimized model.

		Oxygen (wt.%)					
Particular	Biomass Species	R^2_C	R^2_P	Slope _C	Slope _P	Intercept _C	Intercept _P
Wood	Euca	0.3842	1.0000	0.5993	0.3416	18.5080	29.7010
	Pine	0.2854	1.0000	0.3913	-0.0362	27.5290	47.1430
	Alnu	0.4993	1.0000	0.5014	0.9362	23.0630	4.5052
	Bombax	0.7459	1.0000	1.3490	-1.1972	-15.4990	100.4800
Non-Wood	<i>Zea mays</i> -Cob	0.6501	1.0000	1.3700	8.9169	-17.1250	-368.0300
	<i>Zea mays</i> -Stover	0.8611	1.0000	1.5098	-0.3972	-22.8340	64.3960
	<i>Zea mays</i> -Shell	0.3063	0.7989	0.8399	2.0886	6.9934	-48.2230
	Ricehusk	0.9499	1.0000	1.0623	0.3529	-2.3570	25.9720
	Bagass	1.0000	NA	0.0784	NA	42.8950	NA
	Bamboo	0.9301	1.0000	1.1793	3.0761	-8.5173	-95.5720

5.3.6 Comparison of Model Performance between Using Chipped and Ground Biomass

Spectra

In this section, the model performance of chipped biomass for ultimate analysis parameters to the model of ground biomass [13] derived from the same sample varieties is compared. The comparison is based on the metrics R^2_C , RMSEC, R^2_P , RMSEP, and RPD. The results demonstrate that chipped biomass generally performs less effectively in these models compared to ground biomass, except for wt.% of O.

For wt.% of C and wt.% of H, both chipped and ground biomass models demonstrated better performance when employing the GA-PLSR model. This outcome aligns with expectations, as GA optimizes feature selection to maximize fitness, while PLSR maximizes covariance between absorbance values and areas of interest.

For wt.% of C, the GA-PLSR model applied to ground biomass yielded an R^2_C of 0.7851, RMSEC of 0.9753 wt.%, R^2_P of 0.7217, RMSEP of 0.9740 wt.%, and RPD of 1.93 [13]. In contrast, the model applied to chipped biomass performed less effectively (Table 5.2). Therefore, it is recommended to adopt the GA-PLSR model with sd2 preprocessing on ground biomass when evaluating wt.% of C.

Similarly, the GA-PLSR model applied to ground biomass outperforms that of chipped biomass for wt.% of H. Ground biomass yielded an R^2_C of 0.8814, RMSEC of 0.1041 wt.%, R^2_P of 0.7678, RMSEP of 0.1434 wt.%, and RPD of 2.14 [13], whereas chipped biomass lagged behind (Table 5.2). Hence, for wt.% of H, the GA-PLSR model with spectral preprocessing from SNV on ground biomass is recommended.

Regarding wt.% of N, the MP PLSR 5-range method exhibited superior model performance on ground biomass, as evidenced by R^2_C , RMSEC, R^2_P , RMSEP, and RPD values of 0.8682, 0.0675 wt.%, 0.8410, 0.0973 wt.%, and 2.65, respectively [13], when compared to chipped biomass performance obtained from the MP PLSR 3-range method (Table 5.2). This underscores the suitability of ground biomass for evaluating wt.% of N.

Surprisingly, in contrast, for wt.% of O, the model derived from chipped biomass excelled, despite both models utilizing the MP PLSR 5-range method. In the ground biomass, R^2_C , RMSEC, R^2_P , RMSEP, and RPD values were 0.6674, 1.4461wt.%, 0.6289, 1.5275 wt.%, and 1.71, respectively [13], which fell short of chipped biomass results. Hence, it is recommended to adopt the MP

PLSR-5 range method with the preprocessing combination set of 2, 5, 2, 1, and 5 for assessing wt.% of O in chipped biomass. This could be due to ash determination, where ash directly influences %O determination based on Eq 1. Also, ash is typically accumulating in small particles, i.e., the time of grinding in conjunction with subsampling can have an influence on ash determination.

All the above comparisons and findings underscore the importance of selecting the appropriate PLSR-based model for precise analysis of ultimate analysis parameters, depending on the specific parameter of interest. There could be several factors that contribute to the lower performance of the chipped biomass model, which can be addressed to improve the model performance. The key contributing factor to this performance difference is obviously the particle size of the biomass samples. Chipped biomass typically consists of larger and different sizes of particles, leading to increased scattering of NIR light during sample scanning. Consequently, the spectra generated from chipped biomass can be of lower quality, resulting in weaker correlations between spectral data and reference data [38]. Additionally, ground biomass exhibits a more compact and uniform sample structure, reducing the likelihood of NIR light leakage during scanning. Another significant factor affecting the lower model performance is the moisture content in biomass samples. Chipped biomass often contains higher moisture levels, and water has the property of absorbing NIR light in the near-infrared region [39]. This NIR absorption interferes with the measurements and can introduce inaccuracies, particularly for elements like C, H, O, and N. In the chipped biomass models, it is evident that the performance of the prediction set consistently lags behind that of the calibration set. This suggests that the model closely overfits the calibration data, capturing both valuable information and noise or random variations [40]. In the machine learning context, Cawley and Talbot [37] emphasized that overfitting in model selection is likely to be most severe when the sample size is small and the number of hyperparameters to be tuned is relatively large [41]. In our case, the number of latent variables of the best models was high.

Consequently, when new samples are introduced into the prediction set, the model may struggle to generalize and provide accurate predictions. Furthermore, the presence of outliers in the prediction set, which were not accounted for in the calibration set, can further negatively impact the model performance [42].

The performance of ground biomass is better compared to chipped biomass due to several factors. Ground biomass allows for better sample homogenization, ensuring uniformity and consistent composition. Additionally, it offers more control over sample thickness, as chips may vary in thickness, affecting accuracy. Moreover, ground samples reduce light-scattering effects and enable improved penetration of the NIRS signal, allowing for precise and accurate logging of spectral information.

5.4 Conclusions

In this study, PLSR-based models were developed and compared using FT-NIRS to analyze the ultimate analysis parameters of combined non-wood and wood chip biomass, specifically focusing on wt.% of C, H, O, and N content. All chipped biomass samples were scanned within $3594.87\text{--}12489.48\text{ cm}^{-1}$ on the diffuse reflectance with sphere macro sample rotating mode, with a particular emphasis on their suitability for energy application. The model with the optimum performance was selected based on trade-off parameters of R^2_C , RMSEC, R^2_P , RMSEP, RPD, and bias.

The optimum model performance analysis reveals that the model selected for predicting the wt.% of C, H, N, and O in chipped biomass is suitable primarily for initial rough screening. It is recommended to adopt the multi-preprocessing PLSR 5-range method chipped biomass model for wt.% of O content analysis as an alternative method for rapid assessment. However, for the evaluation of wt.% of C, H, and N content, the chipped biomass model performance falls short of the model developed for ground biomass by Shrestha et al. [13]. Thus, it is advisable to use the chipped biomass model solely for initial screening before biomass trading. For a more comprehensive and accurate analysis, it is recommended to grind the chip biomass samples within the range of 0.01 to 3080 μm and employ the GA-PLSR model with sd1 for wt.% of C, GA-PLSR with SNV for wt.% of H, and the MP PLSR 5-range method with combination set of 4, 4, 5, 3, and 4 for wt.% of N, as developed by Shrestha et al. [13]. The LOQ values for C, H, and O were below the model minimum reference value, demonstrating high model sensitivity. However, the LOQ value for N exceeds the minimum reference value, indicating the model detection limit to the minimum value in the calibration sample set range.

By analysis of scatter plots of measured constituent and NIR predicted constituent, the effect of including different biomass species (non-wood and wood species) in the modeling samples was studied. It was concluded that to include different species in a model, the species had to be not only in the different values of the constituents to be predicted to make a wider range for a robust model, but also the different sample species must provide the same rate of change of NIR predicted values with the measured values in the scatter plot (same slope and slope approached to 1, and intercept is same (no gap) and approached zero) for the high-performance model if R is approached to one. The results show that the different species affected the model performance of each parameter prediction in a different manner, and by scatter plot analysis, which of the species affecting the model negatively were identified and dictated how to improve the model performance.

To ensure the model robustness and reliability, it is crucial to expand it by incorporating a wider array of representative non-wood and wood species biomass samples, but the different species must provide the same rate of change of NIR predicted values with the measured values in the scatter plot. Validation and updation using additional unknown samples of the same species are essential for the model effective applicability. Furthermore, exploring alternative machine learning algorithms alongside the recommended model could enhance its practicability. These steps will contribute to not just a more comprehensive and versatile model but also increase its ability for real-world application and improve its overall reliability.

References

1. IRENA. *Bioenergy for the Energy Transition: Ensuring Sustainability and Overcoming Barriers*; International Renewable Energy Agency: Abu Dhabi, United Arab Emirates, 2022.
2. Ness, J.E.; Ravi, V.; Heath, G. *An Overview of Policies Influencing Air Pollution from the Electricity Sector in South Asia*; National Renewable Energy Laboratory: Golden, CO, USA, 2021.
3. Buonocore, J.J.; Salimifard, P.; Michanowicz, D.R.; Allen, J.G. A decade of the US energy mix transitioning away from coal: Historical reconstruction of the reductions in the public health burden of energy. *Environ. Res. Lett.* **2021**, *16*, 054030.
4. Fullerton, D.G.; Bruce, N.; Gordon, S.B. Indoor air pollution from biomass fuel smoke is a major health concern in the developing world. *R. Soc. Trop. Med. Hyg.* **2008**, *102*, 843–851.
5. Liu, T.; Chen, R.; Zheng, R.; Li, L.; Wang, S. Household air pollution from solid cooking fuel combustion and female breast cancer. *Front. Public Health* **2021**, *9*, 677851.
6. Jin, R.; Zheng, M.; Yang, L.; Zhang, Q.; Fu, J.; Yang, R.; Liu, Q.; Shi, J.; Liu, G.; Jiang, G. Indoor exposure to products of incomplete combustion of household fuels in rural Tibetan Plateau. *Environ. Sci. Technol.* **2021**, *56*, 4711–4714.
7. Adamovics, A.; Platea, R.; Gulbe, I.; Ivanovs, S. The content of carbon and hydrogen in grass biomass and its influence on heating value. *Eng. Rural. Dev.* **2018**, *17*, 1277–1281.
8. Jia, Y.; Li, Z.; Wang, Y.; Wang, X.; Lou, C.; Xiao, B.; Lim, M. Visualization of combustion phases of biomass particles: Effects of fuel properties. *ACS Omega.* **2021**, *6*, 27702–27710.
9. Kalinci, Y.; Hepbasli, A.; Dincer, I. Biomass-based hydrogen production: A review and analysis. *Int. J. Hydrog. Energy* **2009**, *34*, 8799–8817.
10. Silva, D.A.d.; Eloy, E.; Caron, B.O.; Trugilho, P.F. Elemental chemical composition of forest biomass at different ages for energy purposes. *Floresta Ambient.* **2019**, *26*. <https://doi.org/10.1590/2179-8087.020116>
11. Ren, X.; Sun, R.; Meng, X.; Vorobiev, N.; Schiemann, M.; Levendis, Y.A. Carbon, sulfur and nitrogen oxide emissions from combustion of pulverized raw and torrefied biomass. *Fuel* **2017**, *188*, 310–323.
12. Vainio, E. Fate of Fuel-Bound Nitrogen and Sulfur in Biomass-Fired Industrial Boilers. Ph.D. Thesis, Åbo Akademi University, Turku, Finland, 2014.

13. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* **2023**, *16*, 5351.
14. Sirisomboon, P.; Funke, A.; Posom, J. Improvement of proximate data and calorific value assessment of bamboo through near infrared wood chips acquisition. *Renew. Energy* **2020**, *147*, 1921–1931.
15. Uddin, M.N.; Ferdous, T.; Islam, Z.; Jahan, M.S.; Quaiyyum, M.A. Development of chemometric model for characterization of non-wood by FT-NIR data. *J. Bioresour. Bioprod.* **2020**, *5*, 196–203.
16. Kumar, P.; Barrett, D.M.; Delwiche, M.J.; Stroeve, P. Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. *Ind. Eng. Chem. Res.* **2009**, *48*, 3713–3729.
17. Worku, L.A.; Bachheti, A.; Bachheti, R.K.; Rodrigues Reis, C.E.; Chandel, A.K. Agricultural residues as raw materials for pulp and paper production: Overview and applications on membrane fabrication. *Membr. J.* **2023**, *13*, 228.
18. Hawanis, H.S.N.; Ilyas, R.A.; Jalil, D.R.; Ibrahim, D.R.; Abdul Majid, D.R.; Ab Hamid, D.N.H. Insights into Lignocellulosic Fiber Feedstock and its Impact on Pulp and Paper Manufacturing: A Comprehensive Review. 2023. Available online: <https://ssrn.com/abstract=4583258> (accessed on 11 November 2023).
19. Aripin, A.M. Potential of Non-Wood Fibres for Pulp and Paper-Based Industries. Ph.D. Thesis, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia, 2014.
20. Rousu, P.; Rousu, P.; Anttila, J. Sustainable pulp production from agricultural waste. *Resour. Conserv. Recycl.* **2002**, *35*, 85–103.
21. Kissinger, M.; Fix, J.; Rees, W.E. Wood and non-wood pulp production: Comparative ecological footprinting on the Canadian prairies. *Ecol. Econ.* **2007**, *62*, 552–558.
22. Channiwala, S.; Parikh, P. A unified correlation for estimating HHV of solid, liquid and gaseous fuels. *Fuel* **2002**, *81*, 1051–1063.
23. Posom, J.; Sirisomboon, P. Evaluation of lower heating value and elemental composition of bamboo using near infrared spectroscopy. *Energy* **2017**, *121*, 147–158.

24. 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.
25. Huang, C.; Han, L.; Yang, Z.; Liu, X. Ultimate analysis and heating value prediction of straw by near infrared spectroscopy. *J. Waste Manag.* **2009**, *29*, 1793–1797.
26. Saha, U.K.; Sonon, L.; Kane, M. Prediction of calorific values, moisture, ash, carbon, nitrogen, and sulfur content of pine tree biomass using near infrared spectroscopy. *J. Near Infrared Spectrosc.* **2017**, *25*, 242–255.
27. 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*, 119698.
28. Shrestha, B.; Shrestha, Z.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Evaluating limit of detection and quantification for higher heating value and ultimate analysis of fast-growing trees and agricultural residues biomass using NIRS. *Eng. Appl. Sci. Res.* **2023**, *50*, 612–618.
29. Maraphum, K.; Ounkaew, A.; Kasemsiri, P.; Hızıroglu, 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. *Eng. Appl. Sci. Res.* **2021**, *49*, 119–126.
30. Chen, Y.M.; Lin, P.; He, Y.; He, J.Q.; Zhang, J.; Li, X.L. Fast quantifying collision strength index of ethylene-vinyl acetate copolymer coverings on the fields based on near infrared hyperspectral imaging techniques. *Sci. Rep.* **2016**, *6*, 20843.
31. Li, C.; He, M.; Cai, Z.; Qi, H.; Zhang, J.; Zhang, C. Hyperspectral Imaging with Machine Learning Approaches for Assessing Soluble Solids Content of Tribute Citrus. *Foods* **2023**, *12*, 247.
32. Araújo, M.C.U.; Saldanha, T.C.B.; Galvão, R.K.H.; Yoneyama, T.; Chame, H.C.; Visani, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometr. Intell. Lab. Syst.* **2001**, *57*, 65–73.
33. Armbruster, D.A.; Pry, T. Limit of blank, limit of detection and limit of quantitation. *Clin. Biochem. Rev.* **2008**, *29*, S49.
34. Williams, P.; Manley, M.; Antoniszyn, J. *Near Infrared Technology: Getting the Best out of Light*; African Sun Media: Stellenbosch, South Africa, 2019.

35. 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 Biol. Biochem.* **2008**, *40*, 1923–1930.
36. Workman, J., Jr.; Weyer, L. *Practical Guide to Interpretive Near-Infrared Spectroscopy*; CRC Press: Boca Raton, FL, USA, 2007.
37. Cawley, G.C.; Talbot, N.L.C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
38. Hans, G.; Allison, B. On-line characterization of wood chip brightness and chemical composition by means of visible and near-infrared spectroscopy. *Holzforschung* **2021**, *75*, 989–1000.
39. Liang, L.; Fang, G.; Deng, Y.; Xiong, Z.; Wu, T. Determination of moisture content and basic density of poplar wood chips under various moisture conditions by near-infrared spectroscopy. *For. Sci.* **2019**, *65*, 548–555.
40. Gillespie, G.D.; Everard, C.D.; McDonnell, K.P. Prediction of biomass pellet quality indices using near infrared spectroscopy. *Energy* **2015**, *80*, 582–588.
41. Ludwig, B.; Murugan, R.; Parama, V.R.; Vohland, M. Accuracy of estimating soil properties with mid-infrared spectroscopy: Implications of different chemometric approaches and software packages related to calibration sample size. *Soil Sci. Soc. Am. J.* **2019**, *83*, 1542–1552.
42. Toscano, G.; Leoni, E.; Gasperini, T.; Picchi, G. Performance of a portable NIR spectrometer for the determination of moisture content of industrial wood chips fuel. *Fuel* **2022**, *320*, 123948.

Chapter 6

NIR spectroscopy as an alternative to thermogravimetric analyzer for biomass proximate analysis : Comparison of chip and ground biomass models

This study investigates the non-destructive analysis of proximate parameters (moisture content, MC; volatile matter, VM; fixed carbon, FC; ash content) in various chipped and ground biomass using a combination of destructive thermogravimetric analysis (TGA) and non-destructive near-infrared spectroscopy (NIRS) with partial least squares regression (PLSR). The thermogravimetric method determines proximate analysis data through TG and DTG curves, tracking biomass mass loss over time or temperature. NIRS scans chipped biomass in diffuse reflectance, and ground biomass in transreflectance mode, covering the wavenumber range from 3595 to 12,489 cm^{-1} . PLSR-based models (Full-PLSR, GA-PLSR, SPA-PLSR, MP PLSR 5-range method, and MP PLSR 3-range method) are developed and evaluated based on R^2P , RMSEP, and RPD. MC and FC models for chip biomass exhibit satisfactory performance, making them cautiously applicable in various applications, including research. Optimal models for MC and FC in chip biomass, constructed using GA-PLSR with the second derivative and Full-PLSR with a constant offset, yield high R^2P values (0.8654 and 0.8773), low RMSEP values (0.85% and 2.12%), and high RPD values (2.9 and 3.0), indicating applicative capabilities. Other parameters such as MC and FC in ground biomass, as well as VM and ash content in both chip and ground biomass, are found suitable for rough screening. Model sensitivity, assessed by calculating LOQ, indicates high sensitivity for VM in both chip and ground biomass and FC in chip biomass, as the calculated LOQ value is lower than the minimum reference values used during model development. However, for the remaining parameters, LOQ values surpass the established minimum reference value, suggesting limitations in predicting samples below the calibration range. Continuous model enhancement incorporating an ample number of representative biomass samples and consistent validation with unknown samples are imperative for ensuring accurate predictions.

Keywords: biomass; proximate analysis; thermogravimetry; near-infrared spectroscopy; partial least squares regression

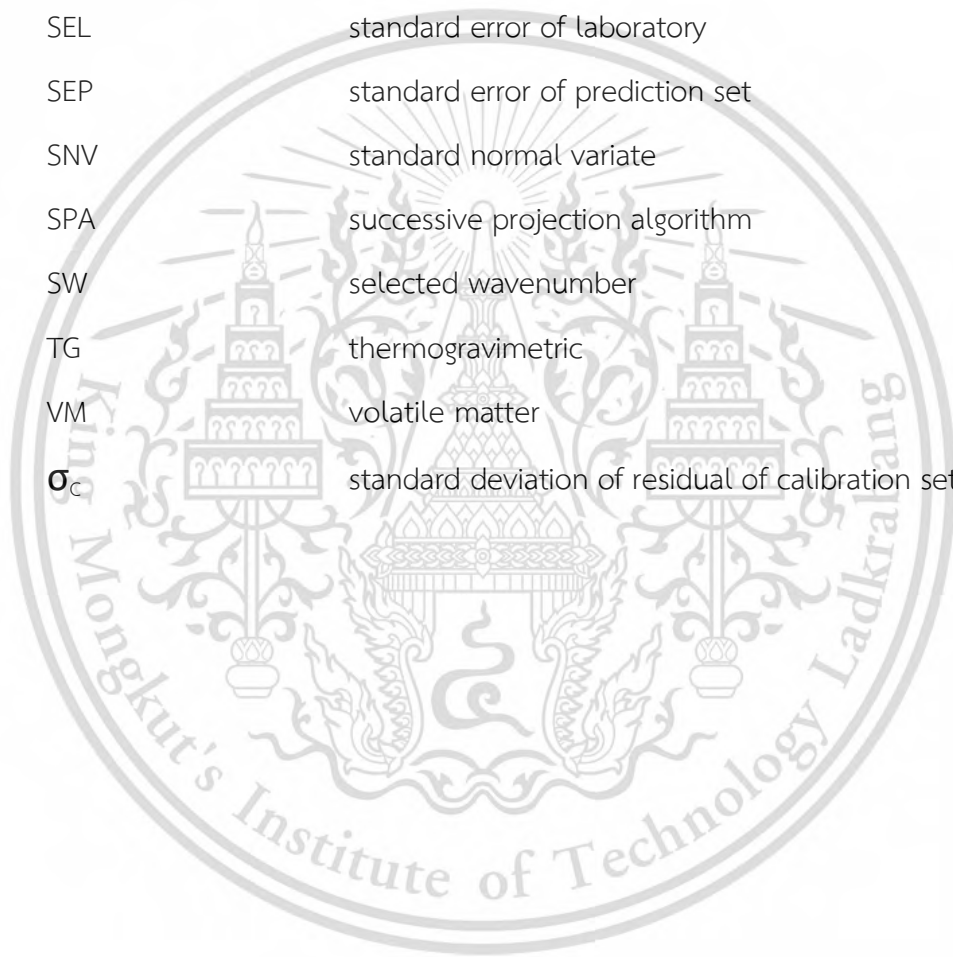


*This chapter constituted the publication article: Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Pornchaloempong, P.; Funke, A. NIR Spectroscopy as an Alternative to Thermogravimetric Analyzer for Biomass Proximate Analysis: Comparison of Chip and Ground Biomass Models. *Energies* 2024, 17, 800. <https://doi.org/10.3390/en17040800>

Abbreviations

D1	first derivative
D2	second derivative
DTG	derivative thermogravimetric
FC	fixed carbon
FT	Fourier transform
g	gap
GA	genetic algorithm
LVs	number of latent variables
LOQ	limit of quantification
Max	maximum
MC	moisture content
Min	minimum
Mean	average
MSC	multiplicative scatter correction
MP	multi-preprocessing
N	nitrogen
N_T	total number of sample
N_c	number of sample in calibration set
N_p	number of sample in prediction set
NIRS	near-infrared spectroscopy
O	oxygen
PLSR	partial least squares regression
R^2	coefficient of determination
R^2_c	coefficient of determination of calibration set
R^2_p	coefficient of determination of prediction set

RPD	ratio of prediction to deviation
RMSE	root mean square error
s	segment
S_c	slope of the regression line
SD	standard deviation
SEC	standard error of calibration set
SEL	standard error of laboratory
SEP	standard error of prediction set
SNV	standard normal variate
SPA	successive projection algorithm
SW	selected wavenumber
TG	thermogravimetric
VM	volatile matter
σ_c	standard deviation of residual of calibration set



6.1 Introduction

Biomass is a renewable, carbon neutral, abundantly available, and sustainable energy source that holds the potential to reduce or replace reliance on fossil fuels, contributing to energy savings and environmental preservation [1]. Studies indicate that the global production of lignocellulosic biomass, including crop residues, wood, and grass, amounts to approximately 181.5 billion tonnes each year [2]. This highlights the substantial potential of utilizing agricultural residues and wood from fast-growing trees for bio-conversion. A primary concern is that biomass generally has a lower energy density compared to fossil fuels. Nevertheless, the application of diverse biomass energy technologies rooted in thermochemical treatments, namely, pyrolysis (300 to 1000 °C), gasification (500 to 1300 °C), liquefaction (250 to 550 °C), torrefaction (200 to 300 °C), and combustion (700 to 1000 °C) displays the potential to effectively transform biomass into valuable energy resources [3,4].

The wide range of biomass sources, the diverse assortment of agricultural and tree waste materials, and the continuous variations in their properties during processing, storage, and transportation collectively pose challenges in effectively characterizing and understanding biomass as an efficient fuel. The characterization of biomass is crucial for reliably predicting its behavior as a fuel [5]. In the context of the thermal conversion of biomass, proximate analysis emerges as a pivotal method for characterization [4]. Proximate analysis is used to determine the mass percentages on a wet basis of MC, VM, FC, and ash content, all of which significantly influence both combustion behavior and plant design.

During the biomass combustion process, the rate of heat release is primarily influenced by proximate analysis parameters, including MC, VM, FC, and ash content [6]. Adjustments in these proximate analysis values can yield variations in crucial combustion performance indicators such as ignition time and temperature, burnout time and temperature, and the overall heat release. Higher moisture levels decrease the higher heating value of biomass [7] and lead to reduced combustion efficiency, whereas higher ratios from VM to FC correspond to increased fuel reactivity [8]. VM refers to the components in biomass that are converted into a gaseous form through

thermal decomposition, whereas FC represents the non-volatile segment of the biomass. A high VM content does not necessarily ensure a high calorific value because some VM constituents are derived from non-combustible gases like CO₂ and H₂O. Lower MC and VM contents contribute to the improvement of the FC content in biomass samples [9]. A high FC signifies a higher carbon content in the biomass, resulting in higher energy content and generating maximum heat during combustion [10]. Therefore, the FC content of biomass positively influences its energy potential and calorific value. Conversely, a higher presence of VM in the biomass results in lower FC, consequently impacting the energy potential of the biomass negatively.

Ash is an inorganic residue that remains after ignition of combustible substance [11,12]. The energy required for the thermal breakdown and phase transition of ash-forming inorganics is extracted from the biomass burning energy, resulting in a reduction in calorific value [13]. The presence of high MC and elevated ash content in biomass leads to ignition and combustion problems [5]. Also, the presence of ash influences corrosion and slag formation [8].

Various standard procedures exist to measure these proximate analysis parameters. ASTM E871-82 (2019) [14] addresses moisture, ASTM E872-82 (2019) [15] handles volatile matter, and ASTM E1755-01 (2020) [16] tackles ash in biomass. FC is usually determined by calculating the difference between 100% and the sum of MC, VM, and ash content [12,17]. These standard methods are involved in heating a sample under specific conditions and assessing the change in weight. For proximate analysis of biomass, experimentation using a simple oven to determine MC and a furnace to determine VM and ash contents. The experimental techniques employed to estimate these parameters often consume considerable time, are financially demanding, and carry an increased likelihood of experimental error. Additionally, considering the large number of samples needed to determine the proximate analysis data for biomass characterization, the process is both tedious and destructive in nature. Hence, various researchers have recommended the use of TGA for the proximate analysis of biomass. For instance, Posom et al. (2020) recommend the use of TGA to determine the proximate analysis of bamboo chip and leucaena pellets biomass [18], Torquato et al. (2017) investigated the appropriate conditions

for determining proximate parameters of lignocellulosic agricultural residue using TGA [19], and García et al. (2013) analyzed 13 different types of biomass for proximate analysis using TGA [8]. Similarly, previous research studies have demonstrated the potential of utilizing NIR spectroscopy, as an alternative method, for assessing biomass proximate analysis parameters across the wavelength range between 780 and 2500 nm. Posom et al. (2022) developed a reliable and accurate model for online measurement of MC in sugarcane [20]. Toscano et al. (2022) estimated the MC of industrial wood chip fuel using a portable NIR spectrometer [21]. Sirisomboon et al. (2020) compared the performance of proximate data models, namely, MC, VM, FC, and ash, through direct scanning of bamboo chips [22]. Shrestha et al. (2018) evaluated MC in bamboo chips using diode array near-infrared spectroscopy [23]. Posom et al. (2017) determined the HHV, VM, FC, and ash content of ground bamboo using FT-NIRS [24]. Adnan et al. (2017) assessed the feasibility of NIRS and chemometrics for rapid and non-destructive prediction of MC in intact green coffee beans [25]. The above research demonstrates the potential of TGA for determining proximate parameters and NIRS for rapidly predicting biomass proximate parameters. Therefore, this study integrates TGA and NIRS, developing NIRS as an alternative method for biomass proximate analysis using calibration models with chemometrics, specifically PLSR, for rapid analysis. This enables real-time measurements and high-throughput screening of biomass samples. This approach eliminates the need for sample preparation and the use of consumables, thereby reducing costs and minimizing negative environmental impact. Additionally, it is a safe method with less operator dependency.

Although NIRS is a rapid, reliable, and non-destructive analytical method, it requires the development of individual calibration models based on spectral data and each reference parameter. This process can be time consuming and costly. However, in the long term, it proves to be beneficial for the accurate assessment of proximate analysis parameters. To the best of our knowledge, there have been no publications reporting the development of a global NIRS model for proximate parameters using fast-growing trees and agricultural residue which can be categorized into wood and non-wood biomass. Many studies on biomass and NIRS modeling have

found that PLSR is a very effective and widely used method for quantitative analysis [26–29]. Therefore, in this work, we elucidate the rationale for combining both types of biomass materials for model development. Additionally, this study marks the first application of the recent MP methods, specifically the 5-range and 3-range methods, for the NIRS evaluation of proximate parameters in biomass. Therefore, the main aims of this study are as follows:

1. To determine the proximate analysis parameters, i.e., MC, VM, FC, and ash content of biomass, using TGA.
2. To develop separate calibration models using Full-PLSR, GA-PLSR, SPA-PLSR, the multi-reprocessing 5-range PLSR, and the multi-preprocessing 3-range PLSR for the non-destructive assessment of each proximate analysis parameter in both chipped and ground biomass.
3. To compare and select the best performing PLSR-based model for each proximate analysis parameter from chipped and ground biomass, and establish it as a rapid, reliable, non-destructive alternative method for assessment of proximate analysis parameters.
4. To determine the LOQ value for each proximate parameter using the calibration set of the proposed high-performance model, both for chip and ground biomass.

6.2 Materials and Methods

Figure 6.1 illustrates the overall research methodology employed to assess the proximate analysis parameters and select the best performance PLSR-based models using TGA (35–700 °C) and NIRS (3594.87–12,489.48 cm^{-1}).

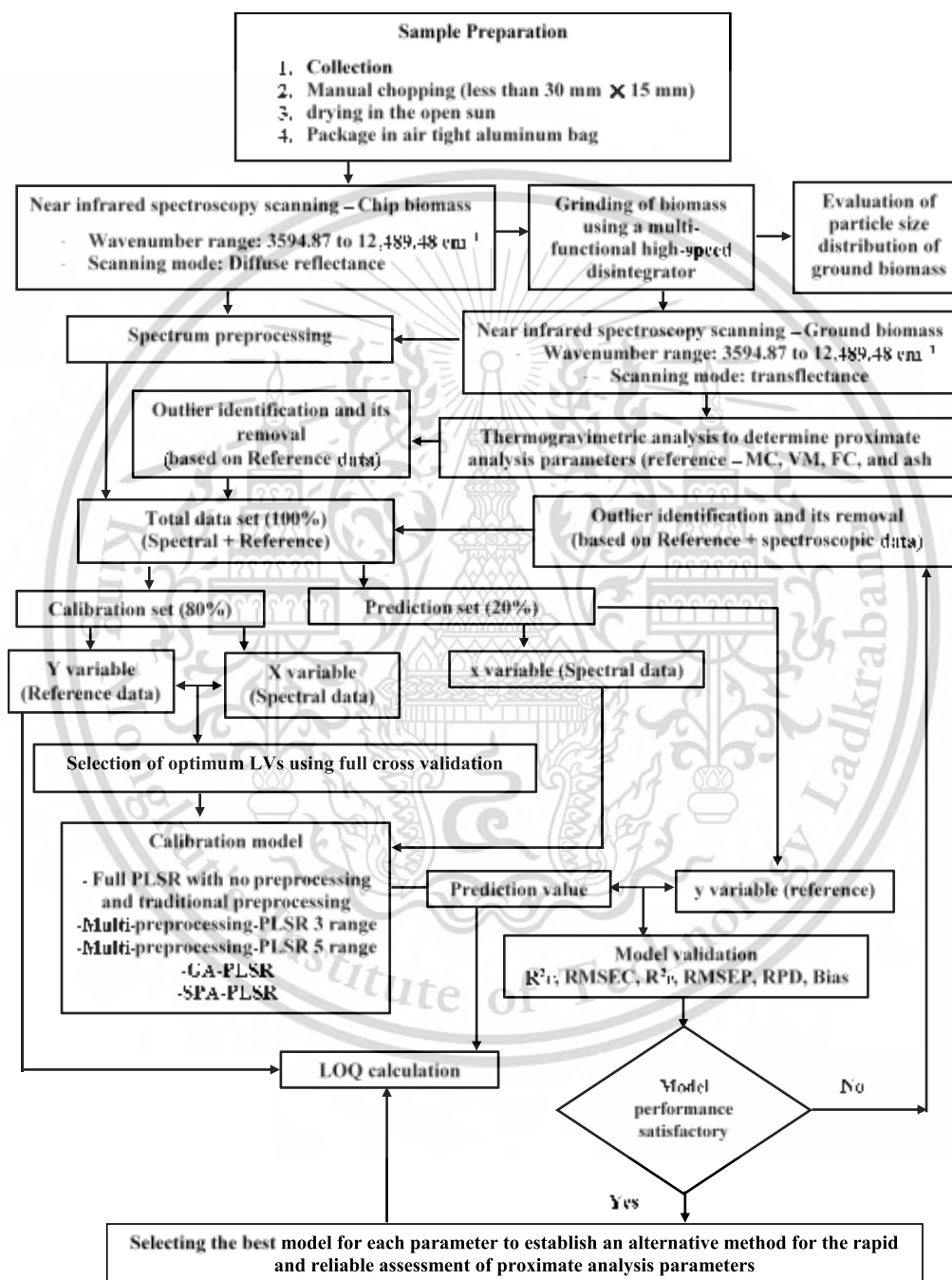


Figure 6.1 Flowchart depicting the overall research methodology for the rapid assessment of biomass proximate analysis parameters using TGA, while using NIRS with PLSR as an alternative.

6.2.1 Sample Preparation

Ten different varieties of biomass samples were collected from various locations in Nepal. Wood samples included four fast-growing species: (1) *Alnus nepalensis*, (2) *Pinus roxburghii*, (3) *Bombax ceiba*, and (4) *Eucalyptus camaldulensis*. Non-wood samples included five agricultural residues: (1) *Zea mays* (cob), (2) *Zea mays* (shell), (3) *Zea mays* (stover), (4) *Oryza sativa*, and (5) *Saccharum officinarum* and one fast-growing tree (6) *Bombusa vulagris*. These samples were manually chopped, dried under the open sun, and then transported to the laboratory for NIRS and TGA experiments to record the spectral data and proximate analysis data for the development of PLSR-based models [26].

6.2.2 FT-NIRS Scanning

In Figure 6.2a, chip biomass samples with dimensions less than 30 mm × 15 mm were scanned using an FT-NIR spectrometer (MPA, Bruker, Ettlingen, Germany) in diffuse reflectance and sphere macro sample rotating mode. The samples were fully filled into a quartz-bottom sample cup with a diameter of 90 mm and a height of 90 mm. They were then covered with aluminum sheet to prevent NIR leakage. Subsequently, the chip biomass samples were ground using a multifunctional high-speed disintegrator (WF-04, Thai grinder, Thailand). The ground biomass samples were scanned in a transmittance mode (Figure 6.2b). These samples were placed in glass vials with a diameter of 20 mm and a height of 40 mm.

NIR scanning of both chip and ground biomass was performed twice without changing their positions under a controlled laboratory environment, maintaining a temperature of 25 ± 2 °C. For both forms of biomass samples, the spectrometer operated at a resolution of 16 cm^{-1} , with a background scan time and sample scan time of 32 scans (average), logging absorbance data as $\text{Log}(1/R)$ within a wavenumber range from 3594.87 to $12,489.48 \text{ cm}^{-1}$. The average absorbance value was calculated and applied for model development. A gold plate scan was performed for every new scanning, and aluminum plates/handles were used to prevent possible NIR leakage. Figure 6.3 shows the raw spectra obtained by scanning the chip biomass and ground biomass using an FT-NIR spectrometer. The ground sample was randomly selected and its particle size distribution was assessed at the Scientific and Technological Research Equipment Center at Chulalongkorn University, Bangkok, Thailand, using the Mastersizer 3000 instrument (MAL1099267,

Hydro MV). Figure 6.4 shows the representative particle size distribution of the ground biomass used in this research. The particle sizes range from 0.01 to 3080 μm , with an average particle size of 251.18 μm .

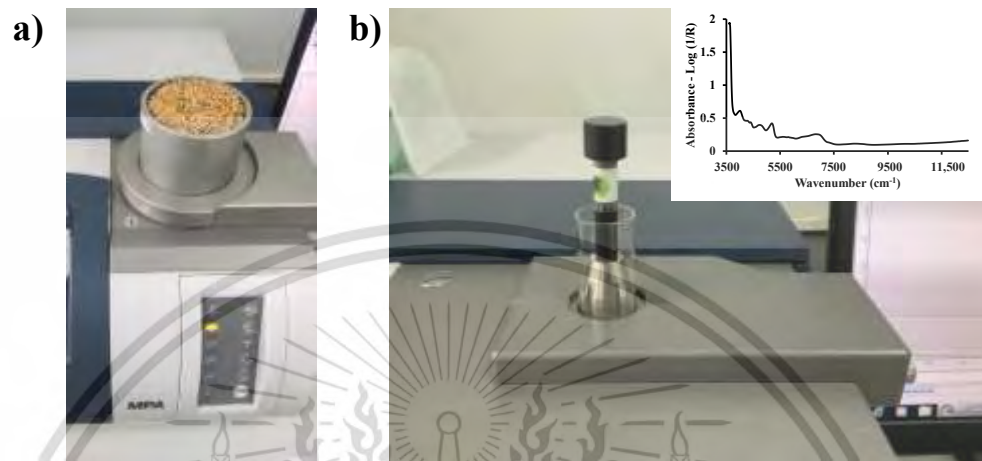


Figure 6.2 FT-NIR spectrometer scanning of (a) chip biomass in diffuse reflectance and sphere macro sample rotating mode using a quartz-bottom sample cup, and (b) ground biomass in transmittance mode using a glass vial.

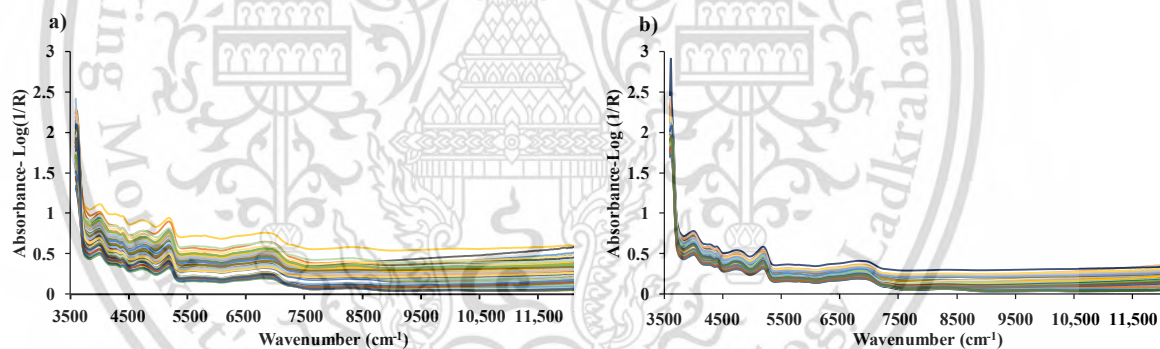


Figure 6.3 Raw spectra of (a) chip biomass obtained from diffuse reflectance mode of FT-NIRS scanning and (b) ground biomass obtained from transmittance mode of FT-NIRS scanning across the entire wavenumber range from 3594.87 to 12,489.48 cm^{-1} .

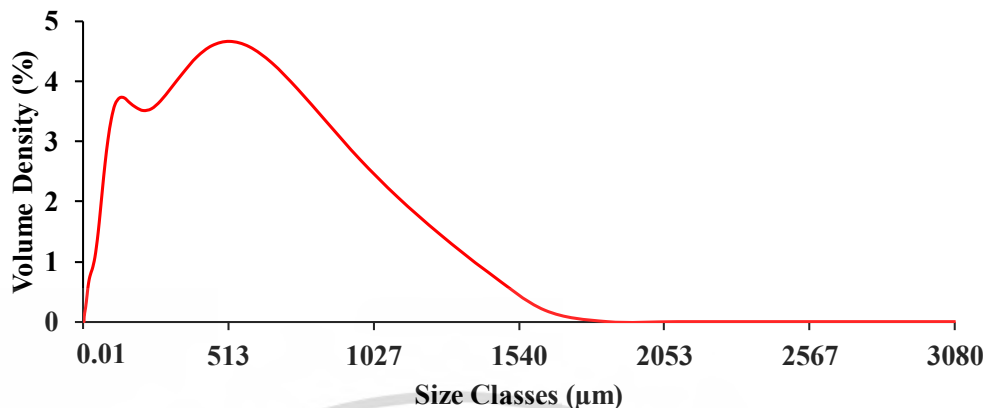


Figure 6.4 Representative particle size distribution of the ground biomass ranging from 0.01 to 3080 µm.

6.2.3 Reference Analysis (Proximate Analysis)

TGA (TG 209 F3 Tarsus, Netzsch, Germany) with a microbalance sensitivity of 0.1 µg resolution was employed for direct combustion in air atmosphere [30], i.e., O₂ and N₂ in 1:4 ratio, within a temperature range from 35 °C to 700 °C. A heat flow rate of 10 °C/min was utilized to record the mass loss of ground biomass sample of 6 ± 0.5 mg with respect to time or temperature. Prior to thermal biomass degradation, all samples were held isothermally at 35 °C for 10 min. N₂ is used as a protective layer to create a stable and inert environment [31]. Proteus 6.0.0 (NETZSCH software, Germany) was used to generate TG and DTG curves. The TG and DTG curves were typical of the degradation process, which is generally divided into four significant stages: MC removal (<110—Stage I), release and combustion of VM (197–350 °C—Stage II), char combustion (360–600 °C—Stage III), and ash lifting (>600 °C—Stage IV) [24] (Figure 6.5). The representative TG and DTG curves of wood and non-wood biomass used in this study are shown in Figure 6.6a and Figure 6.6b, respectively.

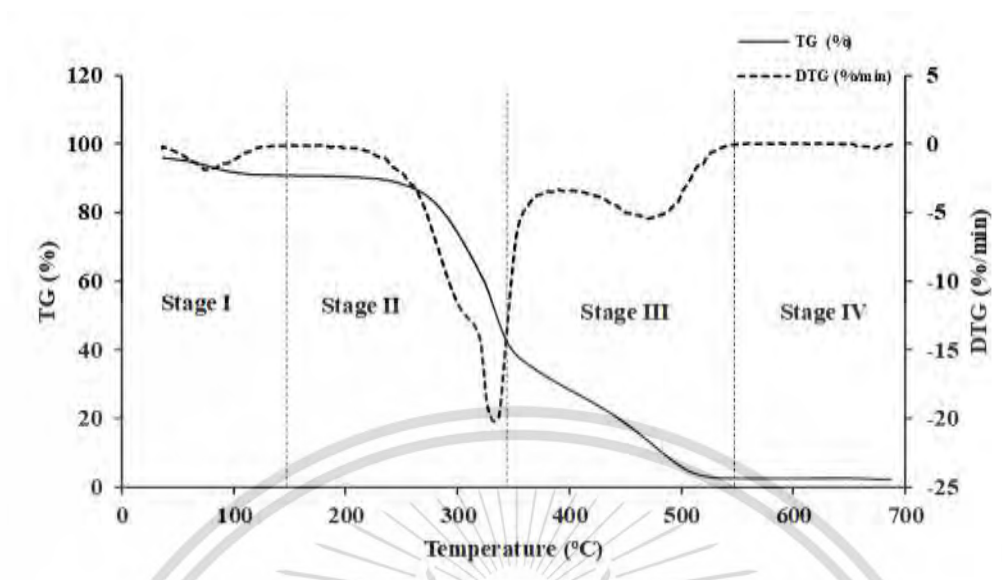


Figure 6.5 The TG and DTG curves obtained from TGA analysis of Nepalese biomass

Biomass is composed of cellulose, hemicellulose, and lignin [32]. As evident from Figure 6.6b, the combustion of lignocellulosic biomass exhibited two distinct major decomposition peaks. The first zone is within a temperature range of approximately 180–400 °C, where the rate of weight loss of biomass is at its maximum and the peak is prominently displayed. This range corresponds to the decomposition of cellulose and hemicellulose. During this phase, volatiles are released and subsequently ignited, resulting in the formation of char [33]. The second zone spans the temperature range from approximately 380 to 600 °C, where a broader peak was observed, associated with char oxidation [33]. In this stage, lignin assumes a key role as the primary contributor to biomass char formation. Beyond 535 °C, the decomposition primarily relates to inorganic matter, specifically the ash content.

In our reference test using the TGA (TG 209 F3 Tarsus, Netzsch, Germany), a small sample of approximately 6 mg was employed. Consequently, all the FT-NIRS scanned chip samples were ground to achieve a homogeneous mixture for obtaining the proximate parameters. It is important to note that this grinding method might slightly affect the measured constituent quantity in the sample, especially concerning MC. However, it is the only viable approach we have to compromise for estimating the MC in the chip sample using a small sample size for TG 209 F3 Tarsus.

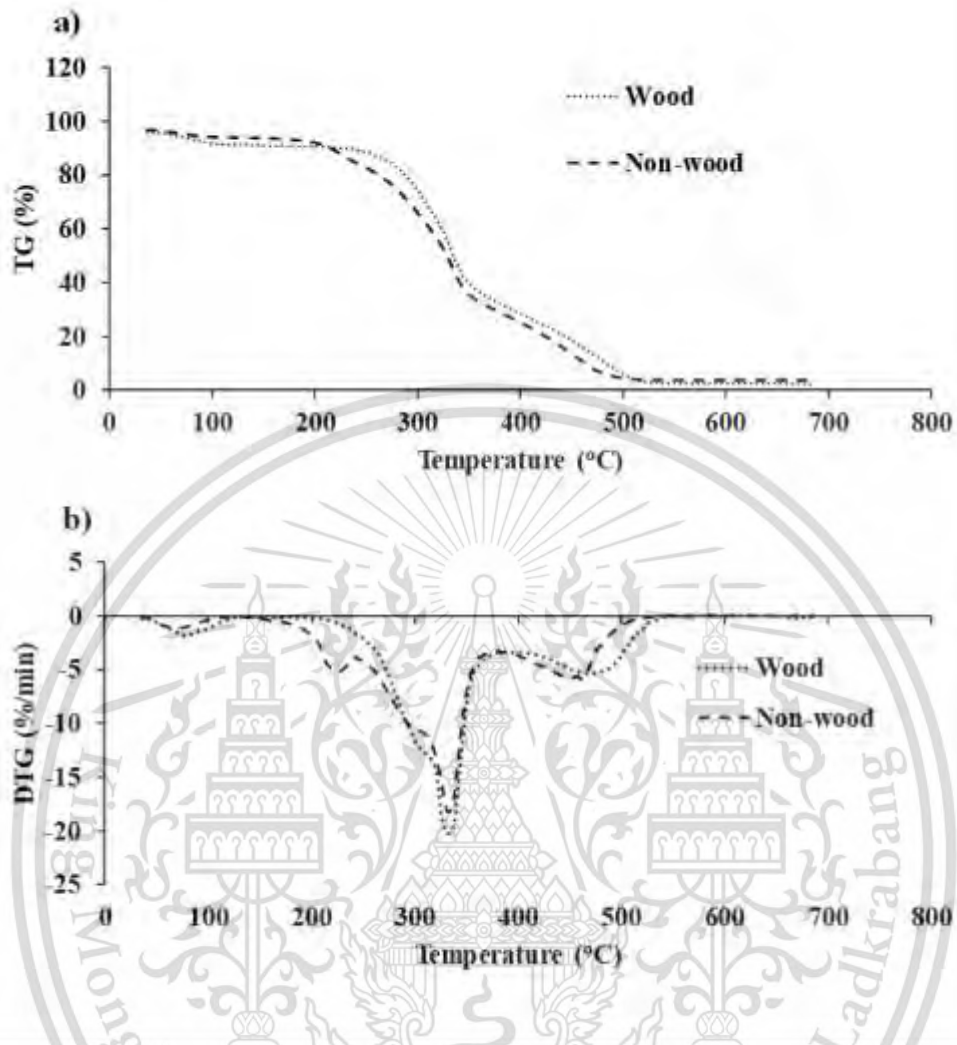


Figure 6.6 (a) TG and (b) DTG profiles of fast-growing tree and agricultural residue biomass obtained using TGA within the temperature range from 35 to 700 °C with a heat flow rate of 10 °C/min.

6.2.4 Outlier Identification

To achieve a high-performance predictive model, the identification and removal of outliers from the complete dataset is one of the most crucial procedures in NIR modeling. In this study, outliers were identified using two methods: one based solely on reference data and the other utilizing both reference and spectroscopic data.

Outlier for the reference data is calculated using the following equation:

$$\frac{(X_i - \bar{X})}{SD} \geq |\pm 3|, \quad (1)$$

where X_i is the measured reference value of sample i , and \bar{X} and SD denote the average and standard deviation of the measured values across all samples. If Equation (1) is satisfied for any sample i , the sample is considered an outlier and is removed from the complete dataset [26,34].

Following the removal of outliers based on Equation (1), the complete dataset was used to develop a model. If the model performance was deemed unsatisfactory, a further outlier identification process was undertaken using both reference and spectroscopic data. This method involved a full cross-validation using PLS regression on total dataset to identify outliers. A scatter plot was then created to compare the measured and predicted values. Outliers were carefully identified and subsequently removed if their behavior significantly deviated from the rest of the data points. This meticulous outlier detection process ensures the model accuracy by eliminating potential data anomalies that could impact the reliability of predictive results.

6.2.5 Spectral Preprocessing

Spectral preprocessing holds significant importance in NIR modeling to achieve high-performance predictions of the reference values using NIR spectral data. NIR raw spectral data often contain instrumental noise, baseline shift, overlapping peaks, scattering, and other errors [26]. Therefore, to enhance spectral features and extract the maximum relevant information, spectral preprocessing, which involves mathematical transformations, is vital prior to model development [35,36].

Prior to model development, raw spectra obtained from both chip and ground biomasses were subjected to ten different types of traditional spectral preprocessing, which include D1, D2, constant offset, SNV, MSC, min-max normalization, mean centering, vector normalization, D1 + MSC, and D1 + vector normalization. Additionally, the raw spectra were pretreated using MP techniques with a 5-range and 3-range to assess prediction performance within the same wavenumber range. In the MP 5-range technique, the entire spectral range is divided into five equal sections, while in the MP 3-range technique, the entire spectral range is divided into three different sections. It is preprocessed with different combination sets obtained from seven different preprocessing techniques, labeled as: 0 = empty (absorbance values = 0), 1 = raw spectra, 2 = SNV, 3 = MSC, 4 = D1, 5 = D2, and 6 = constant offset, covering the range from 3594.87 to

12,489.48 cm^{-1} [26]. These preprocessing steps could potentially improve and or even reduce the predictive accuracy [37] of proximate parameters for both chip and ground biomass.

The MATLAB–R2020b (MathWorks, Natick, MA, USA) with an in-house code was employed for conducting both traditional spectral preprocessing and multi-spectral preprocessing techniques.

6.2.6 Partial Least Squares Regression Model Development

After the identification and removal of outliers from the complete dataset, as shown in Figure 6.1, PLSR models were developed by manually splitting the total dataset which was run in descending order into a calibration set (80%) and a validation set (20%), ensuring that both sets include all varieties of biomass that have been considered. The calibration set is used to establish a mathematical model that correlates the reference data and spectral data of biomass samples. It includes the maximum and minimum reference values of the identified proximate parameters, ensuring the model was adequately trained to handle the entire spectrum of data. On the other hand, the prediction set employs the developed model to predict the properties of an unknown sample. In our study, $\text{Log}(1/R)$ at different wavenumbers were used as independent variables in the regression model, and the regression coefficient represents the weight of each independent variable in series, where each weight was the change for a one-unit change in $\text{Log}(1/R)$ at each wavenumber for prediction of a dependent variable.

The NIR spectra of each biomass sample contained multiple independent variables, resulting in collinearity issues, redundant information, and heightened computational complexity [20]. To tackle these challenges, uninformative wavelength variables, having either negligible or adverse impacts on model performance, need identification and elimination. In this study, two wavelength selection processes, GA and SPA, were employed to develop a PLSR model. The primary objective in choosing these wavelength selection methods is to improve predictive capacity, achieve higher model efficiency, and create simpler models that are more interpretable [38].

GA is an evolutionary optimization technique that mimics natural selection and genetic mechanisms to find optimal solutions, utilizing three key operators: selection, crossover, and mutation [39,40]. In feature wavelength selection, it begins by generating a population based on

available features. Subsequently, the algorithm creates successive populations by selecting individuals based on their fitness level in the problem domain. These individuals undergo recombination through selection, crossover, and mutation across multiple generations, progressively eliminating redundant information from the chosen wavelengths [41]. The optimal member of the population becomes the selected feature.

SPA is a forward selection method that selects variables with minimal redundant information from the informative variable [38]. The variable selected by SPA has the greatest projection onto the orthogonal subspace of the previously selected variable [42]. In SPA, the initial phase involves projecting onto the columns of the spectral matrix, creating a candidate subset of variable with minimal collinearity. The subsequent step assesses these candidate subsets based on the RMSE values obtained from the validation set of the PLSR calibration. The final step involves removing uninformative variables through a variable elimination procedure that does not significantly compromise predictive ability [42–44].

In this study, five different PLSR-based models were constructed: Full-PLSR, MP PLSR 5-range, MP PLSR 3-range, GA-PLSR, and SPA-PLSR. These models were developed using an in-house code in MATLAB–R2020b (MathWorks, Natick, MA, USA) and their performances were subsequently compared.

The models were evaluated based on R^2_c , RMSEC, R^2_p , RMSEP, RPD, and bias. These parameters were calculated using Equations (2)–(5).

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

$$\text{RMSEC}, \text{RMSEP} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}, \quad (3)$$

$$\text{RPD} = \frac{\text{SD}}{\text{RMSEP}}, \quad (4)$$

$$\text{Bias} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}, \quad (5)$$

where y is the measured value, \hat{y} is the predicted value, i is the subscript used to indicate the number of the sample, \bar{y} is the mean of the measured value, N is the number of samples of the respective set, SD is the standard deviation of the measured values of the validation set, and n is the number of samples in the validation set.

The optimum model is selected based on higher R^2_c , R^2_p , and RPD values, as well as lower RMSEC and RMSEP values. In the case of similar performance, the model with fewer LVs is selected

as an alternative method for rapid non-destructive evaluation of biomass proximate analysis parameters [26].

Williams et al. (2019) [45] and Zornoza et al. (2008) [46] guidelines have been followed for the interpretation of R^2 and RPD values, respectively.

6.2.7 Limit of Quantification (LOQ)

The sensitivity of the selected high-performance PLSR model is a critical factor for reliably and accurately determining biomass proximate analysis parameters. To address this, the LOQ of the selected model is calculated, representing the minimum concentration of the analyte that can be detected and quantified with high sensitivity and reliability. There are different methods to calculate the LOQ, which is based on (1) visual evaluation, (2) signal-to-noise, and (3) the SD of the response to slope [47]. In this study, LOQ is calculated based on the SD of the response to slope method obtained from the calibration set. The LOQ is calculated using the following equation:

$$\text{LOQ} = 10 \frac{\sigma_C}{S_C}, \quad (6)$$

6.3 Results and Discussion

6.3.1 Spectral Preprocessing

Figure 6.7 displays the average pretreated spectrum obtained from the best-performing model for proximate analysis of chip or ground biomass. This model utilizes the following techniques: (a) MC with the second derivative from chip biomass, (b) VM with a MP combination set of 4, 3, 4, 6, and 0 from chip biomass, (c) FC with a constant offset from chip biomass, and (d)

ash content with the first derivative from the ground biomass.

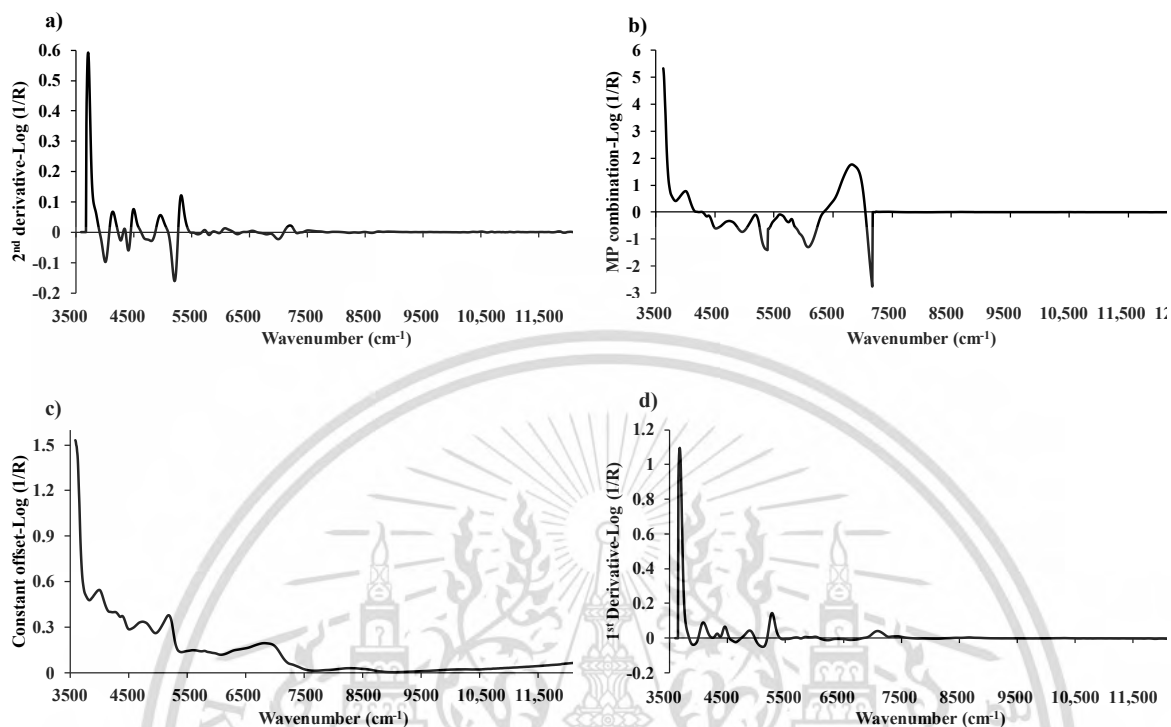


Figure 6.7 The average preprocessed spectra from the best performing model for: (a) MC with the second derivative, (b) VM with a multi-preprocessing combination set of 4, 4, 3, 2, and 0, (c) FC with a constant offset, and (d) ash content with the first derivative.

6.3.2 Reference Data

The proximate analysis data for ten different varieties of biomass, including four wood varieties of fast-growing trees and six non-wood varieties (five agricultural residues and one fast-growing tree) were obtained by analyzing the TG and DTG profiles obtained from TGA, as explained in Section 2.3. Figure 6.8 displays the comprehensive results of proximate analysis for biomass, encompassing MC, VM, FC, and ash content, all of which were obtained from TGA.

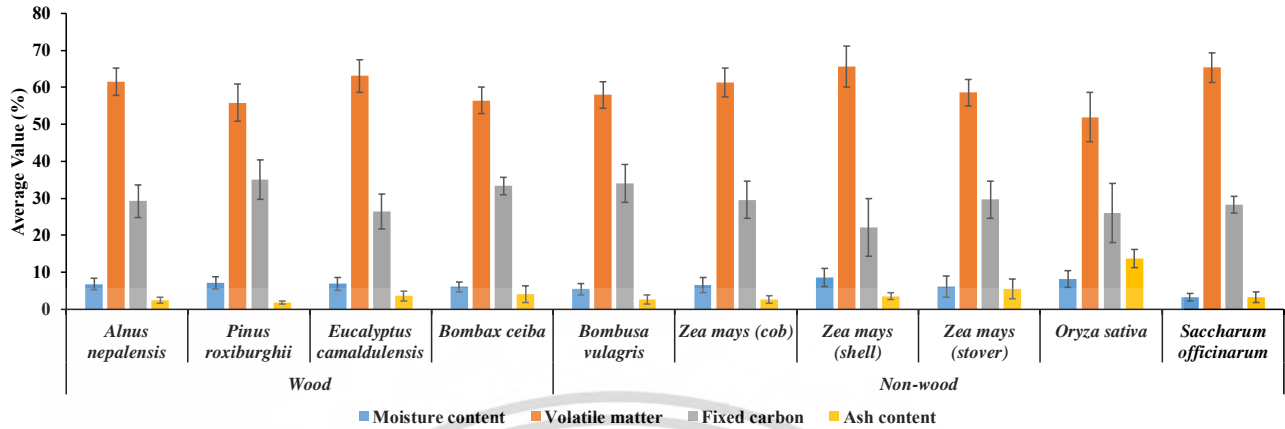


Figure 6.8 Overall proximate analysis results of ten different biomass varieties, including four wood varieties of fast-growing trees and six non-wood varieties (five agricultural residues and one fast-growing tree) obtained from the TGA analysis.

Under the TGA, as explained in Section 2.3, the average MC, VM, FC, and ash content in the biomass samples were recorded as 6.47%, 58.99%, 31.63%, and 2.91%, respectively, and 6.55%, 60.57%, 27.17%, and 5.72%, respectively. Here, FC is calculated as:

$$FC (\%) = 100 - MC (\%) - VM (\%) - \text{ash} (\%), \quad (7)$$

The analysis revealed notable variations in MC, VM, FC, and ash content among the studied biomass varieties. The maximum MC of 8.61% and VM of 65.64% were recorded for *Zea mays* (shell), whereas the minimum MC of 3.19% was found in *Saccharum officinarum*, and the minimum VM of 51.96% was observed in *Oryza sativa*. Additionally, *Pinus roxburghii* exhibited the highest FC content at 35.11%, while *Zea mays* (shell) displayed the lowest FC content of 22.21%. Similarly, *Oryza sativa* had the maximum ash content of 13.72%, and *Pinus roxburghii* had the lowest ash content at 1.82%.

The MC, VM, FC, and ash content in fast-growing trees range from 3.50% to 10.39%, 48.31% to 69.74%, 18.95% to 42.05%, and 0.32% to 7.90%, respectively. For agricultural residues, the respective ranges are 2.21–12.21% (MC), 44.62–75.10% (VM), 10.40–35.83% (FC), and 0.88–17.59% (ash content). Interestingly, there seems to be an overlap in the ranges. This was confirmed by the data in Table 5.7 and scatter plots of the developed models shown in Figure 6.9, which indicated that the proximate parameter spans were wider when the wood and non-wood biomass of fast-growing trees and agricultural residue were combined for modeling.

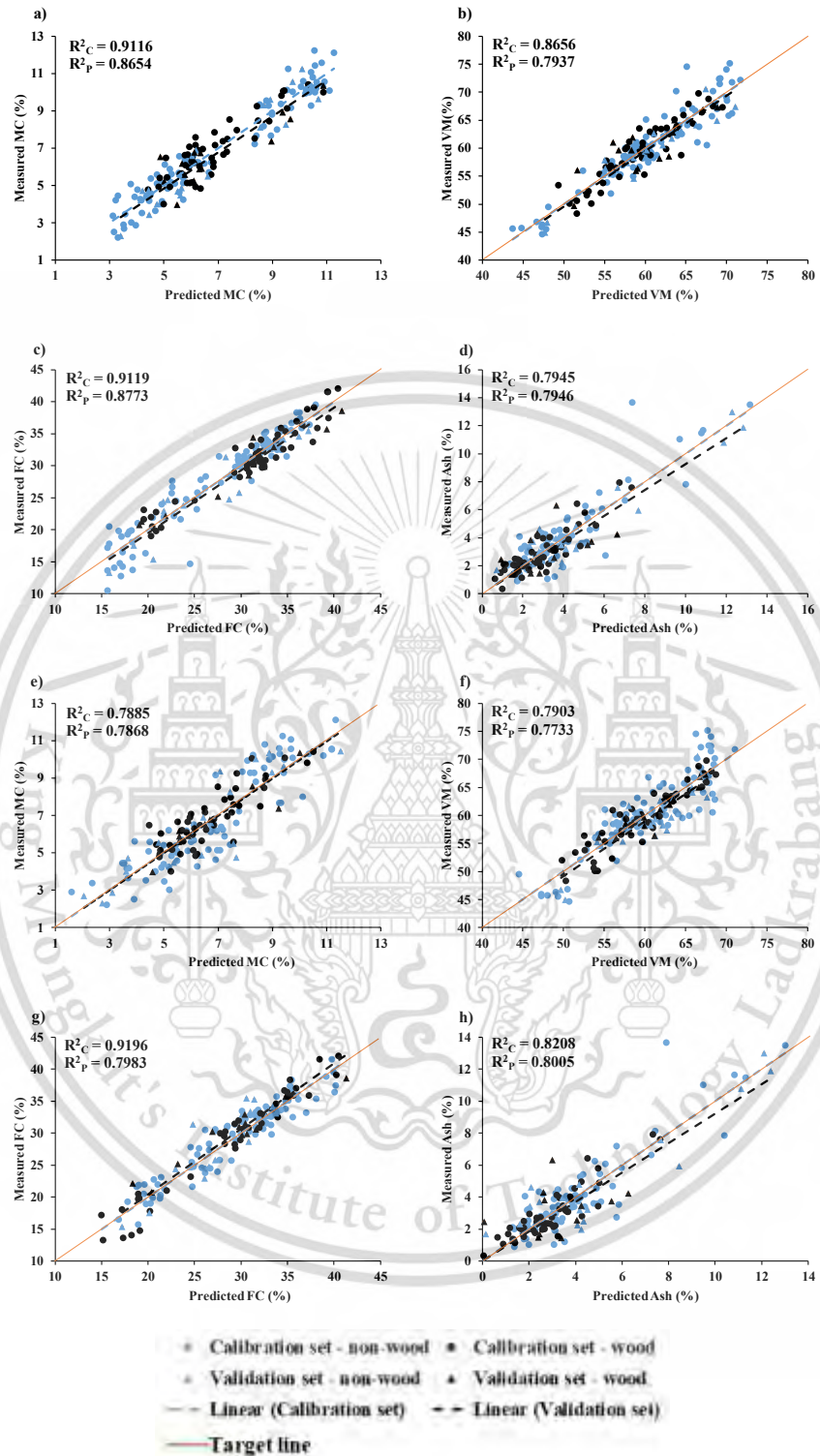


Figure 6.9 Measured versus predicted value in calibration set and validation set for chip biomass (a) MC, (b) VM, (c) FC, and (d) ash and for ground biomass (e) MC, (f) VM, (g) FC, and (h) ash.

The observed variations in MC among the biomass varieties are of significance. Despite all MC values falling below 10%, the notably higher MC in *Zea mays* (shell) highlights the necessity for effective drying procedures prior to utilizing it for energy purposes. Achieving an optimal MC could enhance the combustion efficiency of this biomass [48].

The diverse levels of VM content are indicative of differing combustion behaviors. Biomass with higher VM content tend to ignite more readily, making them favorable for certain combustion-based applications [10]. The presence of both high and low VM biomass, such as *Zea mays* (shell) and *Oryza sativa*, respectively, provides opportunities for specific use in different energy production processes.

The FC content is a crucial factor influencing combustion efficiency. Biomass with higher FC content generally possess better energy density and combustion characteristics [49]. In light of this, *Pinus roxburghii* contained notably higher FC content and is positioned as a promising biomass for energy production, particularly in scenarios where concerns related to ash generation and deposition need to be minimized.

Furthermore, ash content is a critical consideration in combustion processes. Lower ash content, such as in *Pinus roxburghii*, can alleviate operational challenges associated with ash accumulation. Conversely, *Oryza sativa* had higher ash content and might require careful management or additional treatment to mitigate potential issues during combustion.

6.3.3 Prediction of Proximate Analysis Data

Table 6.1 shows the statistical summary data for the proximate parameters: MC, VM, FC, and ash. These parameters were employed in both the calibration and validation sets to develop a PLSR-based model for both chipped and ground biomass, using the full wavenumber range from 3594.87 to 12,489.48 cm^{-1} to assess proximate analysis parameters. Table 6.2 shows the details of outliers identified in chip and ground biomass, which were subsequently removed from the total dataset for PLSR model development. Table 6.3 shows PLSR modeling results for chipped biomass, and Table 6.4 shows the results for ground biomass, with the bolded model showing the best performance in each parameter.

6.3.3.1 Moisture Content (%)

Table 6.2 shows the count of outliers identified for MC in chip (based on reference data) and ground biomass (based on both reference and spectroscopic data). After the removal of these outliers, PLSR-based models were developed to evaluate the percentage of MC in chip and ground biomass. The model performance was assessed across various preprocessing techniques, and the best-performing model, selected as the optimal choice among five different types of PLSR-based models, was highlighted in bold in both Tables 6.3 and 6.4.

Table 6.1 Statistical data of the proximate analysis parameters of the chip and ground biomass used in the PLSR model development.

Category	Parameters (%)	N _T	Calibration Set					Validation Set				
			N _C	Max	Min	Mean	SD	N _p	Max	Min	Mean	SD
Chip	MC	173	138	12.21	2.21	6.83	2.37	35	11.24	2.29	6.81	2.35
	VM	173	138	75.10	44.62	60.12	6.39	35	70.59	44.84	59.42	5.52
	FC	173	138	42.05	10.40	28.91	7.06	35	38.55	13.23	28.32	6.15
	Ash	162	130	13.64	0.32	3.53	2.34	32	12.97	1.37	4.17	2.90
Ground	MC	157	126	12.10	2.21	6.72	2.26	31	10.42	2.29	6.49	2.37
	VM	174	139	75.10	44.62	60.11	6.37	35	70.59	44.84	59.42	5.52
	FC	152	122	42.05	13.23	29.08	6.71	30	38.55	15.33	29.11	5.63
	Ash	163	130	13.64	0.32	3.56	2.35	33	12.97	1.37	4.15	2.86

Table 6.2 Details on the number of identified outliers in chip and ground biomass evaluated before the development of the PLSR model.

Proximate Parameter (%)	Chip Biomass		Ground Biomass	
	Total Number of Sample	Outliers	Total Number of Sample	Outliers
MC	173	0	174	17
VM	173	0	174	0
FC	173	0	174	22
Ash	173	11	174	11

In Figure 6.9a, the scatter plot displays the measured and predicted MC (%) values for wood and non-wood samples in both the calibration and validation sets, using GA-PLSR. The best performance was achieved with GA-PLSR utilizing 10 LVs, D2 spectral pretreatment, and 57 important wavenumbers (Table 6.3). A notable 2.82% improvement in explained variance from Full-PLSR ($R^2_p = 0.8417$) to GA-PLSR ($R^2_p = 0.8654$) was observed. This improvement can be attributed to GA-PLSR efficient feature selection and model parameter optimizations, resulting in a more accurate and precise predictive model. The increased explained variance indicates a stronger correlation between the spectral data (independent variable) and the reference data (dependent variable), ultimately enhancing the predictive capabilities of the model.

Similarly, Figure 6.9e displays a scatter plot depicting the measured and predicted MC (%) of ground biomass using mean centering. The Full-PLSR model, employing 12 LVs and mean centering as spectral preprocessing, exhibited the best performance (Table 6.4). An enhancement of 19.72% in explained variance was observed, transitioning from Full-PLSR using raw spectra ($R^2_p = 0.6752$) to Full-PLSR employing mean centering ($R^2_p = 0.7868$). Furthermore, the use of mean centering indicates that the model accuracy can be improved by removing baseline shifts, aligning spectral data around a common central point, and mitigating systematic variations caused by instrumental effects such as intensity fluctuations.

Table 6.3 Results of the PLS regression-based model for the proximate analysis (%) of chip biomass, with the bolded model showing the best performance.

Parameter	Algorithm	Preprocessing	LVs	Calibration Set		Validation Set			
				R^2_c	RMSEC (%)	R^2_p	RMSEP (%)	RPD	Bias (%)
MC	Full-PLSR	D1	14	0.9498	0.53	0.8417	0.92	2.5	-0.14
	SPA-PLSR	Constant offset (SW: 100)	12	0.8587	0.89	0.8281	0.96	2.4	-0.02
	GA-PLSR	D2 (SW:57)	10	0.9116	0.70	0.8654	0.85	2.9	-0.25
	MP-PLSR-5 range	Combination set: 4, 3, 4, 6, 0	13	0.9280	0.63	0.8260	0.97	2.4	-0.13
	MP-PLSR-3 range	Combination set: 4, 6, 0	11	0.9194	0.67	0.8386	0.93	2.5	-0.13
VM	Full-PLSR	Constant offset	15	0.8719	2.28	0.7725	2.60	2.1	-0.56
	SPA-PLSR	Mean Centering (SW: 971)	12	0.8345	2.59	0.7720	2.60	2.2	-0.61
	GA-PLSR	D1 (SW: 423)	9	0.8460	2.50	0.7801	2.55	2.2	-0.44
	MP-PLSR-5 range	Combination set: 4, 4, 3, 2, 0	15	0.8656	2.33	0.7937	2.47	2.2	-0.48
	MP-PLSR-3 range	Combination set: 4, 4, 0	9	0.8368	2.57	0.7504	2.72	2.0	-0.46
FC	Full-PLSR	Constant offset	15	0.9119	2.09	0.8773	2.12	3.0	-0.64
	SPA-PLSR	Constant offset (SW: 711)	12	0.8831	2.40	0.8759	2.14	3.0	-0.74
	GA-PLSR	Constant offset (SW: 412)	15	0.9081	2.13	0.8762	2.13	3.0	-0.72
	MP-PLSR-5 range	Combination set: 4, 4, 4	10	0.9060	2.16	0.8432	2.40	2.7	-0.77
	MP-PLSR-3 range	Combination set: 5, 1, 6, 6, 6	13	0.9008	2.21	0.8680	2.20	2.9	-0.66
Ash	Full-PLSR	D2	12	0.8074	1.02	0.7596	1.40	2.1	-0.20
	SPA-PLSR	D1 (SW: 458)	14	0.7554	1.15	0.7544	1.42	2.0	-0.21
	GA-PLSR	D1 (SW: 175)	13	0.7805	1.09	0.7638	1.39	2.1	-0.17
	MP-PLSR-5 range	Combination set: 5, 6, 5, 0, 0	15	0.7645	1.13	0.7648	1.39	2.1	-0.27
	MP-PLSR-3 range	Combination set: 5, 4, 4	13	0.7945	1.06	0.7946	1.29	2.2	-0.22

Figure 6.10 displays the average absorbance spectra after undergoing D2 pretreatment, with 57 significant wavenumbers highlighted in red to emphasize their importance as identified

through GA. The highest peaks were observed at 3722, 4525, 5000, and 5285 cm^{-1} , all of which could have the potential to enhance the model performance derived from GA-PLSR for MC (%).

Similarly, Figure 6.11 illustrates the regression coefficients for the MC (%) of ground biomass using Full-PLSR with spectra preprocessed through mean centering. Significant peaks were identified at 3650, 4902, 7042, and 8163 cm^{-1} , which could have a substantial impact on enhancing the model performance.



Table 6.4 Results of the PLS regression-based model for the proximate analysis (%) of ground biomass, with the bolded model showing the best performance.

Parameter	Algorithm	Preprocessing	LVs	Calibration Set		Validation Set			Bias (%)
				R^2_c	RMSEC (%)	R^2_p	RMSEP (%)	RPD	
MC	Full-PLSR	Mean Centering	12	0.7885	1.03	0.7868	1.08	2.2	0.01
	SPA-PLSR	D1 + vector normalization (SW: 125)	15	0.7754	1.07	0.7624	1.14	2.1	0.05
	GA-PLSR	Constant offset (SW: 925)	13	0.8083	0.98	0.7582	1.15	2.0	-0.07
	MP-PLSR-5 range	Combination set: 5, 4, 0, 4, 6	12	0.8338	0.92	0.7727	1.11	2.1	0.18
	MP-PLSR-3 range	Combination set: 4, 4, 1	15	0.8854	0.76	0.7123	1.25	1.9	0.00
VM	Full-PLSR	SNV	15	0.8363	2.57	0.7557	2.69	2.2	-1.17
	SPA-PLSR	SNV (SW: 931)	14	0.8213	2.68	0.7513	2.71	2.2	-1.18
	GA-PLSR	SNV (SW: 383)	9	0.7382	3.25	0.7305	2.83	2.1	-1.24
	MP-PLSR-5 range	Combination set: 3, 5, 4, 5, 0	9	0.7903	2.91	0.7733	2.59	2.3	-1.09
	MP-PLSR-3 range	Combination set: 4, 2, 6	14	0.8245	2.66	0.7323	2.82	2.2	-1.32
FC	Full-PLSR	D1	14	0.9196	1.90	0.7983	2.49	2.3	0.67
	SPA-PLSR	D1 (SW: 945)	14	0.9139	1.96	0.7965	2.50	2.3	0.64
	GA-PLSR	D1 + vector normalization (SW: 74)	14	0.8465	2.62	0.7034	3.02	1.9	0.85
	MP-PLSR-5 range	Combination set: 1, 4, 1, 4, 1	13	0.8743	2.37	0.7932	2.52	2.2	0.19
	MP-PLSR-3 range	Combination set: 6, 5, 1	12	0.8217	2.82	0.7732	2.64	2.1	0.20
Ash	Full-PLSR	D1	13	0.8208	0.99	0.8005	1.26	2.2	0.04
	SPA-PLSR	D1 (SW: 1089)	13	0.8216	0.99	0.8005	1.26	2.2	0.03
	GA-PLSR	SNV (SW: 112)	15	0.7973	1.05	0.7834	1.31	2.2	0.20
	MP-PLSR-5 range	Combination set: 4, 6, 3, 5, 3	15	0.8818	0.81	0.7971	1.27	2.2	0.14
	MP-PLSR-3 range	Combination set: 4, 6, 2	11	0.7765	1.11	0.7705	1.35	2.1	0.07

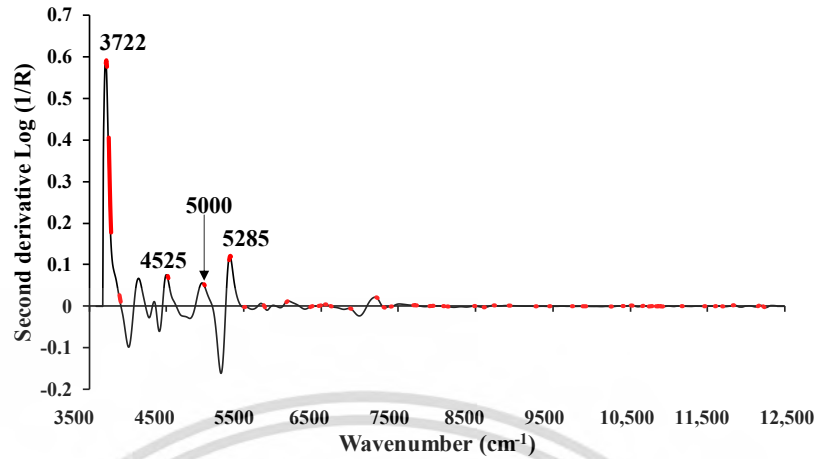


Figure 6.10 The average absorbance value of MC (%) of chip biomass obtained using second derivative preprocessing with a selection of important wavenumbers obtained from GA.

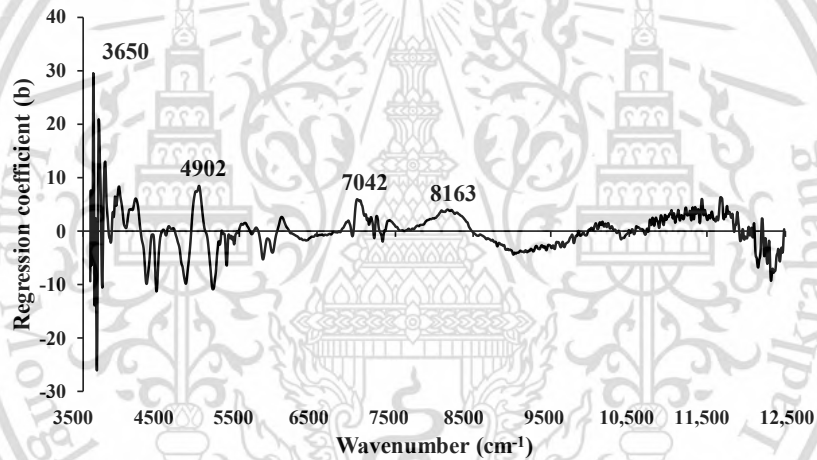


Figure 6.11 The regression coefficient for the MC (%) of ground biomass using the Full-PLSR with spectral preprocessing of mean centering.

Table 6.5 [50] tabulates the bond vibration and corresponding functional group, spectral-structure, and material type which were found in the featured wavelength plot and regression coefficient plots by near-infrared spectroscopy analysis of chip and ground biomass for proximate analysis parameters which are derived from the best performing models in this study. Figures 10, 12, 14 and 16 are for chip models and Figures 11, 13, 15 and 17 are for ground biomass models for MC, VM, FC, and ash, respectively.

6.3.3.2 Volatile Matter (%)

For VM, no outliers were identified in either chip or ground biomass (Table 6.3). Therefore, models were developed as explained in Section 2.6. For both chip and ground biomass, the PLSR model developed using the spectral MP 5-range methods yielded the highest performance.

For chip biomass, the best result was obtained with a spectral preprocessing combination set of 4, 4, 3, 2, 0, i.e., D1, D1, MSC, SNV, and zero, respectively, from the range 3625.72–12,489.48 cm^{-1} , which is equally divided into five sections. Figure 6.9b displays the scatter plot of measured versus predicted percentages of VM in wood and non-wood samples for the calibration and validation sets. When compared with the best model performance (Table 6.3) obtained from Full-PLSR using a constant offset (R^2_p of 0.7725), the MP 5-range PLSR method resulted in a 2.7443% improvement in explained variance.



Table 6.5 The dominant peaks on the regression coefficient plot obtained from the best-performing PLSR-based model.

Proximate Parameter	Biomass Type	Peak Wave-length (cm ⁻¹)	Functional Group	Spectra-Structure	Material Type	Reference
MC	Chip	3722	C-H aromatic C-H (aryl)		C-H aryl	
		4525	N-H ammonia in water	N-H (3V) for NH ₃ (ammonia) in water	Ammonia in water	
		5000	N-H ammonia in water		Ammonia	
		5285	O-H hydrogen bonding between water and exposed polyvinyl alcohol OH		Water and polyvinyl alcohol OH	[50]
	Ground	3650	O-H from primary alcohols as (-CH-OH)	O-H (V)(-CH ₂ -OH), Primary alcohols	Primary alcohols	
		4902	N-H combination band from urea	N-H for primary amides	Urea	
		7042	O-H aromatic (ArO-H)	O-H (2V), O-H	Hydrocarbons, aromatic	
		8163	C-H secondary or tertiary carbon (.CH)	C-H (3V), C-H	Hydrocarbons, aliphatic	
VM	Chip	4019	C-H/C=C (C-H and C-C)	C-H stretching and C-C stretching combination	Cellulose	
		5495	O-H/C-H combination	O-H stretching and C-O stretching combination	Cellulose	
		5865	C-H methyl (.CH ₃)	C-H methyl (.CH ₃)	Hydrocarbons, methyl	
		6944	C-H methylene (.CH ₂)	C-H combination, C-H ₂	Hydrocarbons, aliphatic	
	Ground	3693	C-H aromatic C-H (aryl)		C-H aryl	[50]
4505		N-H combination band from urea (NH ₂ -C=O-NH ₂)	N-H combination	N-H from urea		
5200		O-H assigned to molecular water [O-H (O-H O-H stretching and HOH deformation combination and HOH)]		O-H molecular water		
5735		C-H methyl C-H, aromatic associated (ArCH ₃)	C-H methyl, aromatic (ArCH ₃)	Aromatic (ArCH ₃)		
		5952	C-H aromatic (ArCH)	C-H (2V), aromatic C-H	Hydrocarbons, aromatic	
FC	Chip	3722	C-H aromatic C-H (aryl)		C-H aryl	
		4405	O-H/C-H cellulose (.OH and .C-O)	O-H stretching and C-O stretching combination	Cellulose	
		5952	C-H aromatic (ArCH)	C-H (2V), aromatic C-H	Hydrocarbons, aromatic	
		8754	C-H aromatic (ArCH)	C-H (3V), aromatic C-H	Hydrocarbons, aromatic	[50]
	Ground	3733	C-H aromatic C-H (aryl)		C-H aryl	
4099		C-H aromatic C-H (aryl)		C-H aryl		
5903		C-H methyl C-H, (CH ₃) (Asymmetric)	C-H (2V), methyl	Hydrocarbons, aliphatic		
11,655		C-H aromatic (ArCH)	C-H (4V), aromatic C-H	Hydrocarbons, aromatic		
Ash	Chip	4019	C-H/C=C (C-H and C-C)	C-H stretching and C-C stretching combination	Cellulose	
		5285	O-H hydrogen bonding between water and exposed polyvinyl alcohol OH		Water and polyvinyl alcohol OH	
		11,655	C-H aromatic (ArCH)	C-H (4V), aromatic C-H	Hydrocarbons, aromatic	
		12,048	C-H methylene C-H, associated with linear aliphatic R(CH ₂) _n R	C-H combination	Hydrocarbons, aliphatic	
	Ground	4505	N-H combination band from urea (NH ₂ -C=O-NH ₂)	N-H combination	N-H from urea	
		5285	O-H hydrogen bonding between water and exposed polyvinyl alcohol OH		Water and polyvinyl alcohol OH	
		11,655	C-H aromatic (ArCH)	C-H (4V), aromatic C-H	Hydrocarbons, aromatic	
		12,300	C-H methylene C-H, associated with branched aliphatic RC(CH ₃) ₃	C-H combination	Hydrocarbons, aliphatic	

V : Fundamental stretching vibrational absorption band, 2V : First overtone of fundamental

stretching band, 3ν : Second overtone of fundamental stretching band, 4ν : Third overtone of fundamental stretching band.

Figure 6.12 shows the regression coefficient plot for the % of VM in chip biomass, derived from the MP PLSR 5-range method. The important peaks were observed at the wavenumbers 4019, 5495, 5865, and 6944 cm^{-1} (Table 6.5), which could significantly influence the improved performance of the selected model.

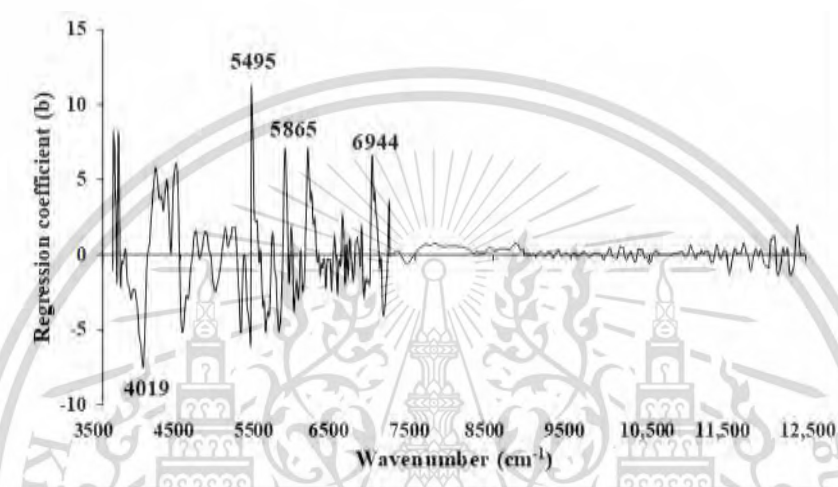


Figure 6.12 Regression coefficient for the % of VM in chip biomass using the MP PLSR 5-range method with a spectral multi-preprocessing combination set of 4, 4, 3, 2, and 0.

The model achieved its best performance for VM (%) in the ground biomass with a spectral preprocessing combination set of 3, 5, 4, 5, and 0, i.e., MSC, D2, D1, D2, and zero absorbance—in the range of 3625.72–12,489.48 cm^{-1} , equally divided into five sections. This MP PLSR 5-range resulted in an R^2_p of 0.7733 (Table 6.4), signifying a notable improvement of 2.3290% compared to Full-PLSR ($R^2_p = 0.7557$).

As shown in Figure 6.13, significant peaks were observed at wavenumbers 3693, 4505, 5200, 5735, and 5952 cm^{-1} (Table 6.5), all of which could exert a notable influence on the enhanced performance of the model using the MP PLSR 5-range method. Within the wavenumber range between 10,722.9 and 12,489.48 cm^{-1} , the regression coefficient was observed to be zero. This suggests that the range lacks relevant spectral information for predicting VM in ground biomass.

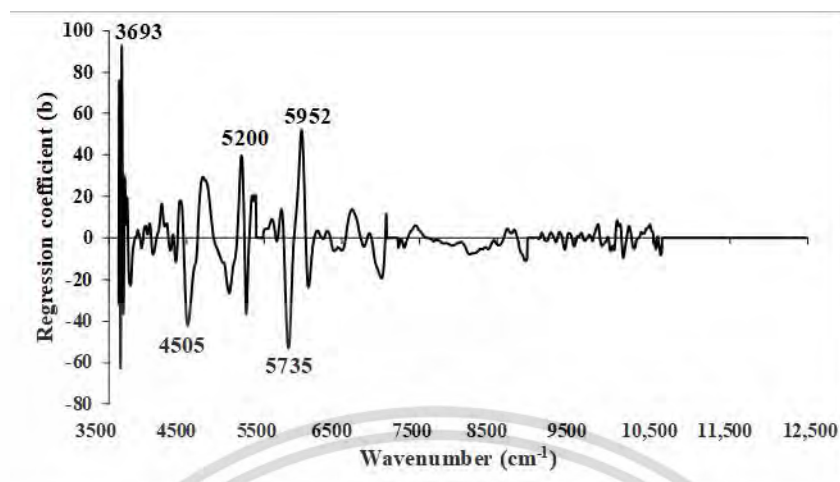


Figure 6.13 Regression coefficient for the % of VM in ground biomass using the MP PLSR 5-range method with a spectral multi-preprocessing combination set of 3, 5, 4, 5, and 0.

6.3.3.3 Fixed Carbon (%)

Table 6.2 presents the count of outliers for FC in chip and ground biomass. No outliers were detected in the chip biomass dataset. However, using the reference and spectroscopic data, 22 ground biomass samples were identified as outliers. These identified outliers were subsequently removed from the total ground biomass dataset, after which PLSR models were developed following the procedures outlined in Section 2.6.

The model developed using Full-PLSR, constant offset spectral preprocessing, and 15 LVs yielded better results for FC (%) in chip biomass (Table 6.3). These results also indicate that model accuracy can be improved by correcting the baseline shift in the raw spectra using a constant offset. Figure 6.9c displays the scatter plot of measured versus predicted % of FC for non-wood and wood samples in the calibration and validation set, obtained from Full-PLSR with constant offset preprocessing.

Figure 6.14 displays the regression coefficient plot for the % of FC in chip biomass using Full-PLSR with spectral preprocessing of constant offset. Significant peaks were observed at 3722, 4405, 5952, and 8754 cm^{-1} (Table 6.5), which are likely to contribute significantly to the enhancement of model performance.

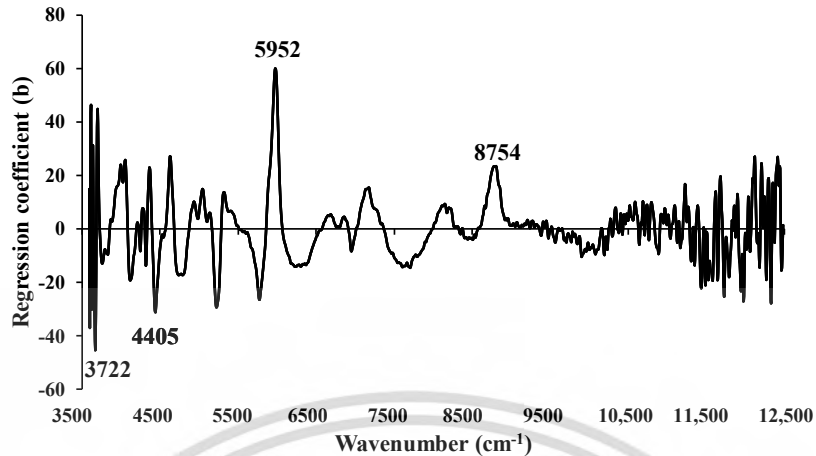


Figure 6.14 Regression coefficient for the % of FC in chip biomass using the Full-PLSR method with a spectral preprocessing of constant offset.

The Full-PLSR model with spectral preprocessing using the D1 provides improved model performance for % of FC in the ground biomass (Table 6.4). These results indicate that the reduction in linear shift and enhancement of the peaks in the raw spectra, achieved by applying the D1 with a gap and segment of 5, each contribute to the improvement in model accuracy.

Figure 6.15 presents the regression coefficient plot for the % of FC in ground biomass using Full-PLSR with D1 spectral preprocessing. Significant peaks were observed at 3733, 4099, 5903, and 11,655 cm^{-1} (Table 6.5), each of which could potentially exert an influence on the model improved performance.

6.3.3.4 Ash Content (%)

Table 6.2 presents the count of outliers for ash content in chip and ground biomass, based on reference and spectroscopic data. These identified outliers were removed from the total dataset, and various PLSR-based models were developed as explained in Section 2.6. For chip biomass, the best-performing model was developed using the MP PLSR 3-range methods with multi-preprocessing combination set of 5, 4, and 4. For ground biomass, the Full-PLSR model with spectral preprocessing from D1 provided the best performance results.

Figure 6.9d shows a scatter plot comparing the measured and predicted % ash content of non-wood and wood samples within the calibration and validation datasets. These results were obtained using the MP PLSR 3-range method. The most favorable outcomes were achieved through a spectral preprocessing combination set, consisting of the D2 from 3594.87 to 5492.59

cm⁻¹, the D1 from 7498.31 to 5500.30 cm⁻¹, and the D1 from 7506.02 to 12,489.48 cm⁻¹, respectively (Table 6.3). The MP PLSR 3-range method enhances the explained variance by 4.6077%, while the Full-PLSR has an R²_p of 0.7596.

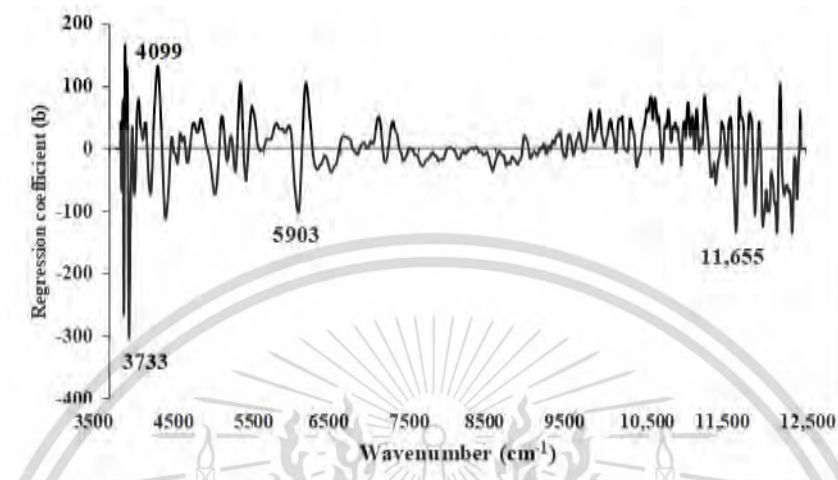


Figure 6.15 Regression coefficient for the % of FC in ground biomass using Full-PLSR method with a spectral preprocessing of first derivative.

Figure 6.16 depicts the regression coefficient plot for the % ash content in chip biomass using the multi-preprocessing PLSR 3-range method, incorporating a combination set of 5, 4, and 4. The plot reveals significant peaks at 4019, 5285, 11,655, and 12,048 cm⁻¹ (Table 6.5). These peaks are expected to have a substantial impact on achieving the highest-performing model for assessing the % ash content in chip biomass.

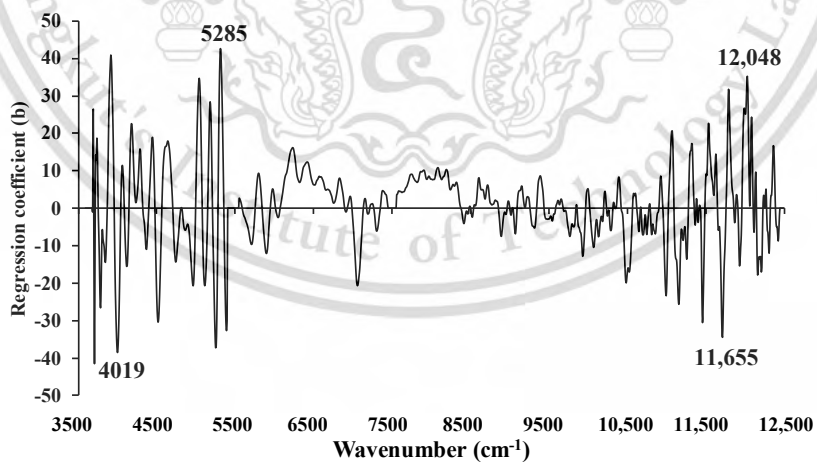


Figure 6.16 Regression coefficient for the % of ash content in chip biomass using the MP PLSR 3-range method with a spectral multi-preprocessing combination set of 5, 4, and 4.

The best performance for % ash content in the ground biomass was achieved by using Full-PLSR with spectral preprocessing involving the D1, along with 13 LVs (Table 6.4). The results

also indicate that removing baseline shift, resolving overlapping peaks, and highlighting the detailed structure in raw spectra will contribute to improving the model accuracy. In Figure 6.9h, a scatter plot illustrates the relationship between measured and predicted ash content percentages of non-wood and wood samples in both the calibration and validation sets, demonstrating the model high performance.

Figure 6.17 displays the regression coefficients for the % of ash content in the ground biomass using Full-PLSR with a spectral preprocessing technique involving the D1 (segment = 5 and gap = 5). The crucial peaks that may significantly impact the model performance are observed at wavenumbers 4505, 5285, 11,655, and 12,300 cm^{-1} (Table 6.5).

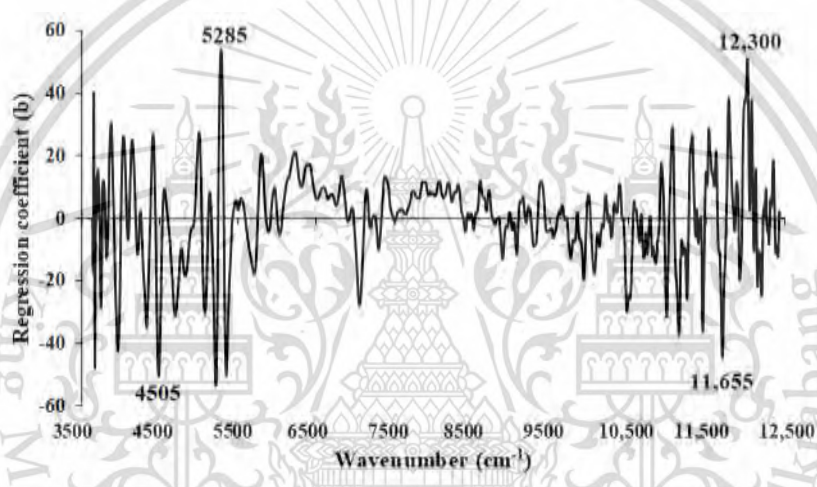


Figure 6.17 Regression coefficient for the % of ash content in ground biomass using Full-PLSR method with a spectral preprocessing of D1.

6.3.4 Limit of Quantification (LOQ)

In this study, the LOQ, defined as the lowest possible concentration of analyte that can be reliably detected and quantified with an acceptable level of accuracy and precision, is determined using the SD of the response and the slope obtained from the calibration set of the selected model [31]. Table 6.6 presents the LOQ results of the PLSR models for the percentages of MC, VM, FC, and ash content in both chip and ground biomass.

The LOQ for VM (%) in both chip and ground biomass, as well as for FC (%) in chip biomass, is below the minimum reference value used in the modeling. This suggests that the selected model for VM and FC in chip biomass has the potential to reliably detect and precisely quantify these parameters. This indicates its high sensitivity. For instance, the chosen MP method (PLSR 5-range) can reliably detect VM in both chip and ground biomass, as indicated by the model LOQ

values, i.e., 23.4294% and 29.1656%, respectively. Similarly, using Full-PLSR with constant offset preprocessing, FC in chip biomass can be reliably detected with an LOQ value of 6.9125%.

Table 6.6 LOQ results of PLSR models for the % of MC, VM, FC, and ash content in chip and ground biomass.

Parameter (%)	Category	Model	No. of Variables	Preprocessing	Reference Range	SD	Slope	LOQ (%)
MC	Chip	GA-PLSR	57	Second derivative	12.21–2.21	0.71	1	7.06
	Ground	Full-PLSR	1154	Mean centering	12.10–2.21	1.08	1	10.38
VM	Chip	MP-PLSR-5 range	1150	Combination set: 4, 4, 3, 2, 0	75.10–44.62	2.34	1	23.43
	Ground	MP-PLSR-5 range	1150	Combination set: 3, 5, 4, 5, 0	75.10–44.62	2.92	1	29.17
FC	Chip	Full-PLSR	1154	Constant offset	42.05–10.40	2.09	1	6.91
	Ground	Full-PLSR	1154	First derivative	42.05–13.23	1.90	1	19.04
Ash	Chip	MP-PLSR-3 range	1154	Combination set: 5, 4, 4	13.64–0.32	1.06	1	10.62
	Ground	Full-PLSR	1154	First derivative	12.64–0.32	1.03	1	10.25

However, for other parameters, such as MC, ash, and FC in ground biomass, the LOQ values exceed the minimum reference value used in model development. This indicates that the selected models have limitations in assessing these parameters for not more than the minimum value of the model development. In Table 6.6, the slope of every model was the same but the SD of MC, ash, and FC was lower than VM, especially, for MC, the SD was very low leading to the high value of LOQ. Although the MC and FC were the constituents which were good absorbers of NIR radiation, in the case of LOQ calculation, this was not concerned with the absorption of radiation. However, the slope of the model and the SD of the data of the calibration set constituents must be high (wide range). Additionally, alternative modeling methods are considered for predicting proximate parameters, especially when their LOQ values surpass the minimum reference value used in modeling, which can enhance sensitivity as illustrated by the NIR model LOQ.

6.3.5 Effect of Combined Wood Samples with Non-Wood Samples in Developed Models

Table 6.7 presents the range of reference values for the % of MC, VM, FC, and ash content found in both wood and non-wood biomass samples within the calibration set and validation set. Based on Figure 6.9 and Table 6.7, the non-wood samples are characterized by a wider range in the MC, VM, FC, and ash content in both calibration and prediction sets. It becomes apparent that the range of each proximate parameter, whether in chip and ground form, widens after the inclusion of non-wood samples alongside the wood samples. This expansion aims to establish a more resilient PLSR model for predicting proximate parameters.

In Figure 6.9a,b, the range of MC (%) and VM (%) in wood chip biomass samples was narrower than that of non-wood samples, with the latter extending to both higher and lower percentages. Figure 6.9c illustrates that the range of FC values for wood samples was higher and narrower compared to that of the non-wood samples. The inclusion of non-wood samples expanded the range, mostly towards the lower percentage values. In contrast, Figure 6.9d, depicting ash content in chip biomass, reveals a lower and narrower range for both wood and non-wood samples compared to other proximate parameters. Consequently, the performance of the optimal model was comparatively lower for ash content in comparison to MC, VM, and FC.

Table 6.7 The range of % of MC, VM, FC, and ash content of wood and non-wood samples in calibration and validation sets.

Biomass	Parameter (%)	Calibration Set		Validation Set	
		Wood	Non-Wood	Wood	Non-Wood
Chip	MC	3.99–10.39	2.21–12.21	3.97–10.35	2.29–11.24
	VM	48.31–69.74	44.62–75.10	49.63–67.29	44.84–70.59
	FC	18.95–42.05	10.4–39.55	20.67–38.55	13.23–35.46
	Ash	0.32–7.90	0.88–13.64	1.37–6.29	1.68–12.97
Ground	MC	3.99–10.39	2.21–12.10	3.97–10.35	2.29–10.42
	VM	48.31–69.74	44.62–75.10	49.63–67.29	44.84–70.59
	FC	13.23–42.05	15.69–41.48	20.67–38.55	15.33–35.46
	Ash	0.32–7.90	0.88–13.64	1.37–6.29	1.68–12.97

Similarly, in ground biomass, the range of non-wood samples is higher compared to wood samples, except for FC. Similar to chip biomass, the range of MC (%) (Figure 6.9e) and VM (%) (Figure 6.9f) values for wood samples in ground biomass is narrower compared to that of the non-wood samples. Therefore, the inclusion of non-wood samples contributes to expanding the range to both higher and lower percentages. In Figure 6.9g, the percentage range of FC in wood samples is higher and wider compared to that of the non-wood samples. The inclusion of both wood and non-wood samples with a wider range has contributed to a better performance of the model compared to other parameters. From Figure 6.9h, the ash content (%) range in wood samples appears lower and narrower. However, the inclusion of non-wood samples towards higher % has notably expanded the ash content range, ultimately enhancing the model performance.

6.3.6 Comparison of the Model Performance between Chip and Ground Biomass

In this section, the performance of the best models for analyzing proximate parameters in chip and ground biomass was compared, as detailed in Tables 6.3 and 6.4. The primary difference in the NIR scanning process was that chip biomass was scanned using the diffuse reflectance sphere macro sample rotating mode in a large amount, while ground biomass was scanned using the transmittance mode in a small amount. All other scanning conditions, including resolution, background scan time, sample scan time, and controlled laboratory settings, remained consistent, particularly in terms of temperature. The results revealed contrasting outcomes for chip and ground biomass, with chip biomass outperforming in all proximate parameters except for ash content. This observation is particularly significant when considering that ground biomass samples are typically more homogeneous. One possible explanation for this difference may be attributed to the scanning mode employed for biomass analysis, specifically the use of diffuse reflectance sphere macro sample rotating mode. Chip biomass samples inherently vary in particle size and shape, and introducing the sphere macro sample rotating mode during diffuse reflectance scanning allows the samples to rotate in a larger cup with more constituent variation, facilitating high-throughput screening and aiding in averaging any inhomogeneities, thereby providing a more representative measurement of the entire sample. Consequently, this enhanced scanning technique contributes to the improved performance observed in chip biomass analysis. During transmittance scanning, although light diffuses better within homogeneous materials, the sample

size is very small and the constituent variation is minimal. Additionally, there may be a possibility of small NIR leakage via the glass vial edge, which might not provide useful information, resulting in comparatively lower performance compared to chip biomass analysis. For model of ash content, the chip model was poorer than the ground biomass model. The ash determination method by TGA in this experiment was by combustion up to 700 °C which is above 550 °C. Therefore, there is a risk in losing some of the potassium, and hence underestimating the ash content, while in chip, there was a full amount of these minerals.

6.3.7 Comparison with Previous Work

Several research studies have utilized NIR spectroscopy for proximate analysis of various biomass types. Typically, these studies have focused on specific biomass varieties. However, this study differs by developing models that encompass ten commonly used wood and non-wood biomass for energy purposes. The goal is to create a unified model for non-destructive assessment of proximate parameters. Unfortunately, the distribution of constituent parameters within both wood and non-wood is linear, resulting in a broad range for each parameter without clear separation into distinct groups. This study marks the first instance of combining these two types of data for estimation of proximate parameter analysis.

For MC analysis, Toscano et al. (2022) [21] conducted an evaluation of MC in industrial wood chips, analyzing a total of 817 samples. They reported R^2_p values ranging from 0.86 to 0.89. Similarly, Posom et al. (2022) [20] employed PLS regression to measure MC in cane bagasse, achieving an R^2_p value of 0.90. Sirisomboon et al. (2020) [22] assessed MC in bamboo chips, obtaining an R^2_p value of 0.96, while Adnan et al. (2017) [25] predicted the MC in intact green coffee beans, resulting in an R^2_p value of 0.96. Shrestha et al. (2018) [23] evaluated MC in bamboo chips using NIR-gun (600–1100 nm) and reported an R^2_p value of 0.92, whereas they achieved an R^2_p value of 0.74 using Micro-NIR (1150–215 nm). Similarly, Xue et al. (2015) [51] measured MC in corn stover with an R^2_c value of 0.81. In this study, the model performance is comparatively lower for both chip ($R^2_p = 0.8654$) and ground ($R^2_p = 0.7868$) biomass compared to the previous research mentioned above.

Similarly, for VM analysis for non-wood sample, previous research recorded R^2_p values of 0.81 for chip bamboo [22], 0.82 for ground bamboo [24], and 0.68 for corn stover [51]. In this

study, the R^2_p for VM in chip biomass (0.79) and ground biomass (0.77) are both lower when compared to the chip bamboo [22] and ground bamboo [24], respectively. However, when compared to the performance of corn stover, both the chip and ground combined wood and non-wood models exhibited better performance.

For FC, previous research recorded R^2_p values of 0.81 for chip bamboo [22], 0.66 for corn stover [51], and 0.85 for ground bamboo [24]. The model performance for FC in this study on chip biomass ($R^2_p = 0.88$) is better compared to the previous study.

For ash, being inorganic, it does not absorb NIR radiation [22,24]. However, the model developed for ash content in wood and non-wood chip (R^2_p of 0.7946) and ground biomass of wood and non-wood (R^2_p of 0.8005) in this study proves to be a reliable predictor. Previous research of non-wood model has reported R^2_p values of 0.86 for chip bamboo [22], 0.85 for corn stover [51], and 0.51 for ground bamboo [24]. Compared to the previous study, the performance of chip combined wood and non-wood is lower compared to chip bamboo [22] and corn stover [51]. However, the performance of ground combined model is comparatively higher than that of ground bamboo [24].

As recommended by Williams et al. (2019) [45], a model achieving an R^2 of 0.83–0.90 can be cautiously utilized for most applications, including research and from 0.66 to 0.81 is considered suitable for rough screening. Moreover, based on Zornoza et al. (2008) [46], an RPD value between 2 and 2.5 enables approximate quantitative predictions and 2.5 and 3 are indicative of good prediction. Therefore, the model selected in this research for non-destructive MC (R^2_p of 0.8654 and RPD of 2.9) and FC (R^2_p of 0.8773 and RPD of 3) assessment in chip biomass can be cautiously employed for most applications, including research while the other models, i.e., VM and ash models from chip and MC, VM, FC, and ash model from ground biomass can be applied for screening purposes only (R^2_p of 0.7733–0.8005 and RPD was 2.2–2.3). Therefore, it is advisable to employ the chosen model derived from chip wood and non-wood when evaluating proximate parameters for practical applications, particularly MC and FC.

6.4 Conclusions

In this study, proximate parameters (MC, VM, FC, and ash content) were determined using a TGA. PLSR models were developed and compared utilizing NIR spectroscopy within a

wavenumber range from 3594.87 to 12,489.48 cm^{-1} to assess these proximate analysis parameters, both in chip and ground biomass. These models were constructed from raw spectra, traditional preprocessing approaches, GA, SPA, and spectral MP with 5-range and 3-range methods.

The findings of this study underscore the potential of NIR spectroscopy as a promising alternative tool for the quantitative prediction of MC and FC in chip biomass. The selected model for MC and FC in chip biomass is suitable for most applications, including research, but should be used with caution. As for the remaining proximate parameters, i.e., MC and FC in ground biomass, and VM and ash content in both chip and ground biomass, the models exhibit fair performance and are primarily suitable for rough screening purposes. The models perform more accurately in chip biomass, highlighting the efficacy of diffuse reflectance sphere macro sample rotating mode compared to the transmittance mode of scanning in ground biomass. In addition, the chip biomass model can be preferred for proximate analysis instead of the ground biomass model, as it eliminates the need for biomass grinding, saving time, labor, and costs prior to scanning.

The model developed in this study proved to be more robust than those in previous studies. This is attributed to the wider variation in different types of biomass, including fast-growing trees and agricultural residues, whereas previous studies were focused on specific biomass species. The accuracy, robustness, and sensitivity of these models can be further enhanced by expanding the dataset with samples from diverse sources and validating their performance with unknown samples before industrial application. Implementing the selected models will benefit researchers, engineers, and industries aiming to design thermochemical conversion systems that can select a suitable biomass and efficiently extract maximum energy, all while minimizing energy consumption, optimized process, and environmental impact.

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*, 7205–7219.
2. Nimmanterdwong, P.; Chalermisinsuwan, B.; Piumsomboon, P. Prediction of lignocellulosic biomass structural components from ultimate/proximate analysis. *Energy* **2021**, *222*, 119945.
3. Saldarriaga, J.F.; Aguado, R.; Pablos, A.; Amutio, M.; Olazar, M.; Bilbao, J. Fast characterization of biomass fuels by thermogravimetric analysis (TGA). *Fuel* **2015**, *140*, 744–751.
4. Teh, J.S.; Teoh, Y.H.; How, H.G.; Sher, F. Thermal Analysis Technologies for Biomass Feedstocks: A State-of-the-Art Review. *Processes* **2021**, *9*, 1610.
5. Demirbas, A. Combustion characteristics of different biomass fuels. *Prog. Energy Combust. Sci.* **2004**, *30*, 219–230.
6. Aich, S.; Behera, D.; Nandi, B.K.; Bhattacharya, S. Relationship between proximate analysis parameters and combustion behaviour of high ash Indian coal. *Int. J. Coal Sci.* **2020**, *7*, 766–777.
7. Demirbas, A. Effects of moisture and hydrogen content on the heating value of fuels. *Energy Source Part A* **2007**, *29*, 649–655.
8. García, R.; Pizarro, C.; Lavín, A.G.; Bueno, J.L. Biomass proximate analysis using thermogravimetry. *Bioresour. Technol.* **2013**, *139*, 1–4.
9. Kongto, P.; Palamanit, A.; Ninduangdee, P.; Singh, Y.; Chanakaewsomboon, I.; Hayat, A.; Wae-hayee, M. Intensive exploration of the fuel characteristics of biomass and biochar from oil palm trunk and oil palm fronds for supporting increasing demand of solid biofuels in Thailand. *Energy Rep.* **2022**, *8*, 5640–5652.
10. Jia, Y.; Li, Z.; Wang, Y.; Wang, X.; Lou, C.; Xiao, B.; Lim, M. Visualization of Combustion Phases of Biomass Particles: Effects of Fuel Properties. *ACS Omega* **2021**, *6*, 27702–27710.
11. Vassilev, S.V.; Vassileva, C.G.; Song, Y.-C.; Li, W.-Y.; Feng, J. Ash contents and ash-forming elements of biomass and their significance for solid biofuel combustion. *Fuel* **2017**, *208*, 377–409.

12. Pazó, J.A.; Granada, E.; Saavedra, A.; Eguía, P.; Collazo, J. Biomass thermogravimetric analysis: Uncertainty determination methodology and sampling maps generation. *Int. J. Mol. Sci.* **2010**, *11*, 2701–2714.
13. Özyüğüran, A.; Yaman, S. Prediction of calorific value of biomass from proximate analysis. *Energy Procedia* **2017**, *107*, 130–136.
14. *ASTM E871-82(2019)*; Standard Test Method for Moisture Analysis of Particulate Wood Fuels. ASTM: West Conshohocken, PA, USA, 2019.
15. *ASTM E872-82(2019)*; Standard Test Method for Volatile Matter in the Analysis of Particulate Wood Fuels. ASTM: West Conshohocken, PA, USA, 2019.
16. *ASTM E1755-01(2020)*; Standard Test Method for Ash in Biomass. ASTM: West Conshohocken, PA, USA, 2020.
17. Sahito, A.R.; Mahar, R.; Siddiqui, Z.; Brohi, K. Estimating calorific values of lignocellulosic biomass from volatile and fixed solids. *Int. J. Renew. Energy Res.* **2013**, *2*, 1–6.
18. Posom, J.; Maraphum, K.; Phuphaphud, A. Rapid Evaluation of Biomass Properties Used for Energy Purposes Using Near-Infrared Spectroscopy. In *Renewable Energy-Technologies and Applications*; IntechOpen: London, UK, 2020.
19. Torquato, L.D.M.; Crnkovic, P.M.; Ribeiro, C.A.; Crespi, M.S. New approach for proximate analysis by thermogravimetry using CO₂ atmosphere: Validation and application to different biomasses. *J. Therm. Anal. Calorim.* **2017**, *128*, 1–14.
20. Posom, J.; Phuphaphud, A.; Saengprachatanarug, K.; Maraphum, K.; Saijan, S.; Pongkan, K.; Srimai, K. Real-time measuring energy characteristics of cane bagasse using NIR spectroscopy. *Sens. Bio-Sens. Res.* **2022**, *38*, 100519.
21. Toscano, G.; Leoni, E.; Gasperini, T.; Picchi, G. Performance of a portable NIR spectrometer for the determination of moisture content of industrial wood chips fuel. *Fuel* **2022**, *320*, 123948.
22. Sirisomboon, P.; Funke, A.; Posom, J. Improvement of proximate data and calorific value assessment of bamboo through near infrared wood chips acquisition. *Renew. Energy* **2020**, *147*, 1921–1931.
23. Shrestha, A.; Sirisomboon, P. Rapid evaluation of moisture content in bamboo chips using diode array near infrared spectroscopy. *MATEC Web Conf. —ICEAST 2018* **2018**, *192*, 03020.

24. Posom, J.; Sirisomboon, P. Evaluation of the higher heating value, volatile matter, fixed carbon and ash content of ground bamboo using near infrared spectroscopy. *J. Near Infrared Spectrosc.* **2017**, *25*, 301–310.
25. Adnan, A.; Hörsten, D.V.; Pawelzik, E.; Mörlein, A.D. Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy. *Foods* **2017**, *6*, 38.
26. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* **2023**, *16*, 5351.
27. Li, M.; He, S.; Wang, J.; Liu, Z.; Xie, G.H. An NIRS-based assay of chemical composition and biomass digestibility for rapid selection of Jerusalem artichoke clones. *Biotechnol. Biofuels* **2018**, *11*, 334.
28. 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*, 2001–2012.
29. 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 Chem. X* **2023**, *17*, 100539.
30. Singh, S.B.; De, M. Alumina based doped templated carbons: A comparative study with zeolite and silica gel templates. *Microporous Mesoporous Mater.* **2018**, *257*, 241–252.
31. Stolov, A.A.; Simoff, D.A.; Li, J. Thermal stability of specialty optical fibers. *J. Light. Technol.* **2008**, *26*, 3443–3451.
32. 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*, 1048.
33. López-González, D.; Fernandez-Lopez, M.; Valverde, J.L.; Sanchez-Silva, L. Thermogravimetric-mass spectrometric analysis on combustion of lignocellulosic biomass. *Bioresour. Technol.* **2013**, *143*, 562–574.

34. Shrestha, B.; Shrestha, Z.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Evaluating limit of detection and quantification for higher heating value and ultimate analysis of fast-growing trees and agricultural residues biomass using NIRS. *EASR* **2023**, *50*, 612–618.
35. Broad, N.; Graham, P.; Hailey, P.; Hardy, A. Guidelines for the development and validation of near-infrared spectroscopic methods in the pharmaceutical industry. *Handb. Vib. Spectrosc.* **2002**, *5*, 3590–3610.
36. Roger, J.-M.; Mallet, A.; Marini, F. Preprocessing NIR Spectra for Aquaphotomics. *Molecules* **2022**, *27*, 6795.
37. Robert, G.; Gosselin, R. Evaluating the impact of NIR pre-processing methods via multiblock partial least-squares. *Anal. Chim. Acta* **2022**, *1189*, 339255.
38. Jiang, W.; Lu, C.; Zhang, Y.; Ju, W.; Wang, J.; Hong, F.; Wang, T.; Ou, C. Moving-window-improved Monte Carlo uninformative variable elimination combining successive projections algorithm for near-infrared spectroscopy (NIRS). *J. Spectrosc.* **2022**, *2020*, 3590301.
39. Koljonen, J.; Nordling, T.E.; Alander, J.T. A review of genetic algorithms in near infrared spectroscopy and chemometrics: Past and future. *J. Near Infrared Spectrosc.* **2008**, *16*, 189–197.
40. Fei, Q.; Li, M.; Wang, B.; Huan, Y.; Feng, G.; Ren, Y. Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic algorithm for wavelength selection. *Chemometr. Intell. Lab. Syst.* **2009**, *97*, 127–131.
41. 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. *Eng. Appl. Sci. Res.* **2021**, *49*, 119–126.
42. Araújo, M.C.U.; Saldanha, T.C.B.; Galvão, R.K.H.; Yoneyama, T.; Chame, H.C.; Visani, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometr. Intell. Lab. Syst.* **2001**, *57*, 65–73.
43. Chen, Y.M.; Lin, P.; He, Y.; He, J.Q.; Zhang, J.; Li, X.L. Fast quantifying collision strength index of ethylene-vinyl acetate co-polymer coverings on the fields based on near infrared hyperspectral imaging techniques. *Sci. Rep.* **2016**, *6*, 20843.

44. Li, C.; He, M.; Cai, Z.; Qi, H.; Zhang, J.; Zhang, C. Hyperspectral Imaging with Machine Learning Approaches for Assessing Soluble Solids Content of Tribute Citrus. *Foods* **2023**, *12*, 247.
45. Williams, P.; Manley, M.; Antoniszyn, J. *Near Infrared Technology: Getting the Best out of Light*; African Sun Media: Stellenbosch, South Africa, 2019.
46. 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 Biol. Biochem.* **2008**, *40*, 1923–1930.
47. Guideline, I.H.T. Validation of analytical procedures: Text and methodology. *Q2 (R1)* **2005**, *1*, 5.
48. Werther, J.; Saenger, M.; Hartge, E.-U.; Ogada, T.; Siagi, Z. Combustion of agricultural residues. *Prog. Energy Combust. Sci.* **2000**, *26*, 1–27.
49. Yuan, Y.; Yong, H.; Tan, J.; Wang, Y.; Kumar, S.; Wang, Z. Co-Combustion characteristics of typical biomass and coal blends by thermogravimetric analysis. *Front. Energy Res.* **2021**, *9*, 753622.
50. Workman Jr, J.; Weyer, L. In *Practical Guide to Interpretive Near-Infrared Spectroscopy*; CRC Press: Boca Raton, FL, USA, 2007.
51. Xue, J.; Yang, Z.; Han, L.; Liu, Y.; Liu, Y.; Zhou, C. On-line measurement of proximates and lignocellulose components of corn stover using NIRS. *Appl. Energy* **2015**, *137*, 18–25.

Chapter 7

Near-Infrared Spectroscopy Modeling of Combustion Characteristics in Chip and Ground Biomass from Fast-Growing Trees and Agricultural Residue

This study focuses on the investigation and comparison of combustion characteristic parameters and combustion performance indices between fast-growing trees and agricultural residues as biomass sources. The investigation is conducted through direct combustion in an air environment using a thermogravimetric analyzer (TGA). Additionally, partial least squares regression (PLSR)-based models were developed to assess combustion performance indices via near-infrared spectroscopy (NIRS), serving as a non-destructive alternative method. The results obtained through the TGA reveal that, specifically, fast-growing trees display higher average ignition temperature (227 °C) and burnout temperature (521 °C) in comparison to agricultural residues, which exhibit the values of 218 °C and 515 °C, respectively. Therefore, fast-growing trees are comparatively difficult to ignite, but sustain combustion over extended periods, yielding higher temperatures. However, despite fast-growing trees having a high ignition index (D_i) and burnout index (D_f), the comprehensive combustion performance (S_i) and flammability index (C_i) of agricultural residue are higher, indicating the latter possess enhanced thermal and combustion reactivity, coupled with improved combustion stability. Five distinct PLSR-based models were developed using 115 biomass samples for both chip and ground forms, spanning the wavenumber range of 3595–12,489 cm^{-1} . The optimal model was selected by evaluating the coefficients of determination in the prediction set (R^2_p), root mean square error of prediction (RMSEP), and RPD values. The results suggest that the proposed model for D_f , obtained through GA-PLSR using the first derivative (D1), and S_i , achieved through full-PLSR with MSC, both in ground biomass, is usable for most applications, including research. The model yielded, respectively, an R^2_p , RMSEP, and RPD, which are 0.8426, 0.4968 $\text{wt}\% \text{ min}^{-4}$, and 2.5; and 0.8808, 0.1566 $\text{wt}\% \text{ min}^{-2} \text{ }^\circ\text{C}^{-3}$, and 3.1. The remaining models (D_i in chip and ground, D_f , and S_i in chip, and C_i in chip and ground biomass) are primarily applicable only for rough screening purposes. However, including more representative samples and exploring a more suitable machine learning algorithm are essential

for updating the model to achieve a better nondestructive assessment of biomass combustion behavior.

Keywords: biomass; combustion; thermogravimetric analyzer; near-infrared spectroscopy; partial least squares regression



*This chapter constituted the publication article: Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Pornchaloempong, P.; Funke, A. NIR Spectroscopy as an Alternative to Thermogravimetric Analyzer for Biomass Proximate Analysis: Comparison of Chip and Ground Biomass Models. *Energies* 2024, 17, 800. <https://doi.org/10.3390/en17040800>

Abbreviations

C_i	flammability index
D1	first derivative
D2	second derivative
D_i	ignition index
Df	burnout index
DTG	derivative thermogravimetric
FT	Fourier transform
GA	genetic algorithm
LV_s	number of latent variables
Max	maximum
Min	minimum
Mean	average
MSC	multiplicative scatter correction
MP	multi-preprocessing
NIRS	near-infrared spectroscopy
PLSR	partial least squares regression
R^2	coefficient of determination
R^2_c	coefficient of determination of calibration set
R^2_p	coefficient of determination of prediction set
RPD	ratio of prediction to deviation
RMSEC	root mean square error of calibration set
RMSEP	root mean square error of prediction set
S_i	comprehensive combustion performance
SD	standard deviation
SEC	standard error of calibration set

SEP	standard error of prediction set
SNV	standard normal variate
SPA	successive projection algorithm
TG	thermogravimetric
TGA	thermogravimetric analysis
T_i	ignition temperature
T_f	burnout temperature
t_i	ignition time
t_f	burnout time



7.1. Introduction

Global energy demand continues to escalate, prompting the exploration of diverse energy sources to meet this growing need, while mitigating the negative impacts on both energy availability and the environment. The predominant reliance on non-renewable fossil fuels not only contributes to environmental degradation but also raises concerns about future energy security due to the finite reserves. According to the International Energy Agency, as of 2020, 80% of global primary energy consumption was attributed to fossil fuels, resulting in a substantial carbon footprint [1]. The financial burdens associated with fossil fuels, exacerbated by fluctuating prices and geopolitical uncertainties, have triggered an urgent quest for alternative energy options.

A balanced and sustainable energy portfolio necessitates the promotion of renewable energy sources, primarily including hydro, wind, solar, biomass, and geothermal. Among these, biomass energy stands out as a promising solution, which accounts for 15% of the total energy consumption [2], and is derived from continuously renewable organic materials, such as wood, agricultural residues, and organic waste. Biomass energy conversion occurs predominantly through direct combustion [3], thermochemical processes (specifically pyrolysis and gasification) to produce solid (charcoal) and gaseous (syngas) fuels, as well as biological methods involving fermentation to produce ethanol and anaerobic digestion to yield methane-rich biogas. The utilization of biomass for energy purposes not only reduces reliance on non-renewable sources but also aids in waste management, contributing to rural development.

The fulfillment of global primary energy relies on the direct combustion of biomass and the co-combustion of two or more different fuels within the same combustion system, such as biomass and biochar [4], textile dyeing sludge and waste rubber [5], phytoremediation biomass and textile dyeing sludge [6], calcium-rich oil shale with biomass [7]. Despite biomass being deemed a carbon-neutral fuel [8], it exhibits varied combustion behaviors [9]. Therefore, careful management of the combustion process is vital to minimize emissions of additional pollutants, including particulate matter, sulfur oxides (SO_x), nitrogen oxides (NO_x), and volatile organic compounds [10]. A thorough comprehension of the combustion properties across different types of biomass is imperative to appropriately choose suitable biomass and design efficient combustion systems. Hence, combustion characteristic parameters, such as biomass ignition time (t_i) and ignition temperature (T_i), burnout time (t_f) and burnout temperature (T_f), maximum and

average combustion rate, etc., are essential for evaluating combustion performance indices such as the D_i , D_f , S_i , and C_i [11]. Accurate assessment of these indices can enhance the overall efficiency of the biomass combustion system, reduce environmental impacts, and bring us closer to achieving a sustainable energy future driven by renewable sources.

TGA is typically employed to determine combustion characteristic parameters for evaluating different combustion performance indices [12]. Biomass combustion in TGA mainly consists of three stages: (i) water evaporation, (ii) volatile release and its combustion, and (iii) char combustion [13–15]. TGA logs the mass loss of biomass as a function of time and temperature. As a result, the thermogravimetric (TG) curve obtained through TGA provides information about the mass loss of the biomass sample as it undergoes thermal decomposition and combustion. The DTG curve is derived as the D1 from the TG curves, providing additional information about the rate of mass loss at various times and temperatures [15]. Based on the TG and DTG curves, various combustion characteristic parameters can be identified. These parameters are used to evaluate the D_i , D_f , S_i , and C_i . TGA has been employed in a various range of studies, covering diverse aspects of combustion and thermal behavior. It has been used to assess the self-ignition potential of woody biomass and wheat straw [2], investigate the thermal behavior of Malaysian oil palm biomass, low-rank coal, and their respective blends under oxidative atmosphere [16], and identify thermo-chemical characteristics data for date palm biomass [17]. TGA has also been instrumental in studying the ignition behavior of straw pellets [18] and investigating ignition and burnout in bamboo and sugarcane bagasse [19]. Furthermore, TGA has been utilized to analyze the combustion characteristics of various biomass pellet types, including rubberwood sawdust pellets, teak sawdust pellets, eucalyptus bark pellets, cassava rhizome pellets [20], as well as agricultural solid waste torrefied pellets [21] and briquettes [15]. These studies collectively provide valuable insights into the reactivity, flammability, and thermal properties of these biomass materials, exhibiting their potential as fuels and their role in sustainable energy solutions.

The T_i is the lowest temperature at which solid fuel initiates ignition in air without requiring an external ignition source [2]. Ignition of biomass is a pivotal stage that initiates combustion. A lower D_i indicates that the biomass can be easily ignited and combusted at lower temperatures, while a higher D_i indicates that the biomass requires higher temperature to ignite and combust [22], making it more challenging to start the combustion process. Biomass with a higher volatile

matter poses a lower T_i and lower D_i , exhibiting ease of combustion [23]. It is important for biomass to ignite neither too quickly nor too slowly. Therefore, calculating the D_i is essential for understanding biomass ignition properties. The T_f indicates the temperature at which the combustion process of the biomass is completed [19]. A high D_f signifies complete combustion, leaving minimal unburned fuel or ash content [19,22]. A higher D_i and D_f indicate greater reactivity of the biomass, making it more suitable and flammable as a fuel [21]. The peak temperature is the point on the TGA curve at which the rate of weight loss of biomass due to combustion is at its maximum. This value typically varies around 280–300 °C [8]. For a thorough assessment of combustion behavior, it is essential to consider the S_i , which integrates three main properties of biomass combustion: ignition, burnout, and combustion characteristics [12]. A higher value of the S_i indicates efficient combustion, characterized by early ignition and thorough burnout [12,24,25]. Similarly, C_i is a crucial factor in assessing the fire risk and combustion behavior of biomass fuels. A higher C_i will have better combustion stability. It indicates that biomass can ignite easily at lower temperatures, releasing excess heat during combustion and supporting strong flames [26]. All of these indices provide valuable insights into the combustion characteristics of various biomass samples, enabling informed decisions when selecting suitable biomass and optimizing combustion system designs for efficient energy production and the effective use of the biomass as a fuel source, all while carefully considering safety aspects.

NIRS is one of the non-destructive, rapid, and low operation cost methods that do not require the employment of chemicals and chemical expertise. A mathematical correlation is established between the spectral and reference data of samples, containing either full wavelength ranges or a few significant wavelengths. This correlation is used to create the calibration equation for the prediction and evaluation of properties of biomass [27], such as elemental compositions (C, H, N, and S), determined by ultimate analysis [28,29], as well as moisture, volatile matter, fixed carbon, and ash content, assessed by proximate analysis [29,30]. The approach demonstrates acceptable performance and serves as an alternative to reference analysis, i.e., ultimate analysis and proximate analysis, which are characterized by their destructive nature, complexity, time-consuming process, and high operational costs, requiring chemicals and chemical expertise. The proximate constituents affect combustion performance [31], as well as the elemental composition, e.g., ignition temperature, which is determined by the H/C ratio and

some other parameters [32], indicating the possibility of using NIRS to determine the combustion performance of biomass or fuel.

To the best of our knowledge, no study has been conducted to non-destructively evaluate combustion performance indices, such as the D_i , D_f , C_i , and S_i in chipped and ground biomass using FT-NIRS. Therefore, this research is structured into two main sections. The first section involves determining the combustion parameters, including t_i and T_i , t_f and T_f , the maximum combustion rate ($\frac{dw}{dt}_{\max}$), and the average combustion rate ($\frac{dw}{dt}_{\text{mean}}$), using TGA to calculate the D_i , D_f , S_i , and C_i of biomass from fast-growing trees and agricultural residues. The second section focuses on developing calibration models using Full-PLSR, GA-PLSR, SPA-PLSR, the MP 5 range-PLSR, and the MP 3 range-PLSR for the non-destructive assessment of the D_i , D_f , S_i , and C_i in both chipped and ground biomass. Then, the best-performing PLSR-based model for each index is selected, establishing it as a rapid, reliable, non-destructive alternative method for assessing combustion performance indexes in both chipped and ground biomass.

The research outcomes will assist industries in selecting the most suitable biomass for cost-effective energy production and resource optimization. Additionally, the developed non-destructive evaluation methods will serve as an alternative method to other destructive thermal analysis methods. Furthermore, they will provide a foundation for designing safe, economical, and environmentally balanced biomass combustion systems.

7.2. Materials and Methods

Figure 7.1 illustrates the comprehensive research methodology employed to ascertain combustion performance indices and develop their predictive model utilizing TGA and NIRS.

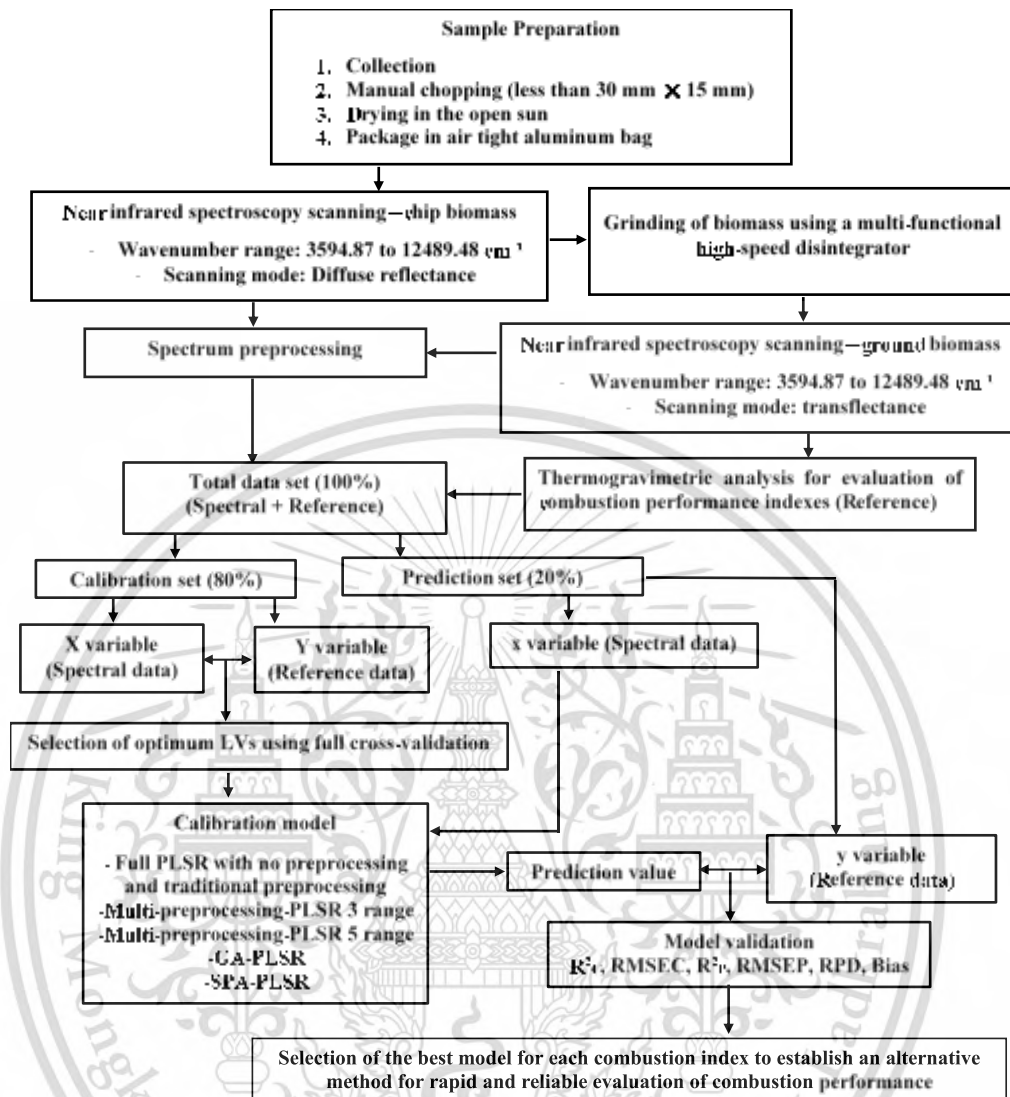


Figure 7.1 Flowchart illustrating the comprehensive research methodology for evaluating combustion performance indices of fast-growing trees and agricultural residues, using TGA in conjunction with NIRS combined with PLSR.

7.2.1. Sample Preparation

A total of ten different varieties of biomass samples were collected from the terai and mid-hill regions of Nepal, ranging from 86 to 1940 m above sea level, as representative samples. These biomass varieties are locally available and are commonly used in households and the industrial sector to fulfill their energy requirements. The biomass samples include fast-growing trees, i.e., *Alnus nepalensis* (11), *Pinus roxburghii* (14), *Bombusa vulagris* (13), *Bombax ceiba* (11), and *Eucalyptus camaldulensis* (12), as well as agricultural residues, i.e., *Zea mays* (cob) (13), *Zea mays* (shell) (11), *Zea mays* (stover) (11), *Oryza sativa* (10), and *Saccharum officinarum* (9). A total of 115 samples were collected for this experiment. The samples were manually chopped, sun-dried until the weight of the samples reached equilibrium, and then approximately 350 gm of each sample were sealed in airtight aluminum bags to prevent air and moisture exchange [33]. They were transported to the Near-Infrared Spectroscopy Research Center for Agricultural Products and Food at the Department of Agricultural Engineering, School of Engineering, King Mongkut's Institute of Technology, Ladkrabang, Thailand, for FT-NIRS sample scanning and TGA experiments.

Initially, all the biomass samples were scanned in chip form using FT-NIRS. Afterward, the samples were ground using a multi-functional high-speed disintegrator (EF-04, Thai grinder, Thailand) and sealed in plastic zip-lock bags to allow the samples to have equilibrium moisture content with the laboratory surroundings and to prevent any contamination. In this study, three ground biomass samples were randomly selected and their particle size distribution was analyzed at Chulalongkorn University's Scientific and Technological Research Equipment Center in Bangkok, Thailand. This assessment was conducted using the Mastersizer 3000 instrument (MAL1099267, Hydro MV). The average particle size distribution of the ground biomass ranges between 0.01 and 3080 μm . All the ground samples were subsequently scanned again using the same FT-NIRS instrument to record the absorbance value at each wavenumber. The ground samples, which were sealed in plastic bags, were opened only during the TGA experiments.

7.2.2. FT-NIRS Scanning

All the biomass samples were scanned non-destructively using FT-NIRS (MPA, Bruker, Ettlingen, Germany) within the wavenumber range of 3594.87 to 12,489.48 cm^{-1} . Biomass chips were scanned using diffuse reflectance and sphere macro sample rotating mode, whereas ground biomass was scanned in the transmittance mode, both at a resolution of 16 cm^{-1} . Background and sample scans were set at 32 scans (average), with absorbance data logged as $\log(1/R)$, where R stands for reflectance. Both the chipped and ground biomass were scanned twice in a controlled, air-conditioned laboratory environment, with the temperature maintained at 25 ± 2 °C, without altering their positions. To obtain accurate and informative results without interference from background spectral data in the biomass samples, a gold plate scan was performed for every new sample, and aluminum plates and handles were used to prevent the leakage of near-infrared radiation.

7.2.3. Thermogravimetric Analysis Experiment

TGA is a destructive yet an effective method for studying the thermal behavior of biomass and for evaluating the combustion performance indices [3]. The TGA investigation is based on the mass loss of biomass samples during the entire experimental duration [21]. The combustion setting in TGA (TG 209 F3 Tarsus, Netzsch, Germany), with a microbalance sensitivity of 0.1 μg resolution, is programmed to simulate biomass direct combustion in air, i.e., with oxygen (99.7%) and nitrogen (99.99%) in a 1:4 ratio. The TGA experiment utilized ground biomass samples collected from the bottom of the glass vial that were used during NIRS scanning. These samples had an approximate average mass ranging from 6 to 29 mg, or one-third of the crucible volume, and were used for direct combustion. The biomass samples were combusted in a 6.8 mm diameter aluminum oxide (Al_2O_3) crucible within a temperature range of 35 °C to 700 °C, with a heat flow rate of 10 °C/min. Initially, the samples were isothermally held at 35 °C for 10 min. Nitrogen (99.99%) was utilized as a protective layer in TGA to create a stable and inert environment, shielding the sample and preventing unwanted reactions with the surrounding air during thermal analysis, ensuring accurate analysis [34]. The thermal behavior of the biomass samples was analyzed by TG and DTG curves. Al_2O_3 crucibles were cleaned thoroughly using

distilled water, followed by oven drying for 24 h to remove moisture content. The TGA instrument was calibrated regularly with an empty, clean Al₂O₃ crucible.

7.2.4. Reference Data Calculation

The TG and DTG profiles were analyzed using Proteus 6.0.0 (NETZSCH software, Germany) to determine the key combustion parameters, including the maximum combustion rate ($\frac{dw}{dt}_{max}$), its corresponding time (t_p), and temperature (T_{max}) at which the maximum combustion rate occurs. The analysis also involved calculating the average combustion rate ($\frac{dw}{dt}_{mean}$), t_i , T_i , t_f , T_f , and $\Delta t_{1/2}$, representing the time range at which the combustion rate is half of the ($\frac{dw}{dt}_{max}$) value, measured in minutes. These parameters were used to compute combustion performance indices, such as D_i , D_f , S_i , and C_i , considering only one maximum mass loss peak, which collectively characterize the properties and performance of biomass combustion. The above-mentioned combustion performance indexes were calculated as follows [11]:

$$D_i = \frac{\frac{dw}{dt}_{max}}{t_p t_i} \quad (1)$$

$$D_f = \frac{\frac{dw}{dt}_{max}}{\Delta t_{1/2} t_p t_f} \quad (2)$$

$$S_i = \frac{\frac{dw}{dt}_{max} \times \frac{dw}{dt}_{mean}}{T_i^2 T_f} \quad (3)$$

$$C_i = \frac{\frac{dw}{dt}_{max}}{T_i^2} \quad (4)$$

7.2.5. Outlier Identification

Identification and removal of outliers from the total dataset is a critical step before developing any calibration model. In this study, outliers for reference data are identified using the following equation [35]:

$$\frac{(X_i - \bar{X})}{SD} \geq |\pm 3| \quad (5)$$

where X_i is the measured value of sample i , and \bar{X} and SD denote the average and standard deviation of the measured values across all samples. If the outlier equation is satisfied, the sample is considered an outlier and is subsequently removed from the total data set.

In addition, if performance of the model was not satisfactory, outliers were further identified using the reference and NIR absorbance data. To achieve this, a comprehensive full cross-validation was conducted to obtain the prediction values for each biomass sample. A scatter plot was then created, comparing the measured and predicted values of the calibration set. The rigorous outliers were carefully identified and subsequently removed if their patterns notably diverged from the majority of data points to improve model accuracy.

7.2.6. Partial Least Squares Regression Modeling

After the NIRS scanning (optical data) and the calculation of combustion performance indices (reference data) based on TG and DTG curves obtained through TGA, PLSR-based regression models were developed. Five different types of PLSR-based models were employed, namely Full-PLSR, multi-preprocessing PLSR-5 range, multi-preprocessing PLSR-3 range, GA-PLSR, and SPA-PLSR (refer to Figure 7.1). In this study, after running the data in ascending order, the total data set was manually divided into an 80% calibration set and a 20% validation set, where the first 8 samples were assigned to the calibration set and the following 2 were assigned to the validation set. The process was repeated until every sample was assigned. Both maximum and minimum reference data must be included in the calibration set, ensuring coverage across a wide range [33,35].

Full-PLSR includes the traditional approach of employing various spectral preprocessing techniques to develop a PLSR model. These techniques include raw spectra, constant offset, SNV, MSC, D1, D2, vector normalization, min-max normalization, mean centering, D1+vector normalization, and D1+MSC. In the multi-preprocessing 5-range method and the multi-preprocessing 3-range method, the entire available wavenumber range is divided into five and three distinct sections, respectively. Entire divided sections undergo pre-treatment using a series of the most effective combinations of various preprocessing techniques within the range of 3595 to 12,489 cm^{-1} . Under the multi-preprocessing techniques, seven different types of preprocessing techniques have been employed and labeled as follows: (0) Zero, indicating a zero absorbance

value for all the wavenumbers in the particular section, (1) raw spectra, (2) SNV, (3) MSC, (4) D1, (5) D2, and (6) constant offset. All possible preprocessing combination sets are created, and a full cross-validation is performed using PLSR on the total data set to identify the best preprocessing combination set. PLSR models are then developed based on this optimal combination set [33]. GA-PLSR and SPA-PLSR are optimization techniques that select the most influential wavenumbers for the development of a PLSR model [36]. The NIRS total dataset in this study contains 1154 dependent variables, which can potentially lead to issues of multicollinearity and overfitting during modeling. By efficiently identifying the most relevant wavenumbers, these optimization techniques address these challenges, resulting in a more accurate and efficient predictive model. After the models were optimized, they were externally validated using a validation set comprising 20% of the total samples collected. The validation was done by subjecting the validation sample spectra to the models and comparing the true (measured) values of the samples to the predicted values.

The performance of the models was compared based on the following statistical parameters: R^2_C and R^2_P , RMSEC and RMSEP, RPD and bias.

In this study, the interpretation of the coefficient of determination was performed based on Williams et al.'s (2019) guidelines [37], and the RPD value was assessed using the guideline proposed by Zornoza et al. (2008) [38]. The selection of the best model was based on higher values for R^2_C , R^2_P , and RPD, as well as lower RMSEC and RMSEP values. However, in the case of similar performance, the model with a lower number of LV_s was selected as the best-performing model. For the overall modeling, a built-in code from MATLAB-R2020b (MathWorks, Natick, MA, USA) was utilized.

7.3. Results and Discussion

7.3.1. NIR Spectra of Fast-Growing Trees and Agricultural Residues

Figure 7.2 shows the average raw spectra of fast-growing trees and agricultural residues from (a) chip biomass obtained through the diffuse reflectance mode of FT-NIRS scanning and (b) ground biomass obtained through the transreflectance mode of FT-NIRS scanning, covering the full wavenumber range from 3595 to 12,489 cm^{-1} , under a controlled air-conditioned laboratory environment. The temperature was maintained at 25 ± 2 °C and the moisture content inside the

spectrometer was absorbed by molecular sieve pellets. A significant variation is notable in the raw spectra between the chip and ground biomass samples. The ground biomass exhibited lower signal intensities, sharper and better-defined absorption peaks, as well as a reduced presence of baseline variability. These distinct observations are attributed to the small particle size and homogeneous nature of the biomass sample.

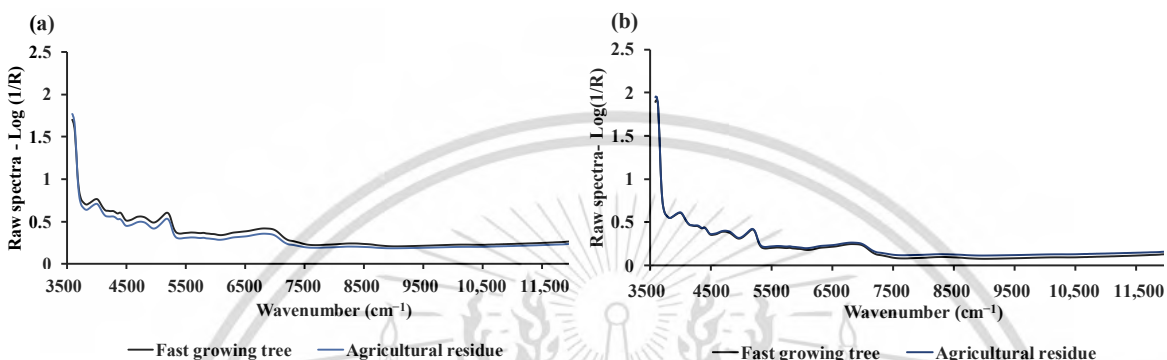


Figure 7.2 Average raw spectra of fast-growing trees and agricultural residue for (a) chip biomass obtained through diffuse reflectance mode and (b) ground biomass obtained through transmittance mode of FT-NIRS scanning.

7.3.2. Combustion Characteristic Parameters and Combustion Performance Indices from TGA

Figure 7.3 shows the typical TG and DTG curves of ten distinct biomass samples obtained via TGA, which has been utilized to calculate combustion performance indices and for PLSR modeling.

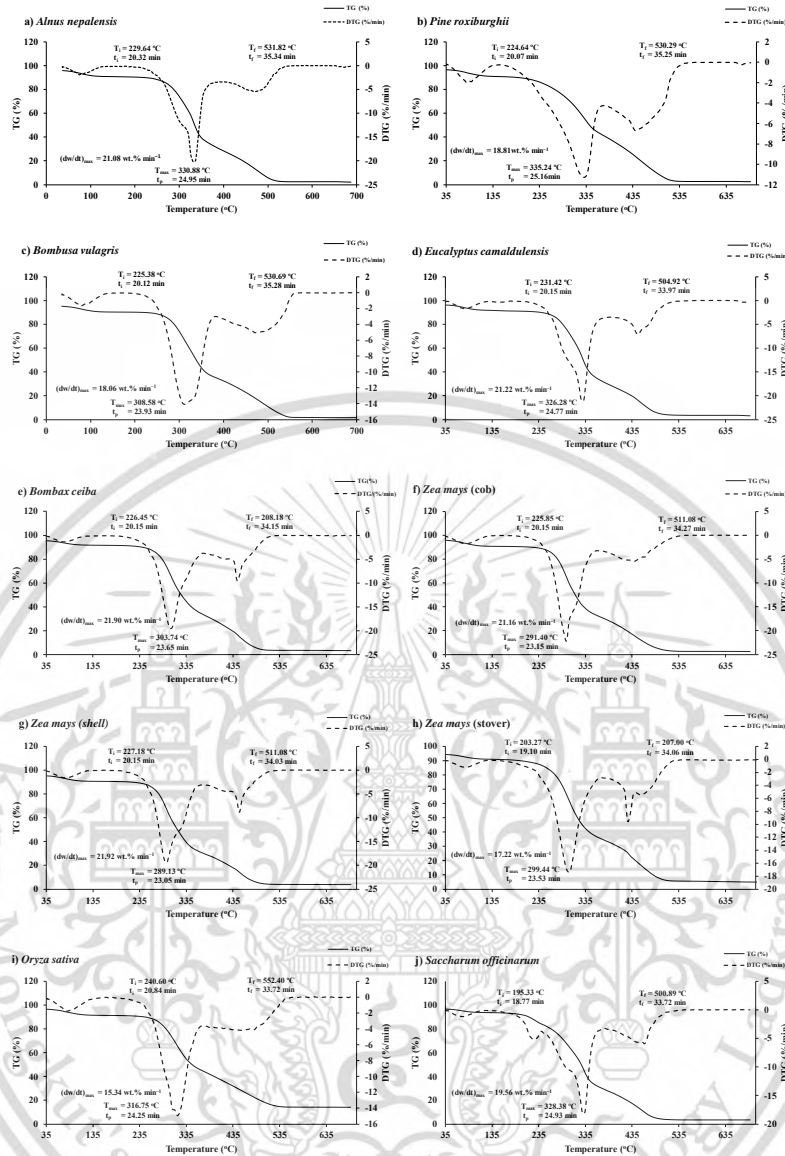


Figure 7.3 TG and DTG curves obtained via TGA within the temperature range of 35 to 700 °C for (a) *Alnus nepalensis*, (b) *Pinus roxburghii*, (c) *Bombusa vulagris*, (d) *Eucalyptus camaldulensis*, (e) *Bombax ceiba*, (f) *Zea mays* (cob), (g) *Zea mays* (shell), (h) *Zea mays* (stover), (i) *Oryza sativa*, and (j) *Saccharum officinarum*.

The selection of an appropriate biomass fuel is a crucial decision, heavily reliant on various essential parameters, such as energy production potential, the efficiency of the combustion process, required burning duration, compatibility with the system’s specifications, environmental considerations, and availability [39,40]. Therefore, it is of utmost importance to understand the biomass combustion characteristic parameters at different times and temperatures, as well as the overall combustion performance indices, before designing and developing a combustion system to fulfill energy needs and demands. TG and DTG curves obtained via TGA are instrumental in

investigating combustion characteristics parameters and their indices. The TG curve represents mass loss as a function of time or temperature, whereas the DTG curve represents the time derivative of the sample mass loss [20]. With a combined analysis of TG and DTG curves, a comprehensive evaluation of the combustion characteristic parameters and combustion performance indices can be achieved. T_i on the DTG curve is the point where the mass loss rate reaches 1%/min after the initial weight loss caused by the moisture. T_f marks the point at which the burning rate reaches 1%/min at the end of the DTG curve [41,42]. T_{max} , also known as peak temperature, is represented on the DTG curve where the biomass mass loss rate is the highest. Correspondingly, t_i , t_f , and t_p denote the recorded times for T_i , T_f , and T_{max} . $\Delta t_{1/2}$ is the time range at which the combustion rate is half of the $(\frac{dw}{dt}_{max})$ value, and $\frac{dw}{dt}_{mean}$ is the average conversion rate between T_i and T_f [41].

The normal distribution of all the combustion performance reference data, including D_i , D_f , S_i , and C_i was analyzed using a one-sample Kolmogorov–Smirnov test in SPSS 29.0. The corresponding p -values for D_i , D_f , S_i , and C_i were calculated as 0.893, 0.033, 0.000, and 0.608, respectively.

Considering the significance level of 0.05, it is observed that the p -values for D_i and C_i are greater than 0.05. As a result, the reference data for D_i and C_i , utilized in the PLSR modeling study, are considered to exhibit a uniform distribution. In contrast, the obtained p -values for D_f and S_i are less than 0.05, indicating a departure from a uniform distribution for these variables. Therefore, as explained in Section 2.5, the identification and removal of outliers from the total dataset of D_f and S_i become imperative. This step is crucial for enhancing the validity and reliability of the model developed in this research, as outliers can significantly impact the accuracy and robustness of the findings.

Table 7.1 summarizes the average combustion characteristic parameters (t_p , T_{max} , t_i , T_i , t_f , T_f , $\frac{dw}{dt}_{max}$, $\frac{dw}{dt}_{mean}$, $\Delta t_{1/2}$) and combustion performance indices (D_i , D_f , S_i , C_i) of the fast-growing trees and agricultural residues obtained through the combined analysis of TG and TGA curves (refer to Figure 7.3). During direct combustion, the first stage involves the removal of moisture from the biomass, a process represented by the gradual thermal degradation of the biomass. This typically takes place within a temperature range of 35–140 °C. The second stage involves devolatilization and its combustion, which occurs at temperatures around 150–405 °C and is

characterized by a rapid loss of mass. The third stage involves the char combustion, during which the rate of mass loss decreases and gradually slows down until the sample eventually turns into ash [21].

From Table 7.1, it is evident that fast-growing trees and agricultural residues are slightly different in both active combustion temperature ranges and time ranges. For fast-growing trees, the average active combustion temperature range was 227.51–521.18 °C, with a corresponding average time range of 20.22–34.80 min. For agricultural residues, the average active combustion temperature range was 218.45–515.51 °C, and the average time range was 19.8–34.49 min. The average maximum combustion rates recorded were 20.21 wt.% min⁻¹ at 320.94 °C for fast-growing trees and 19.04 wt.% min⁻¹ at 305.02 °C for agricultural residues.

The values of T_i and T_f for fast-growing trees ranged from 224.64 to 231.42 °C and from 504.92 to 531.82 °C, respectively. Similarly, for agricultural residues, the values ranged from 195.33 to 240.60 °C and from 500.89 to 552.40 °C. The higher values of T_i and T_f in fast-growing trees signify that fast-growing trees are more difficult to ignite, but they combust for a longer period and turn into ash more slowly than agricultural residues. The presence of high lignin content and low volatile matter in fast-growing trees may contribute to the elevated T_i and T_f [42,43]. The D_i of both fast-growing trees and agricultural residue is similar, while the D_f of fast-growing trees is comparatively higher than that of agricultural residues. The S_i , which assesses both the ignition and burnout characteristics of the fuel, indicating the efficiency of combustion conversion, is high for agricultural residues. This demonstrates that agricultural residues are easier to burn, indicating their higher thermal and combustion reactivity as a fuel source. Additionally, the higher C_i of agricultural residues indicates its better combustion stability.

7.3.3. Modeling for Combustion Performance Indices

Table 7.2 presents statistical data on the combustion performance indices: D_i , D_f , S_i , and C_i . These indices were employed in the development of the PLSR-based model for both chipped and ground biomass. Prior to model development, outliers were identified and were subsequently removed from the total dataset (refer to Section 2.5). The entire dataset was then partitioned, with 80% of the data being allocated to the calibration set containing the highest and lowest combustion performance index values, and 20% to the prediction set for both chipped and

ground biomass. As elaborated in Section 2.6, five distinct PLSR-based regression models were formulated for each index: the full-PLSR model, the multi-preprocessing PLSR-5 range model, the multi-preprocessing PLSR-3 range model, the GA-PLSR model, and the SPA-PLSR model. These models incorporated various preprocessing techniques. The performance of each model was compared, and the best model for each technique is listed. Tables 7.3 and 7.4 display the overall performance of the PLSR-based model for each index in chipped and ground biomass, respectively. The model with the best performance is indicated in bold. Figure 7.4 shows the average spectrum preprocessing for each combustion performance indices obtained for chip and ground biomass from the best performance model.

7.3.3.1. Ignition Index (D_i)

From the data in Tables 7.3 and 7.4, by R^2_p determination, the performance of D_i using the multi-preprocessing PLSR-5 range method for ground biomass has improved by 4.5645% compared to that of the traditional approach, i.e., the full-PLSR method.

Figure 7.5a,e show the scatter plots of the measured and predicted D_i values from the calibration and prediction sets for chip and ground biomass, respectively, using the full-PLSR and the multi-preprocessing PLSR-5 range methods.

Table 7.1 Combustion parameters and performance indices for fast-growing trees and agricultural residue subjected to direct combustion using TGA.

Category	Biomass Sample	Combustion Parameters										Combustion Performance Indices			
		(dw/dt) _{max}	(dw/dt) _{mean}	T _i	T _f	T _{max}	t _i	t _f	t _p	Δt _{1/2}	D _i (10 ⁻²)	D _f (10 ⁻³)	S _i (10 ⁻⁶)	C _i (10 ⁻⁴)	
		(wt.% min ⁻¹)	(wt.% min ⁻¹)	(°C)	(°C)	(°C)	(min)	(min)	(min)	(min)	wt.%.min ⁻³	wt.%.min ⁻⁴	wt.% ² .min ⁻² .C ⁻³	wt.%min ⁻¹ .°C ⁻²	
Fast-growing trees	<i>Alnus nepalensis</i>	21.08	2.83	229.64	531.82	330.88	20.32	35.34	24.95	5.19	4.15	4.79	2.54	3.99	
	<i>Pinus roxiburghii</i>	18.81	2.85	224.64	530.29	335.24	20.07	35.25	25.16	5.04	3.71	4.43	2.71	3.71	
	<i>Bombusa vulagris</i>	18.06	2.67	225.38	530.69	308.58	20.12	35.28	23.93	5.68	3.76	4.60	2.41	3.60	
	<i>Eucalyptus camaldulensis</i>	21.22	2.77	231.42	504.92	326.28	20.43	33.97	24.77	4.60	4.20	5.09	2.54	3.96	
	<i>Bombax ceiba</i>	21.90	2.65	226.45	508.18	303.74	20.15	34.15	23.65	5.25	4.61	6.05	2.41	4.30	
Agricultural residues	<i>Zea mays</i> (cob)	21.16	2.80	225.85	511.08	291.40	20.15	34.27	23.15	5.56	4.54	6.18	2.49	4.15	
	<i>Zea mays</i> (shell)	21.92	2.78	227.18	506.18	289.13	20.19	34.03	23.05	28.54	4.71	1.25	2.46	4.25	
	<i>Zea mays</i> (stover)	17.22	2.48	203.27	507.00	299.44	19.10	34.06	23.53	5.27	3.84	4.26	2.87	4.30	
	<i>Oryza sativa</i>	15.34	2.49	240.60	552.40	316.75	20.84	36.37	24.25	6.06	3.04	3.88	1.89	2.65	
	<i>Saccharum officinarum</i>	19.56	2.82	195.33	500.89	328.38	18.77	33.72	24.93	4.39	4.18	4.31	3.75	5.20	

Table 7.2 Statistical data of combustion performance parameters for ground and chipped biomass of fast-growing trees and agricultural residue used in model development (after outliers were removed from the total of 115 samples).

Biomass	Parameter (Ground)	Units	N _T	Calibration Set					Validation Set				
				Nc	Max	Min	Mean	SD	Np	Max	Min	Mean	SD
Ground	Ignition index D _i (10 ⁻²)	wt.%·min ⁻³	103	82	5.3496	2.4171	4.0699	0.6510	21	5.0998	2.8155	3.8740	0.7008
	Burnout index D _f (10 ⁻³)	wt.%·min ⁻⁴	87	70	6.7591	1.0380	4.2231	1.3066	17	6.5259	1.2071	4.2180	1.2905
	Comprehensive combustion index S _i (10 ⁻⁶)wt.% ² ·min ⁻² ·°C ⁻³		107	86	4.0363	1.6140	2.5704	0.4551	21	4.0296	1.7917	2.5502	0.4649
	Flammability index C _i (10 ⁻⁴)	wt.%·min ⁻¹ ·°C ⁻²	114	91	6.5187	2.3349	3.9879	0.8590	23	5.3362	2.4757	3.8578	0.6980
Chip	Ignition index D _i (10 ⁻²)	wt.%·min ⁻³	102	82	5.3500	2.7000	4.0532	0.6295	20	5.1000	2.8200	3.8975	0.7098
	Burnout index D _f (10 ⁻³)	wt.%·min ⁻⁴	94	75	7.1715	1.0380	4.4178	1.3070	19	6.9777	1.1030	4.5240	1.4880
	Comprehensive combustion index S _i (10 ⁻⁶)wt.% ² ·min ⁻² ·°C ⁻³		102	82	4.0363	1.7584	2.5577	0.4478	20	4.0296	1.7917	2.5325	0.4697
	Flammability index C _i (10 ⁻⁴)	wt.%·min ⁻¹ ·°C ⁻²	112	90	6.2216	2.3349	3.9384	0.7779	22	5.3362	2.4757	3.8255	0.6966

Table 7.3 Results of the PLSR-based model for the combustion performance indices of chip biomass, with the model showing the best performance highlighted in bold.

Parameter (Chip)	Units	Algorithm	Preprocessing	LVs	Calibration Set		Validation Set			
					R ² _C	RMSEC	R ² _P	RMSEP	RPD	Bias
D _i	wt.%·min ⁻³	Full-PLSR	Second derivative	6	0.6491	0.3706	0.6100	0.4321	1.6	-0.0996
		SPA-PLSR	Vector normalization (SW: 130)	9	0.6101	0.3907	0.5994	0.4379	1.6	-0.0770
		GA-PLSR	Vector normalization (SW: 518)	8	0.6479	0.3713	0.6073	0.4335	1.6	-0.1094
		MP-PLSR: 5-Range	Combination set: 2,4,0,5,5	4	0.5962	0.3976	0.5929	0.4414	1.6	-0.1071
		MP-PLSR: 3-Range	Combination set: 2,5,4	4	0.6015	0.3950	0.6008	0.4371	1.6	-0.0764
D _f	wt.%·min ⁻⁴	Full-PLSR	Constant offset	9	0.7470	0.6531	0.6920	0.8045	1.9	0.2043
		SPA-PLSR	Constant offset (SW: 717)	8	0.7335	0.6704	0.6738	0.8279	1.8	0.2549
		GA-PLSR	Min-max normalization (SW: 64)	10	0.7141	0.6943	0.7019	0.7914	1.9	0.1245
		MP-PLSR: 5-Range	Combination set: 6,6,4,6,0	9	0.7420	0.6596	0.6361	0.8744	1.7	0.1619
		MP-PLSR: 3-Range	Combination set: 1,6,6	10	0.7533	0.6450	0.6550	0.8515	1.8	0.2290
S _i	wt.% ² ·min ⁻² ·°C ⁻³	Full-PLSR	Raw spectra	9	0.7700	0.2136	0.7699	0.2196	2.1	0.0372
		SPA-PLSR	First derivative+MSC (SW: 346)	12	0.8153	0.1914	0.7484	0.2296	2.0	-0.0122
		GA-PLSR	First derivative (SW: 18)	11	0.8006	0.1989	0.7812	0.2141	2.2	0.0535
		MP-PLSR: 5-Range	Combination set: 3,5,3,6,0	9	0.8068	0.1958	0.7721	0.2185	2.2	0.0533
		MP-PLSR: 3-Range	Combination set: 6,2,4	3	0.6047	0.2800	0.5126	0.3196	1.4	-0.0414
C _i	wt.%·min ⁻¹ ·°C ⁻²	Full-PLSR	SNV	14	0.8215	0.3267	0.6119	0.4240	1.6	0.0523
		SPA-PLSR	Second derivative (SW: 213)	11	0.6797	0.4377	0.6439	0.4061	1.7	-0.0297
		GA-PLSR	Mean centering (SW: 16)	13	0.5744	0.5045	0.5666	0.4481	1.5	0.0823
		MP-PLSR: 5-Range	Combination set: 2,2,1,6,5	9	0.6469	0.4595	0.6853	0.3818	1.8	-0.0652
		MP-PLSR: 3-Range	Combination set: 2,5,0	14	0.6903	0.4304	0.6766	0.3871	1.8	-0.0343

Refer to the unit column for the RMSEC, RMSEP, and bias units for D_i, D_f, S_i, and C_i.

Table 7.4 Results of the PLSR-based model for the combustion performance indices of ground biomass, with the model showing the best performance highlighted in bold.

Parameter (Ground)	Units	Algorithm	Preprocessing	LVs	Calibration Set		Validation Set			
					R^2_c	RMSEC	R^2_p	RMSEP	RPD	Bias
D_i	wt.%·min ⁻³	Full-PLSR	Raw spectra	8	0.6533	0.3810	0.6466	0.4064	1.7	-0.0898
		SPA-PLSR	Raw (SW: 1132)	8	0.6542	0.3805	0.6472	0.4062	1.7	-0.0898
		GA-PLSR	Mean centering (SW:523)	9	0.6442	0.3859	0.6071	0.4286	1.6	-0.0743
		MP-PLSR: 5-Range	Combination set: 3, 5, 3, 1, 0	9	0.7039	0.3521	0.6782	0.3879	1.8	-0.0016
		MP-PLSR: 3-Range	Combination set: 1, 4, 6	13	0.7773	0.3053	0.5634	0.4518	1.5	-0.0511
D_f	wt.%·min ⁻⁴	Full-PLSR	First derivative (g = 5, s = 5)	11	0.8449	0.5111	0.8217	0.5286	2.4	0.0678
		SPA-PLSR	Second derivative(SW: 954)	10	0.8139	0.5598	0.8001	0.5598	2.2	0.0206
		GA-PLSR	First derivative (SW:921)	11	0.8417	0.5163	0.8426	0.4968	2.5	0.0631
		MP-PLSR: 5-Range	Combination set: 1,5,4,3,6	12	0.8151	0.5580	0.8018	0.5574	2.3	0.1177
		MP-PLSR: 3-Range	Combination set: 2,2,1	14	0.8240	0.5443	0.8137	0.5405	2.6	0.2432
S_i	wt.% ² ·min ⁻² ·°C ⁻³	Full-PLSR	MSC	14	0.9028	0.1411	0.8808	0.1566	3.1	0.0532
		SPA-PLSR	MSC (SW: 626)	13	0.8849	0.1536	0.8045	0.2005	3.0	0.1298
		GA-PLSR	MSC (SW: 60)	10	0.8567	0.1713	0.8566	0.1717	2.8	-0.0632
		MP-PLSR: 5-Range	Combination set: 4,4,5,6,4	12	0.9449	0.1062	0.8136	0.1958	2.3	0.0102
		MP-PLSR: 3-Range	Combination set: 4,2,1	13	0.9071	0.1380	0.8316	0.1861	2.5	-0.0257
C_i	wt.%·min ⁻¹ ·°C ⁻²	Full-PLSR	MSC	15	0.7881	0.3932	0.6914	0.3792	1.9	-0.1361
		SPA-PLSR	Raw (SW: 13)	15	0.7234	0.4492	0.6524	0.4025	1.8	-0.1162
		GA-PLSR	Raw (SW: 333)	9	0.5822	0.5520	0.5476	0.4592	1.5	-0.0477
		MP-PLSR: 5-Range	Combination set: 3,2,1,1,4	12	0.7576	0.4205	0.7204	0.3610	2.0	-0.1310
		MP-PLSR: 3-Range	Combination set: 1,2,4	15	0.7860	0.3951	0.6919	0.3790	1.9	-0.0884

Refer to the unit column for the RMSEC, RMSEP, and bias units for D_i , D_f , S_i , and C_i .

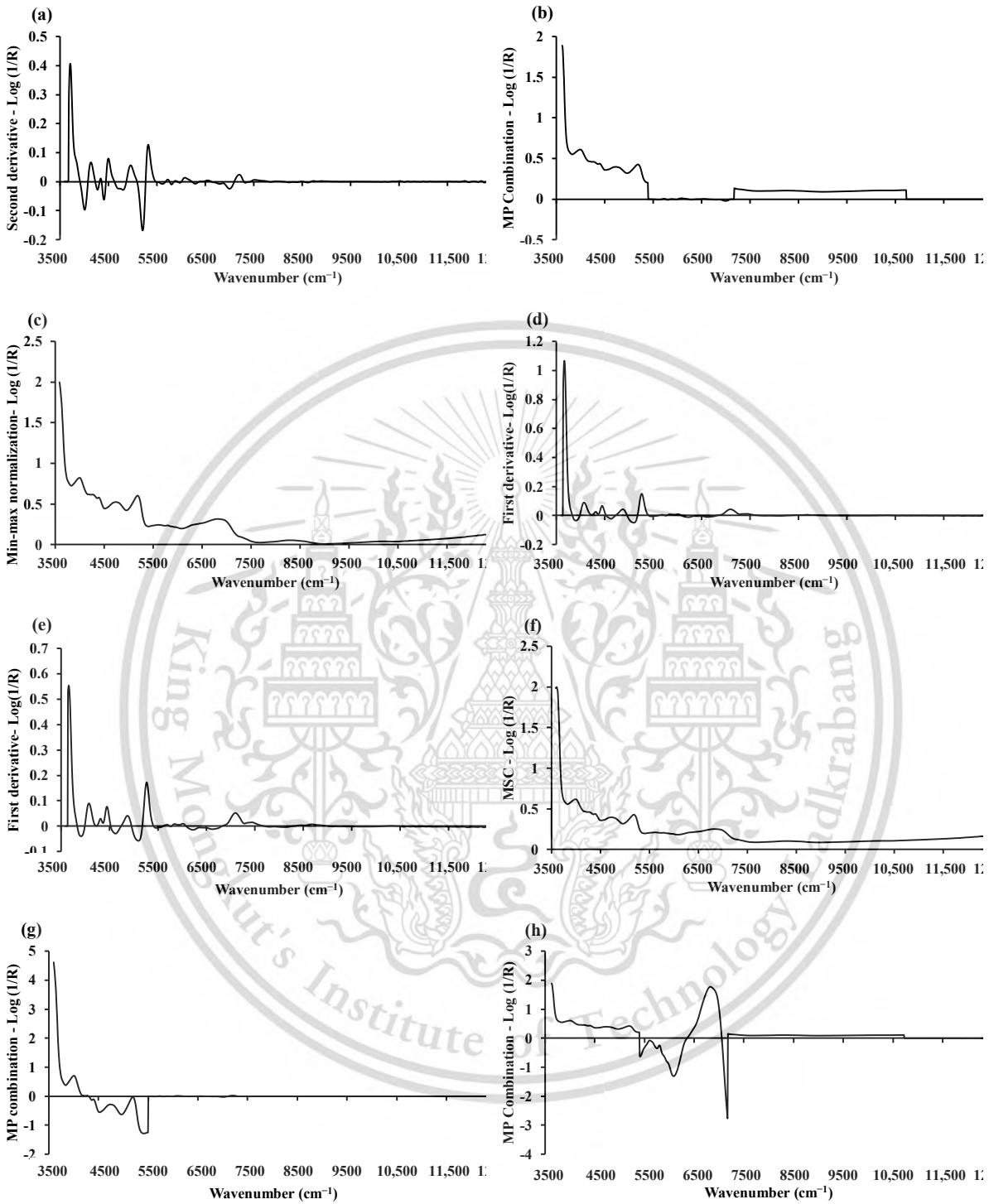


Figure 7.4 The average spectrum for (a) D_i in chip biomass is obtained by using the D2 (5,5). (b) D_i in ground biomass is obtained by using the multi-preprocessing PLSR-5 range with a combination set of 3,5,3,1,0. (c) D_f in chip biomass is achieved by using min-max normalization. (d) D_f in ground biomass is obtained by using D1 (5,5).

(e) S_i in chip biomass is acquired by using D1 (5,5). (f) S_i in ground biomass is acquired by MSC. (g) C_i in chip biomass is calculated by using the multi-preprocessing PLSR-3 range method with a combination set of 2,5,0. (h) C_i in ground biomass is determined by the multi-preprocessing PLSR-5 range method with a combination set of 3,2,1,1,4.

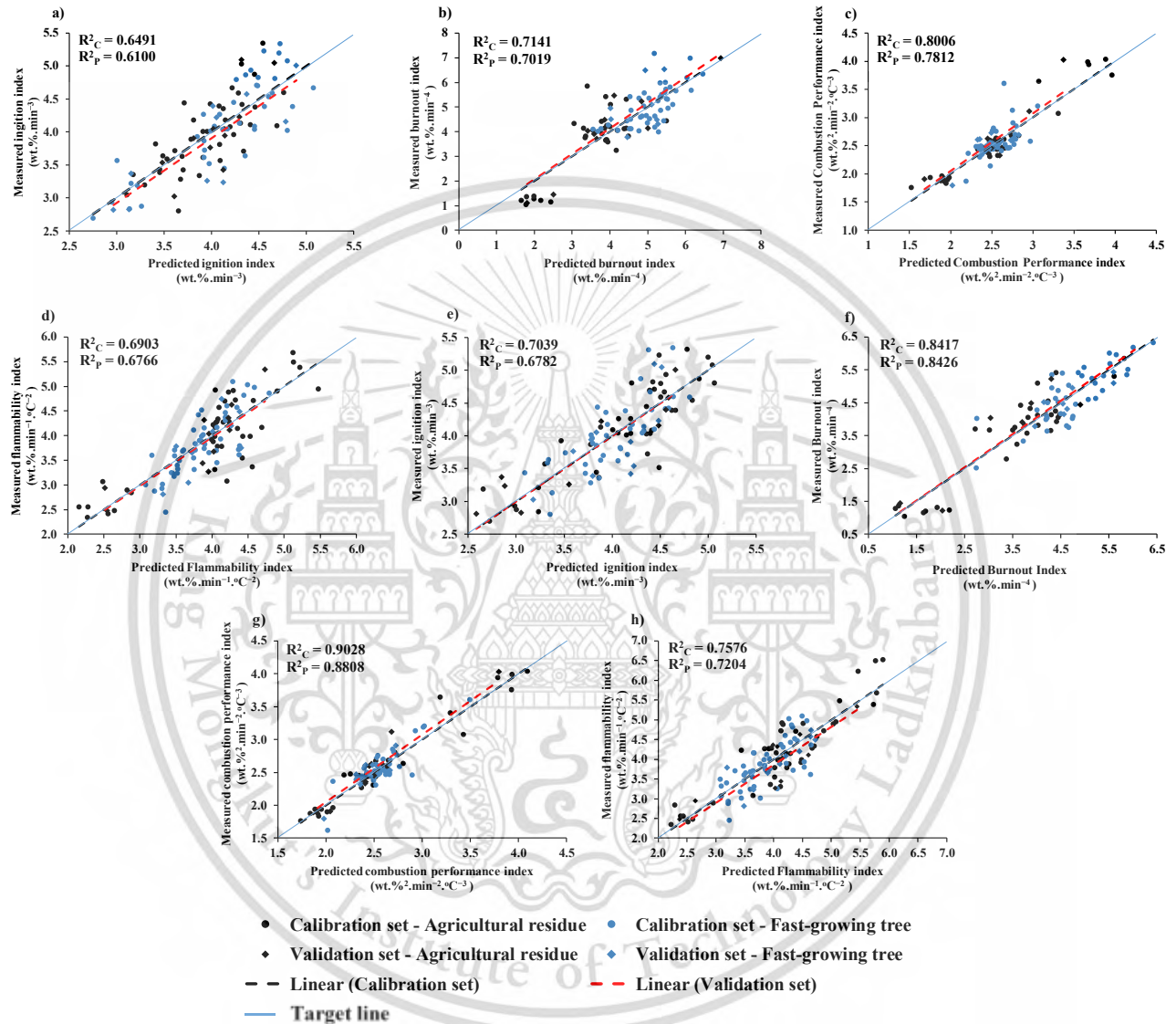


Figure 7.5 Measured versus predicted value in calibration set and validation set for chip biomass: (a) ignition index, (b) burnout index, (c) combustion performance index, and (d) flammability index; and for ground biomass: (e) ignition index, (f) burnout index, (g) combustion performance index, and (h) flammability index.

The regression coefficient plot from the full-PLSR D2 analysis for chip biomass is presented in Figure 7.6. The most important peaks are observed in the regression coefficient plot, specifically at wavenumbers 3722, 4405, 5200, 5787, 12,048, and 12,300 cm^{-1} . These peaks might have a

significant influence on enhancing the performance of the model in evaluating the D_i in chip biomass.

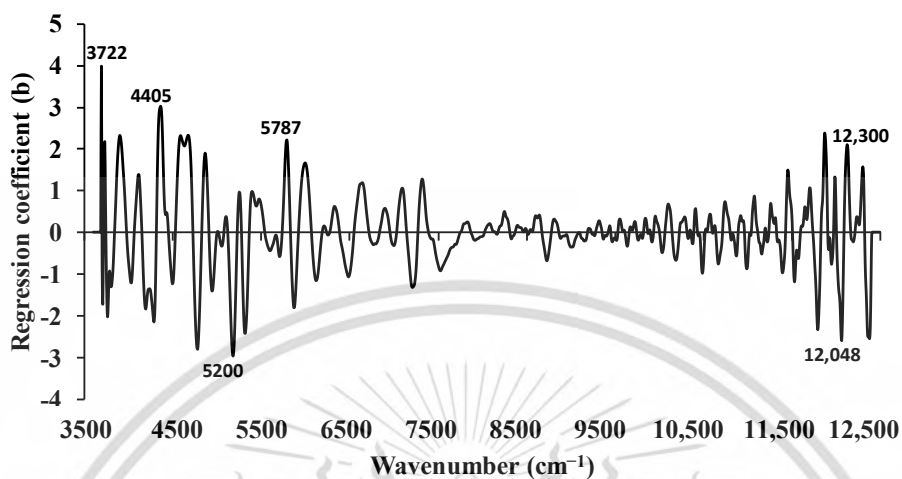


Figure 7.6 The regression coefficient for the D_i of chip biomass using the full-PLSR approach with spectral preprocessing of D2.

Similarly, Figure 7.7 displays the regression coefficient plot for the D_i of ground biomass, obtained using the multi-preprocessing PLSR-5 range method. The important peaks are noticed at wavenumbers 3650, 4608, 5495, and 8754 cm^{-1} , which might significantly influence the model's performance. Here, regression coefficient values within the range of 10,723–12,489 cm^{-1} are observed to be zero. This observation suggests that the variable in this section, which is assigned a zero absorbance using empty preprocessing, may not possess sufficient variation in the dataset to yield meaningful predictive power.

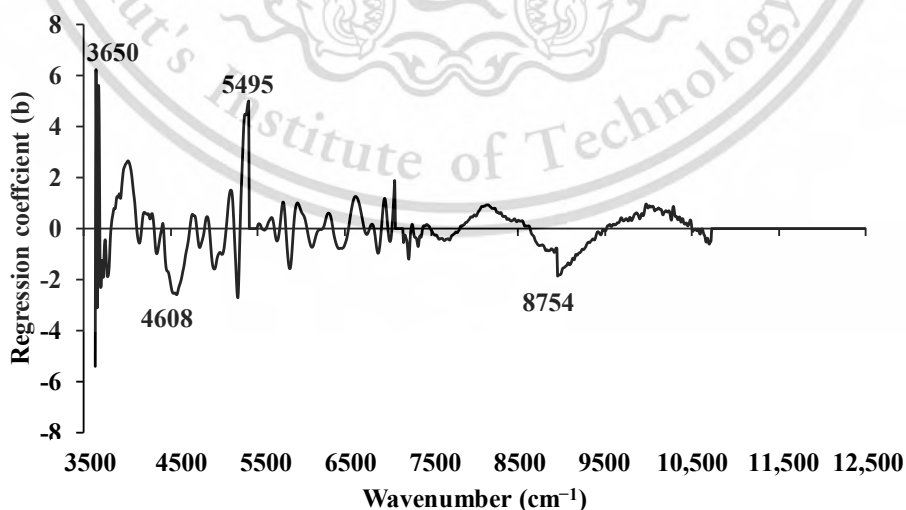


Figure 7.7 The regression coefficient for the D_i of ground biomass using the multi-preprocessing PLSR-5 range method with a spectral preprocessing combination set of 3, 5, 3, 1, 0.

Table 7.5 displays the functional groups, spectra-structure, and material types corresponding to the specific peak wavenumbers observed in the regression coefficient plots through NIRS analysis of chip and ground biomass for D_i [44].



Table 7.5 The dominant peaks on the regression coefficient plot and average absorbance plot obtained via the best-performing PLSR-based model [44].

Combustion Performance Index	Biomass Type	Peak Wave-number (cm ⁻¹)	Functional Group	Spectra-Structure	Material Type	
D _i	Chip	3722	C-H aromatic		C-H aryl	
		4405		O-H stretching and C-O stretching	cellulose	
		5200		O-H stretching and HOH deformation combination	O-H molecular water	
		5787	C-H methylene (-CH ₂) (asymmetric)		Hydrocarbons, aliphatic	
		12,048	C-H methylene C-H		Hydrocarbons, aliphatic	
	12,300		C-H combination	Hydrocarbons, aliphatic		
	Ground	3650	O-H from primary alcohols as (-CH-OH)	O-H (V)	Primary alcohols	
		4608		C-H stretching and C-H deformation combination	Alkenes	
		5495	O-H/C-H combination	O-H stretching and C-O stretching (3V ₂) combination	Cellulose	
		8754	C-H aromatic (ArCH)	C-H (3V ₂), aromatic C-H	Hydrocarbons, aromatic	
D _r		Chip	4019		C-H stretching and C-C stretching combination	Cellulose
	5181			O-H stretching and HOH bending combination	Polysaccharides	
	6319		O-H stretching band, alkyl alcohols or water		Alcohols or water O-H	
	9960	O-H from secondary alcohols as (-CH-OH)	O-H (3V ₂)(-CH-OH)	Secondary alcohols		
	Ground	3650	O-H from primary alcohols as (-CH-OH)	O-H (V)	Primary alcohols	
		4019		C-H stretching and C-C stretching combination	Cellulose	
		5200		O-H stretching and HOH deformation combination	O-H molecular water	
		6897		O-H (2V)	Starch/polymeric alcohol	
	S _i	Chip	4019		C-H stretching and C-C stretching combination	Cellulose
			4292		C-H stretching and CH ₂ deformation combination	Polysaccharides
7092			O-H alcohol (RO-H)	O-H (2V)	Hydrocarbons, aliphatic	
Ground		4525	N-H ammonia in water	N-H (3V) for NH ₃ in water	Ammonia in water	
		4762		O-H bending and C-O stretching combination	Polysaccharides	
		5376		C-Cl (7V)	Chlorinated hydrocarbons	
		5869		C-H (2V), methyl C-H (symmetric)	Hydrocarbons, aliphatic	
7092		O-H alcohol (RO-H)	O-H (2V)	Hydrocarbons, aliphatic		
12,300			C-H combination	Hydrocarbons, aliphatic		
C _i		Chip	4202		C-H stretching and C-C stretching combination	Lipids
	4307			C-H stretching and CH ₂ deformation combination	Polysaccharides	
	5241		P-OH phosphate (P-OH)	O-H (2V)	Phosphate	
	Ground	5495	O-H/C-H combination	O-H stretching and C-O stretching (3V ₂) combination	Cellulose	
		5495	O-H/C-H combination	O-H stretching and C-O stretching (3V ₂) combination	Cellulose	
		5900	C-H methyl (-CH ₃)	C-H (2V), -CH ₃	Hydrocarbons, methyl	
		6666	N-H combination band from urea (NH ₂ -C=O-NH ₂)		N-H from urea	
		6736	N-H band from urea (NH ₂ -C=O-NH ₂)	N-H (2V) symmetric stretching from urea	Urea	

V: Fundamental stretching vibration absorption band, 2V: first overtone of fundamental stretching band, 3V: second overtone of fundamental stretching band, 7V: six overtone of fundamental stretching band.

7.3.3.2. Burnout Index (D_f)

As shown in Tables 7.3 and 7.4, the best-performing PLSR-based models for both chip and ground biomass were obtained using GA-PLSR. For chip biomass, GA selected 64 out of 1154 influential wavenumbers with spectral preprocessing using min-max normalization (refer to Figure 7.5c). For ground biomass, GA selected 921 out of 1154 wavenumbers with spectral preprocessing using the D1 (segment = 5, gap = 5) (refer to Figure 7.5d). Figure 7.5b,f show the scatter plots for measured versus predicted D_f for chip and ground biomass.

By R^2_p determination, the GA-PLSR model performance of D_f in chip biomass has improved by 1.6332% compared to the full-PLSR method.

Figure 7.8 shows the average absorbance spectrum pretreated with min-max normalization, using red marks to emphasize important wavenumbers identified through GA. The selected significant wavenumbers, located at 4019, 5181, 6319, and 9960 cm^{-1} , could potentially exert a notable influence on the model's performance in evaluating the D_f in the chip biomass samples. Similarly, Table 7.5 presents the associated functional groups, spectra-structure, and the material type corresponding to specific peak wavenumbers observed in D_f chip biomass samples [44].

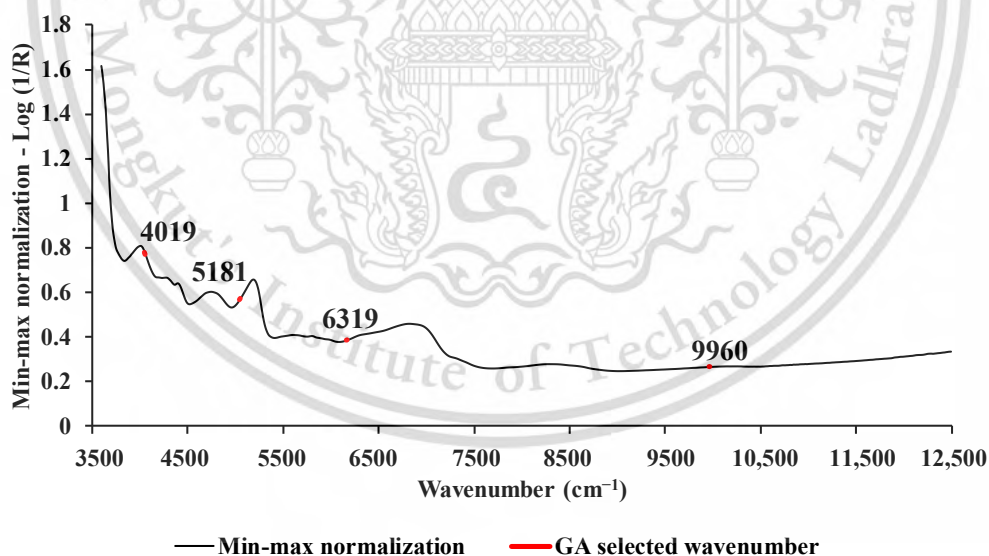


Figure 7.8 The average absorbance value of D_f in chip biomass using min-max normalization with selection of important wavenumbers obtained through GA.

For ground biomass, notably, as determined by R^2_p , the model performance of D_f in ground biomass has improved by 6.0322% compared to the full-PLSR approach.

Figure 7.9 displays the average absorbance values of D_f in ground biomass, obtained after preprocessing with D1. The figure highlights the 921 selected wavenumbers (marked in red) obtained via GA, encompassing the full spectral range of 3594.87–12,489.5 cm^{-1} . The important peaks selected at 3650, 4019, 5200, and 6897 cm^{-1} could significantly influence the model performance in evaluating the D_f in ground biomass. Table 7.5 presents the associated functional groups, spectra-structure, and their material types corresponding to specific peak wavenumbers observed in D_f ground biomass samples [44].

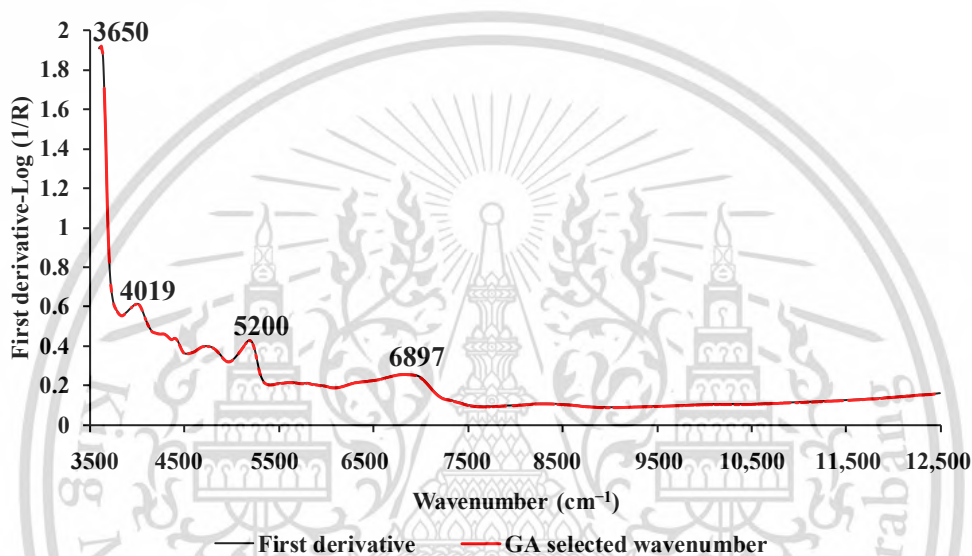


Figure 7.9 The average absorbance value of D_f in ground biomass using D1 with selection of important wavenumbers obtained via GA.

7.3.3.3. Comprehensive Combustion Index (S_i)

As shown in Table 7.3, the best-performing model for chip biomass was obtained through GA-PLSR with spectral preprocessing using D1, with a gap of 5 and a segment of 5 (refer to Figure 7.5e). By R^2 determination, the model-explained variance for S_i in chip biomass improved by 2.4712% compared to the full-PLSR method. Figure 7.5c shows the scatter plot of measured versus predicted S_i using GA-PLSR.

Figure 7.10 shows the average absorbance values of the S_i in chip biomass, obtained after preprocessing with D1. The figure highlights the 18 selected wavenumbers (marked in red) obtained through GA, covering the entire spectral range of 3595 to 12,489 cm^{-1} . Notably, important peaks were observed at wavenumbers 4019, 4292, and 7092 cm^{-1} , respectively, suggesting a potentially pivotal influence on the model's enhanced performance. Refer to Table

7.5, which presents the associate functional groups, spectra-structure, and the material type corresponding to specific peak wavenumbers observed in S_i chip biomass samples [44].

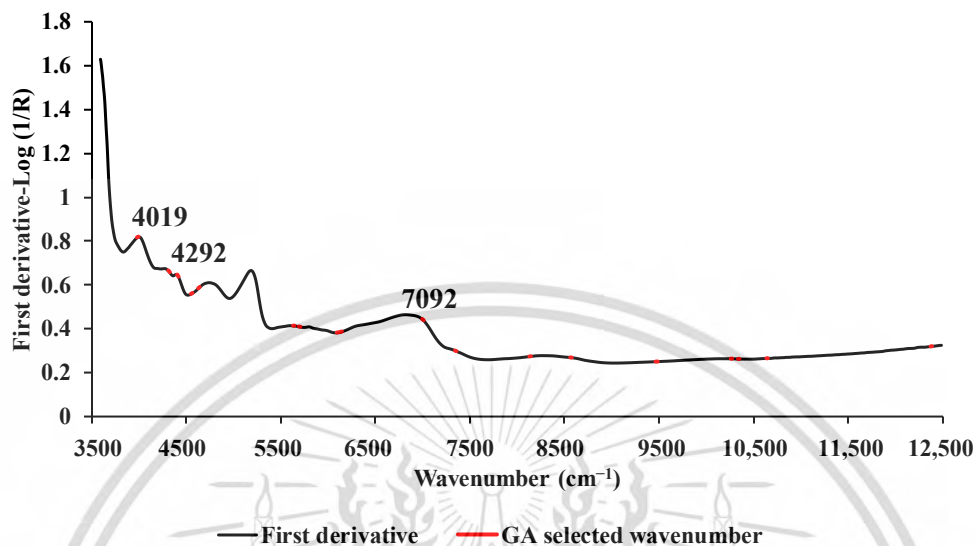


Figure 7.10 The average absorbance value of S_i in chip biomass using D1 with selection of important wavenumbers obtained through GA.

For ground biomass (refer to Table 7.4), the model performance using full-PLSR with MSC (refer to Figure 7.5f) as preprocessing and 14 LVs provides the best assessment for the S_i in terms of R^2_C , RMSEC, R^2_p , RMSEP, RPD, and bias. Figure 7.6g shows the scatter plot of measured versus predicted S_i using full-PLSR with spectrum preprocessing using MSC.

Figure 7.11 displays the regression coefficient for the S_i in ground biomass using full-PLSR with spectral preprocessing of MSC within the wavenumber range of 3595 to 12,489 cm^{-1} . The important peaks are noticed at 4525, 4762, 5376, 5869, 7092, and 12,300 cm^{-1} , which may significantly influence the enhanced performance of the model. Refer to Table 7.5, which presents the associate functional groups, spectra-structure, and the material type corresponding to specific peak wavenumbers observed in S_i ground biomass samples [44].

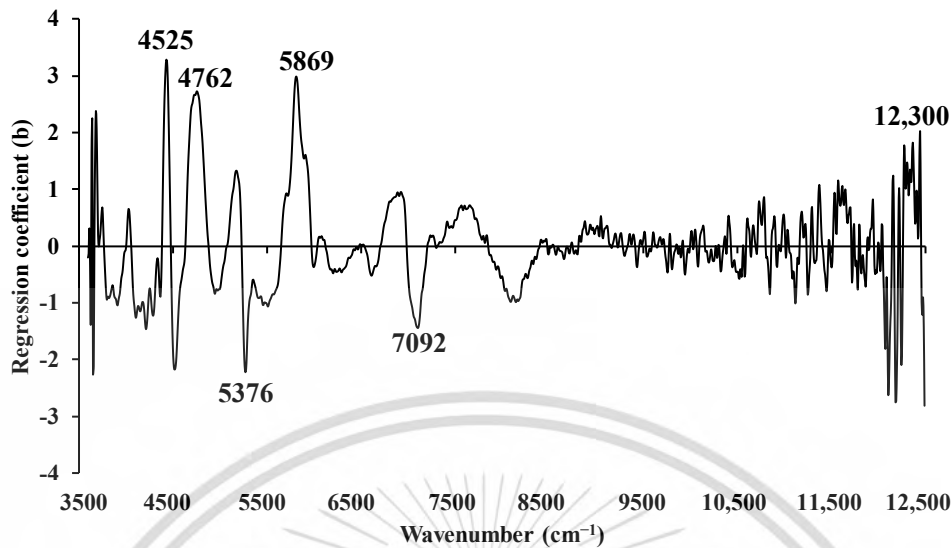


Figure 7.11 The regression coefficient for the S_i of ground biomass using the full-PLSR model with spectral preprocessing of MSC.

7.3.3.4. Flammability Index (C_i)

For the chip biomass model, the multi-preprocessing-PLSR with the 3-range method, employing 14 LVs and utilizing the multi-preprocessing combination set of 2, 5, 0 (i.e., SNV within the wavenumber range 3595–5493 cm^{-1} , D2 (segment = 5, gap = 5) within 7498–5500 cm^{-1} , and empty, i.e., absorbance value equals to zero, within 7506–12,489 cm^{-1}) (refer to Figure 7.4g), provides the best performance for assessing C_i . Figure 7.5d show the scatter plot of measured versus predicted C_i using the multi-preprocessing PLSR-3 range method. In addition, the model performance is improved by 8.7151% compared to the full-PLSR approach.

Figure 7.12 displays the regression coefficient for the C_i of chip biomass, utilizing the multi-preprocessing PLSR-3 range method. Notably, important peaks that might significantly influence the enhancement of the model performance are observed at wavenumbers 4202, 4307, 5241, and 5495 cm^{-1} . Within the wavenumber range of 7500–12,500 cm^{-1} , an observed regression coefficient value of zero suggests that variations in this specific variable are not associated with any changes in the predicted outcome. Table 7.5 presents the associated functional groups, spectra-structure, and the material type corresponding to specific peak wavenumbers observed in C_i chip biomass samples [44].

From Tables 7.3 and 7.4, the overall performance of the C_i model for ground biomass is comparatively better than that for chip biomass. The best model was achieved using the multi-

preprocessing PLSR-5 range method with a preprocessing combination set of 3, 2, 1, 1, 4. This combination includes the MSC from 3626 to 5392 cm^{-1} , the SNV from 5400 to 7167 cm^{-1} , raw from 7174 to 8941 cm^{-1} , raw from 8949 to 10,715 cm^{-1} , and the D1 from 10,723 to 12,489 cm^{-1} (refer to Figure 7.4h). Figure 7.5h shows the scatter plot of measured versus predicted C_i for ground biomass using the multi-preprocessing PLSR-5 range method.

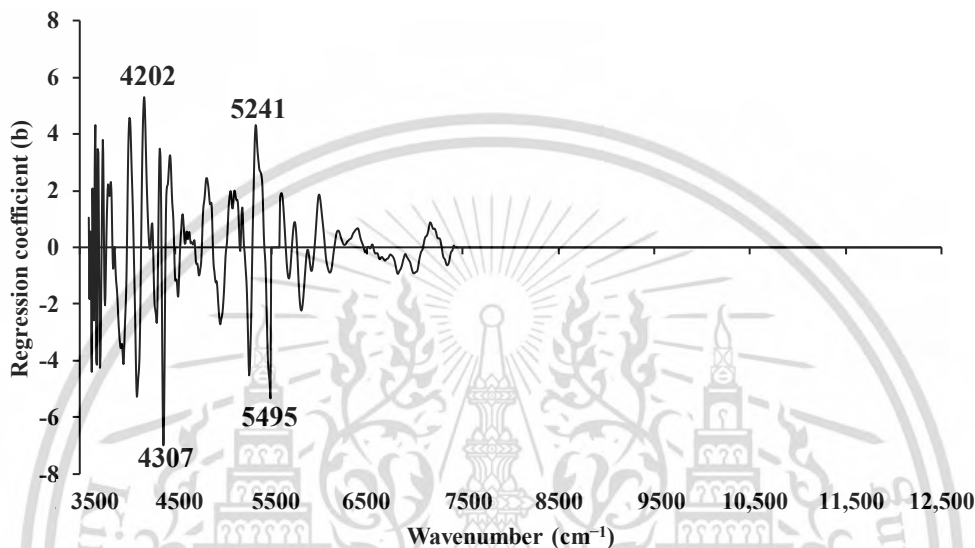


Figure 7.12 The regression coefficient for the C_i of chip biomass using the multi-preprocessing PLSR-3 range method with a spectral preprocessing combination set of 2, 5, 0.

Additionally, with the multi-preprocessing PLSR-5 range method, the model performance for C_i in ground biomass improved by 4.8051% compared to the full-PLSR approach.

Figure 7.13 shows the regression coefficient graph for the C_i of ground biomass. This analysis utilizes the multi-preprocessing PLSR-5 range method with a spectral preprocessing combination set of 5, 0, 4, 2, and 5. Notably, significant peaks are identified at wavenumbers 5495, 5900, 6666, and 6736 cm^{-1} , which are likely to contribute significantly to enhancing the model performance in evaluating C_i within ground biomass. Table 7.5 presents the associated functional groups, spectra-structure, and the material type corresponding to specific peak wavenumbers observed in C_i ground biomass samples [44].

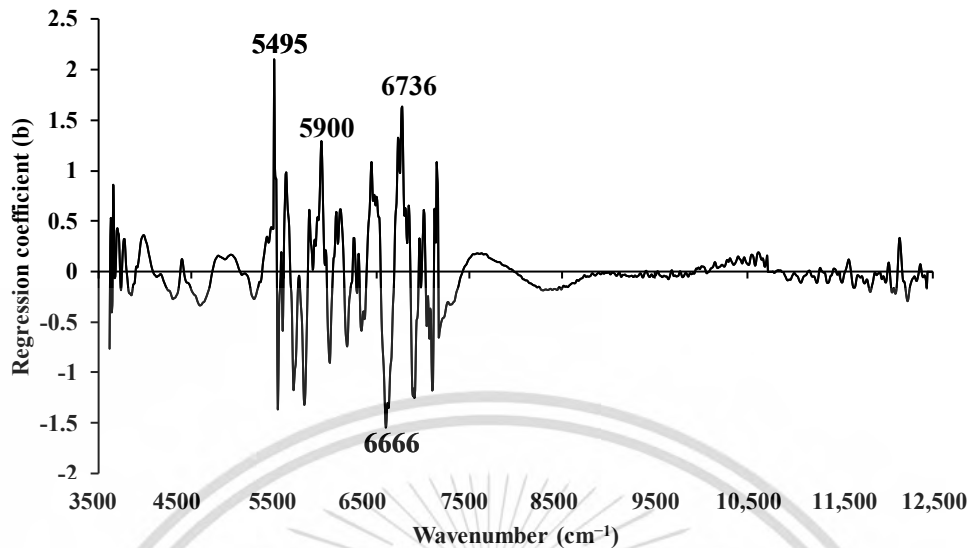


Figure 7.13 The regression coefficient for the C_1 of ground biomass using the multi-preprocessing PLSR-5 range method with a spectral preprocessing combination set of 5, 0, 4, 2, 5.

7.3.4. Comparison with Previous Work

A few studies have assessed the combustion characteristic parameters and performance indices of biomass through TGA. Guohai Jia [13] previously investigated the combustion characteristics of five biomass pellet fuels using TGA at a heating rate of $20\text{ }^\circ\text{C min}^{-1}$. They calculated S_i for masson pine ($1.24 \times 10^{-8}\text{ min}^{-2}\text{ K}^{-3}$), Chinese fir ($2.28 \times 10^{-8}\text{ min}^{-2}\text{ K}^{-3}$), willow ($7.34 \times 10^{-9}\text{ min}^{-2}\text{ K}^{-3}$), slash pine ($5.94 \times 10^{-9}\text{ min}^{-2}\text{ K}^{-3}$), and poplar ($1.83 \times 10^{-8}\text{ min}^{-2}\text{ K}^{-3}$). Shrestha et al. [45] explored the combustion characteristics of *leucaena leucocephala* pellets using TGA at a heating rate of $10\text{ }^\circ\text{C min}^{-1}$, calculating D_i , D_f , and S_i as $6.10 \times 10^{-4}\text{ wt.\% min}^{-3}$, $8.20 \times 10^{-3}\text{ wt.\% min}^{-4}$, and $2.19 \times 10^{-7}\text{ wt.\% min}^{-2}\text{ }^\circ\text{C}^{-3}$, respectively. Similarly, Shrestha et al. [46] evaluated the combustion performance indices for bamboo chips using TGA at a heating rate of $10\text{ }^\circ\text{C min}^{-1}$, deriving D_i , D_f , and S_i values of $88.33 \times 10^{-3}\text{ wt.\% min}^{-3}$, $0.16 \times 10^{-3}\text{ wt.\% min}^{-4}$, and $3.59 \times 10^{-7}\text{ wt.\%}^2\text{ min}^{-2}\text{ }^\circ\text{C}^{-3}$, respectively. The results for combustion characteristic parameters and performance indices vary across different biomass varieties due to distinct heating rates. Consequently, comparisons with previous similar biomass research may lack significance. Furthermore, to date, there have been no reports or publications on the rapid prediction of biomass combustion performance indices using NIRS for comparison.

Following William's guideline [37], if the R^2_p value falls between 0.83–0.90, the model is usable with caution for most applications, including research. For R^2_p values ranging from 0.66 to

0.81, the model can be used for rough screening and other suitable calibration purposes. For R^2_p values ranging from 0.50 to 0.64, the model is only suitable for rough to very rough screening. Following Zornoza et al. [38], any model with an RPD value below 2 is deemed insufficient for any application. If $2.0 < \text{RPD} < 2.5$, it permits approximate prediction. For $2.5 < \text{RPD} < 3$, the model is considered to provide good prediction, and a value higher than 3 represents an excellent model.

According to the recommendation provided by Williams et al. [37], and based on the obtained R^2_p values, along with the consideration of RPD values, as suggested by Zornoza et al. [38], from Tables 7.3 and 7.4, we can conclude that the best models were obtained as follows: for the D_i and C_i of chip biomass, the models were suitable for rough screening, but, when considering the RPD values, they were considered insufficient for any practical applications. For the D_i and C_i of ground biomass, as well as the D_f and S_i of chip biomass, the models were considered acceptable for rough screening and certain other approximate calibrations, based on the obtained R^2_p values. However, when evaluating the RPD values, the models were inadequate for practical applications in the case of the D_i and the D_f models, providing approximate quantitative predictions for the C_i and the S_i models. The best ground biomass models for D_f and S_i can be used with caution for various applications, including research.

7.3.5. Benefit of Combined Agricultural Residue with Fast-Growing Trees in Model

Development

Table 7.1 displays the range of combustion performance indices calculated for both fast-growing trees and agricultural residues, which were utilized in the development of a PLSR-based model. The analysis of Table 7.1 and Figure 7.6 reveals that agricultural residue samples exhibit a broader range in the D_i , D_f , S_i , and C_i . It is evident that the range of each combustion performance index, whether in chip or ground form, expands when agricultural residue samples are incorporated alongside samples from fast-growing trees. This broadening of the range of combustion performance indices is intended to enhance the robustness of the PLSR model for predicting combustion performance indices.

For chip biomass, apart from D_i (Figure 7.5a), the reference value range of D_f , S_i , and C_i (depicted in Figure 7.5b,c,d) in fast-growing trees was narrower compared to that of agricultural residue samples. Integrating agricultural residue samples with fast-growing trees widens their range, resulting in a comparatively enhanced model performance compared to that observed in D_i .

Similarly, concerning ground biomass, Figure 7.5e shows that the range of D_i for fast-growing trees is narrower than that of the agricultural residue sample. The inclusion of agricultural residue samples expanded the range, leading to better model performance compared to chip biomass. In Figure 7.5f, the range of D_f for fast-growing trees was higher and narrower compared to that of the agricultural residue samples. However, the inclusion of agricultural residues samples expanded the range, mostly towards the lower values, contributing to an improved model performance compared to other parameters. Likewise, Figure 7.5g,h illustrate that the range of fast-growing trees is narrower compared to that of the agricultural residue samples. Consequently, the inclusion of agricultural residue samples contributes to expanding the range towards both higher and lower values, ultimately enhancing the model performance.

7.4. Conclusions

The combustion characteristics parameters and combustion performance indices of fast-growing trees and agricultural residues were analyzed through a combined study of TG and DTG curves obtained via TGA. T_i and T_f for fast-growing trees were observed to be higher than those of agricultural residues. This suggests that fast-growing trees were harder to ignite; however, they burnt for a longer duration and produced ash more slowly compared to agricultural residues. While the calculated D_i and D_f were high for fast-growing trees, the S_i and C_i were higher for agricultural residues. This indicates that, even though agricultural residues were easier to ignite and burned more quickly and intensely (exhibiting higher thermal and combustion reactivity), their combustion processes were more controlled and less likely to experience unexpected fluctuations during thermal energy generation.

Similarly, five distinct PLSR-based models were developed and compared using NIRS to assess the D_i , D_f , S_i , and C_i under direct combustion conditions in both chip and ground biomass samples. The models with optimal performance were selected based on higher R^2_c , R^2_p , and RPD values and lower RMSEC, RMSEP, and bias values. The results conclude that the models for D_f and S_i in ground biomass were found to be usable with caution for most applications, including research. All other combustion performance indices, both in chip and ground biomass, were suitable solely for the rough screening purpose. Therefore, a more suitable machine learning algorithm needs to be explored to improve the model performance.

The quality of reference data and spectral data, the inclusion of both agricultural residue samples and fast-growing tree samples to broaden the reference data range, proper identification of outliers, careful selection of the calibration set, and the development and evaluation of models, including spectral pre-treatment and regression methods, all play a pivotal role in establishing a reliable NIR application. Regularly updating calibration and validation procedures, including more representative samples and validating with unknown samples is crucial. Minimizing analytical errors is equally imperative for optimizing the model performance.

This research significantly contributes to the sustainable energy sector and advances our broader understanding of biomass combustion, bridging the gap between research and practical application. With its environmentally friendly behavior, the non-destructive evaluation method by NIR spectroscopy proposed in this study offers an essential and valuable alternative to

traditional thermal destructive techniques, potentially revolutionizing biomass analysis. As NIR models are inherently dynamic, continual improvements and refinements in both experimental methodologies and modeling approach are essential, leading the way for future advancements to be implemented in biomass industries for both production and usage purposes.



References

1. IEA. Greenhouse Gas Emissions from Energy Data Explorer. Available online: <https://www.iea.org/data-and-statistics/data-tools/greenhouse-gas-emissions-from-energy-data-explorer> (accessed on 11 August 2023).
2. Manić, N.; Janković, B.; Stojilković, D.; Radojević, M.; Somoza, B.C.; Medić, L. Self-ignition potential assessment for different biomass feedstocks based on the dynamic thermal analysis. *Clean. Eng. Technol.* **2021**, *2*, 100040.
3. Chen, R.; Sheng, Q.; Dai, X.; Dong, B. Upgrading of sewage sludge by low temperature pyrolysis: Biochar fuel properties and combustion behavior. *Fuel* **2021**, *300*, 121007.
4. Yi, Q.; Qi, F.; Cheng, G.; Zhang, Y.; Xiao, B.; Hu, Z.; Liu, S.; Cai, H.; Xu, S. Thermogravimetric analysis of co-combustion of biomass and biochar. *J. Therm. Anal. Calorim.* **2013**, *112*, 1475–1479.
5. He, Y.; Chen, X.; Tang, X.; Chen, S.; Evrendilek, F.; Chen, T.; Dai, W.; Liu, J. Co-combustion dynamics and products of textile dyeing sludge with waste rubber versus polyurethane tires of shared bikes. *J. Environ. Chem. Eng.* **2023**, *11*, 109196.
6. Wu, X.; Liu, J.; Wei, Z.; Chen, Z.; Evrendilek, F.; Huang, W. Oxy-fuel co-combustion dynamics of phytoremediation biomass and textile dyeing sludge: Gas-to-ash pollution abatement. *Sci. Total Environ.* **2022**, *825*, 153656.
7. Baqain, M.; Neshumayev, D.; Konist, A. TG-MS analysis and kinetic study of co-combustion of ca-rich oil shale with biomass in air and oxy-like conditions. *Carbon Capture Sci. Technol.* **2024**, *10*, 100162.
8. Demirbas, A. Combustion characteristics of different biomass fuels. *Prog. Energy Combust. Sci.* **2004**, *30*, 219–230.
9. Jia, Y.; Li, Z.; Wang, Y.; Wang, X.; Lou, C.; Xiao, B.; Lim, M. Visualization of combustion phases of biomass particles: Effects of fuel properties. *ACS Omega* **2021**, *6*, 27702–27710.
10. Gaba, A.; lordache, S.F. Reduction of air pollution by combustion processes. In *The Impact of Air Pollution on Health, Economy, Environment and Agricultural Sources*; InTech: London, UK, 2011, pp. 119–142.
11. Alves, J.L.F.; da Silva, J.C.G.; Sellin, N.; de Borba Prá, F.; Sapelini, C.; Souza, O.; Marangoni, C. Upgrading of banana leaf waste to produce solid biofuel by torrefaction: Physicochemical

- properties, combustion behaviors, and potential emissions. *Environ. Sci. Pollut. Res.* **2022**, *29*, 25733–25747.
12. Wnorowska, J.; Ciukaj, S.; Kalisz, S. Thermogravimetric analysis of solid biofuels with additive under air atmosphere. *Energies* **2021**, *14*, 2257.
 13. Jia, G. Combustion characteristics and kinetic analysis of biomass pellet fuel using thermogravimetric analysis. *Processes* **2021**, *9*, 868.
 14. Bampenrat, A.; Sukkathanyawat, H.; Seangwattana, T. In *Coal/Biomass Co-Combustion Investigation by Thermogravimetric Analysis, E3S Web of Conferences*; EDP Sciences: Les Ulis, France, 2021; p. 01002.
 15. Liu, J.; Jiang, X.; Cai, H.; Gao, F. Study of combustion characteristics and kinetics of agriculture briquette using thermogravimetric analysis. *ACS Omega* **2021**, *6*, 15827–15833.
 16. Idris, S.S.; Abd Rahman, N.; Ismail, K. Combustion characteristics of Malaysian oil palm biomass, sub-bituminous coal and their respective blends via thermogravimetric analysis (TGA). *Bioresour. Technol.* **2012**, *123*, 581–591.
 17. Sait, H.H.; Hussain, A.; Salema, A.A.; Ani, F.N. Pyrolysis and combustion kinetics of date palm biomass using thermogravimetric analysis. *Bioresour. Technol.* **2012**, *118*, 382–389.
 18. Cao, W.; Li, J.; Lue, L. Study on the ignition behavior and kinetics of combustion of biomass. *Energy Procedia* **2017**, *142*, 136–141.
 19. Lu, J.-J.; Chen, W.-H. Investigation on the ignition and burnout temperatures of bamboo and sugarcane bagasse by thermogravimetric analysis. *Appl. Energy* **2015**, *160*, 49–57.
 20. Arromdee, P.; Ninduangdee, P. Combustion characteristics of pelletized-biomass fuels: A thermogravimetric analysis and combustion study in a fluidized-bed combustor. *Energy Ecol. Environ.* **2023**, *8*, 69–88.
 21. Iryani, D.A.; Rakaseri, I.; Azhar, A.; Haryanto, A.; Hidayat, W.; Hasanudin, U. Thermogravimetric assessment for combustion characteristic of torrefied pellet biomass from agricultural solid waste. *IOP Conf. Ser. Earth Environ. Sci.* **2023**, *1187*, 012019.
 22. Luthfi, N.; Ohkoshi, T.; Tamaru, Y.; Fukushima, T.; Takisawa, K. Investigation into the combustion kinetics and spontaneous ignition of sweet sorghum as energy resource. *Bioresour. Bioprocess.* **2022**, *9*, 49.

23. El-Sayed, S.A.; Mostafa, M.E.; Khass, T.M.; Noseir, E.H.; Ismail, M.A. Combustion and mass loss behavior and characteristics of a single biomass pellet positioning at different orientations in a fixed bed reactor. *Biomass Convers. Biorefinery* **2023**, *2023*, 1–21.
24. Guo, Q.; Cheng, Z.; Chen, G.; Yan, B.; Hou, L.A.; Ronsse, F. Optimal strategy for clean and efficient biomass combustion based on ash deposition tendency and kinetic analysis. *J. Clean. Prod.* **2020**, *271*, 122529.
25. Yuan, Y.; He, Y.; Tan, J.; Wang, Y.; Kumar, S.; Wang, Z. Co-Combustion characteristics of typical biomass and coal blends by thermogravimetric analysis. *Front. Energy Res.* **2021**, *9*, 753622.
26. Chen, G.-B.; Li, J.-W.; Lin, H.-T.; Wu, F.-H.; Chao, Y.-C. A study of the production and combustion characteristics of pyrolytic oil from sewage sludge using the taguchi method. *Energies* **2018**, *11*, 2260.
27. Posom, J.; Shrestha, B.; Maraphum, K.; Pitak, L.; Saengprachatanarug, K.; Sirisomboon, P.; Shrestha, B.P. Near-Infrared Spectroscopy, Hyperspectral, Multispectral Imaging Principles and Applications in Energy Properties of Biomass. In *A Guide to Near-Infrared Spectroscopy*, 1st ed.; Garcia Martin, J.F., Ed.; Nova Science Publishers, Inc.: Hauppauge, NY, USA, 2023.
28. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Funke, A. Effect of Combined Non-Wood and Wood Spectra of Biomass Chips on Rapid Prediction of Ultimate Analysis Parameters Using near Infrared Spectroscopy. *Energies* **2024**, *17*, 439.
29. Reza, M.S.; Taweekun, J.; Afroze, S.; Siddique, S.A.; Islam, M.S.; Wang, C.; Azad, A.K. Investigation of Thermochemical Properties and Pyrolysis of Barley Waste as a Source for Renewable Energy. *Sustainability* **2023**, *15*, 1643. <https://doi.org/10.3390/su15021643>.
30. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Pornchaloempong, P.; Funke, A. NIR Spectroscopy as an Alternative to Thermogravimetric Analyzer for Biomass Proximate Analysis: Comparison of Chip and Ground Biomass Models. *Energies* **2024**, *17*, 800.
31. Yi, B.; Chen, M.; Gao, Y.; Cao, C.; Wei, Q.; Zhang, Z.; Li, L. Investigation on the co-combustion characteristics of multiple biomass and coal under O₂/CO₂ condition and the interaction between different biomass. *J. Environ. Manag.* **2023**, *325*, 116498.
32. Vamvuka, D.; Loukakou, E.; Avgoustidis, C.; Stratakis, A.; Pavloudakis, F.; Sfakiotakis, S. Co-combustion characteristics of lignite/woody biomass blends. Reactivity and fusibility assessment. *Energy Sources Part A Recovery Util. Environ. Eff.* **2023**, *45*, 3916–3930.

33. Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* **2023**, *16*, 5351.
34. Stolov, A.A.; Simoff, D.A.; Li, J. Thermal stability of specialty optical fibers. *J. Light. Technol.* **2008**, *26*, 3443–3451.
35. Shrestha, B.; Shrestha, Z.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Evaluating limit of detection and quantification for higher heating value and ultimate analysis of fast-growing trees and agricultural residues biomass using NIRS. *Eng. Appl. Sci. Res.* **2023**, *50*, 612–618.
36. 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. *Eng. Appl. Sci. Res.* **2022**, *49*, 119–126. Available online: <https://ph01.tci-thaijo.org/index.php/easr/article/view/245217> (accessed on 28 December 2023).
37. Williams, P.; Manley, M.; Antoniszyn, J. *Near Infrared Technology: Getting the Best Out of Light*; African Sun Media: Stellenbosch, South Africa, 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 Biol. Biochem.* **2008**, *40*, 1923–1930. <https://doi.org/10.1016/j.soilbio.2008.04.003>.
39. 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.
40. Madhu, P.; Dhanalakshmi, C.S.; Mathew, M. Multi-criteria decision-making in the selection of a suitable biomass material for maximum bio-oil yield during pyrolysis. *Fuel* **2020**, *277*, 118109.
41. Cardarelli, A.; Pinzi, S.; Barbanera, M. Effect of torrefaction temperature on spent coffee grounds thermal behaviour and kinetics. *Renew. Energy* **2022**, *185*, 704–716.

42. Brassard, P.; Palacios, J.H.; Godbout, S.; Bussi eres, D.; Lagac e, R.; Larouche, J.-P.; Pelletier, F. Comparison of the gaseous and particulate matter emissions from the combustion of agricultural and forest biomasses. *Bioresour. Technol.* **2014**, *155*, 300–306.
43. Mansora, A.M.; Lima, J.S.; Anib, F.N.; Hashima, H.; Hoa, W.S. Characteristics of cellulose, hemicellulose and lignin of MD2 pineapple biomass. *Chem. Eng.* **2019**, *72*, 79–84.
44. Workman, J., Jr.; Weyer, L. *Practical Guide to Interpretive Near-Infrared Spectroscopy*; CRC Press: Boca Raton, FL, USA, 2007.
45. Shrestha, A.; Saechua, W.; Sirisomboon, P. Some physical and combustion characteristic of *Leucaena Leucocephala* pellet. In Proceedings of the 16th TSAE National Conference and 8th TSAE International Conference, Agricultural and Food Engineering Outlook on Sustainable Future, ET-08, Bangkok International Trade & Exhibition Centre, Bangkok, Thailand, 17–19 March 2015; pp. 127–132.
46. Shrestha, A. *Feasibility Study on Near Infrared Spectroscopy for Evaluation of Combustion Performance Parameters and Moisture Content of Bamboo Chips (Dendrocalamus sericeus cl. Phamon)*; King Mongkut’s Institute of Technology: Ladkrabang, Thailand, 2016.

Chapter 8

Conclusion and recommendations

8.1 Conclusion

In this study, five different PLSR-based models were developed using spectral data from both chipped and ground biomass acquired through FT-NIRS. These models incorporated HHV (J/g) values derived from a bomb calorimeter and ultimate analysis (wt.% of C, H, N, and O) performed using a CHNO/S elemental analyzer. Proximate analysis (% of MC, VM, FC, and A) was assessed using TG curves, while combustion performance indices (D_i , D_f , S_i , and C_i) were evaluated using combined TG and DTG curves obtained from TGA. The effectiveness of these models was quantified using parameters such as R^2_C , R^2_P , RMSEC, RMSEP, RPD, and bias. Optimal models selected from either chipped or ground biomass demonstrated high R^2_C , R^2_P , and RPD values with lower RMSEC and RMSEP.

The HHV model from ground biomass, developed using GA-PLSR with the first derivative, is suitable for a wide range of applications, including research. Models predicting the wt.% of C and H from ground biomass, derived using GA-PLSR using the first derivative and SNV respectively, as well as the wt.% of O from chipped biomass derived using MP-PLSR with a 5 range, are primarily intended for preliminary screening. The model for the wt.% of N from ground biomass, developed using MP-PLSR with a 5 range, is generally reliable but should be used with caution. Models determining the % MC from GA-PLSR with the second derivative and the % FC from Full-PLSR with constant offset, both derived from chipped biomass, showed satisfactory performance and are useful in various contexts, including research. The model for the % VM from chipped biomass, developed using MP-PLSR with a 5 range, is only suitable for initial assessments. The model for % ash from ground biomass, obtained from Full-PLSR with the first derivative, is appropriate for preliminary screenings. All the optimal models for combustion performance indices were derived from ground biomass. The models for D_i and C_i using MP-PLSR with a 5 range are suitable only for preliminary screening, while those for D_f from GA-PLSR with the first derivative and S_i from Full-PLSR with MSC are appropriate for broader applications, including research.

Furthermore, a correlation analysis using SPSS 29.0 was conducted to assess the relationship between NIRS and conventional destructive methods. The results showed strong

positive correlations for various parameters, including HHV, C, H, O, N, MC, FC, VM, A, D_i, D_f, C_i, and S_i. The Pearson correlation coefficients were notably high, ranging from 0.838 to 0.976. Specifically, the coefficients were 0.976, 0.882, 0.927, 0.893, 0.930, 0.950, 0.952, 0.925, 0.889, 0.838, 0.918, 0.949, and 0.869, respectively, with associated p-values less than 0.01. These results indicate highly significant correlations and support the validity of using NIRS as an alternative method for evaluating a range of properties in fast-growing trees and agricultural residues, as hypothesized.

This study introduced novel multi-preprocessing 5-range and 3-range methods, applying different spectral preprocessing techniques to specific parts of the NIR spectrum to improve the accuracy of PLSR models. This approach outperformed traditional methods that use a single preprocessing technique across the entire spectrum, indicating that segment-specific preprocessing can enhance model accuracy. The study further explored the impact of including various biomass species (non-wood and wood) in the model samples using scatter plot analysis. It concluded that for robust model performance, not only should the species vary in constituent values, but they must also exhibit a consistent rate of change between NIR-predicted and measured values, aiming for a slope close to 1 with minimal intercept.

In summary, the recommended models provide effective, non-destructive alternative methods for rapidly evaluating the energy properties and combustion characteristics of biomass for energy usage. For enhanced robustness and practicality, it is crucial to expand these models with diverse sample data, validate them with unknown samples, and explore various machine learning algorithms. Additionally, ensuring the quality of reference data, including a variety of sample types such as agricultural residues and fast-growing trees, accurately identifying outliers, and carefully selecting the calibration set are essential for the reliable application of NIR technology and the development of a more comprehensive global model for predicting biomass energy properties. These models have the potential to reduce economic disparities among biomass traders and support stakeholders, including academic institutions, policymakers, industries, and energy companies, in efficiently adopting and managing bio-resources to meet future energy demands.

8.2 Recommendations

Based on the findings and experiences from this study, the following recommendations are proposed to further advance research in biomass energy properties and enhance their practical applications. Firstly, there is potential for improvement in the current models for C, H, VM, ash, D_i , and C_i , which demonstrate satisfactory performance. Future research should explore various alternative machine learning algorithms to enhance the accuracy of these prediction models and conduct comparative analyses to identify the most effective approaches. Additionally, to ensure reliability and robustness, it is crucial to periodically validate these models using sets of unknown samples.

In this study, PLSR-based models were developed for ground biomass using the transmittance mode of FT-NIRS scanning to predict various biomass energy properties. Future research should explore using diffuse reflectance with sphere macro sample rotating scanning mode for ground biomass, maintaining the same resolution, background, and sample scans time, with absorbance data logged as $\log(1/R)$, similar to that used for chip biomass. This approach will enable a better comparison of the PLSR model performance between ground and chip biomass. Such comparative analysis will facilitate the selection of a better model for industrial application.

The traditional method of proximate analysis using TGA is sensitive, costly, time-consuming, and requires skilled operation. Therefore, it is recommended to conduct research that explores the relationships between higher heating value (HHV), ultimate analysis parameters, and proximate analysis parameters. Developing calibration models based on spectral data obtained from FT-NIRS and calculated proximate reference data can facilitate rapid and non-destructive prediction of biomass composition.

Furthermore, integrating machine learning with NIRS to estimate the kinetic triplet of combustion for chipped and ground biomass would extend the utility of NIR technology in biomass research, providing deeper insights into combustion kinetics.

As the global shift towards renewable energy intensifies, particularly within the bioenergy sector, there is a significant opportunity to focus research on the conversion of raw biomass into high-energy content fuels such as pellets. Studying the rapid prediction of energy properties of biomass pellets using NIR technology or hyperspectral imaging could greatly impact the biomass industry, especially in the energy and industrial heating sectors. These recommendations aim to

leverage new technological approaches to meet the evolving demands of the renewable energy sector and contribute to the efficient utilization and management of bio-resources, aligning with worldwide goals for sustainable energy use.



Appendix



Author Biography

Name-Surname	Bijendra Shrestha
Date of birth	28 February 1988
Contact address	Dhulikhel-4, Kavrepalanchok, Bagmati Province, Nepal
Email address	bijendra.shrestha@ku.edu.np
Contact number	+977-9860040628

Education background

D.Eng. (Food and Agricultural Intelligence Engineering) from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand (2021 January – 2024 June, GPA: 4.00)

M. S. by Research (Mechanical Engineering) from Kathmandu University, Dhulikhel, Kavrepalanchok, Bagmati, Nepal (2010 November – 2013 August, GPA: 3.92)

B.Eng. (Mechanical Engineering) from Kathmandu University, Dhulikhel, Kavrepalanchok, Bagmati, Nepal (2006 August – 2010 June, GPA: 3.23)

Training attended

- Workshop on “New and Renewable Energy (Biogas)” held from 20-24 October 2014, Kunming, China organized by Michigan State University, USA and Yunnan Academy of Scientific and Technical Information, China.
- Training on “Market Research and Customer Insights” held from 9-12 March 2016 in Panaruban, Indonesia organized by Siemens Stiftung, Germany.
- Japan-Asia Youth Exchange Program in Science (SAKURA Exchange Program in Science), held from 6-15 March 2017 in Tokyo University of Agriculture, Japan.
- Senior Manager Level Training Program for the project title “Integrated Rural Development of Nepal through Strengthening Research and Development Capacity of Kathmandu University” held from 12-23 June 2017 in South Korea organized by Korean International Cooperation Agency (KOICA), Korea and Science and Technology Policy Institute (STEPI), Korea.

Scholarships

- King Mongkut's Institute of Technology Ladkrabang (KMITL) doctoral scholarship, Thailand [Grant number KDS 2020/052].
- Erasmus Mundus Mobility with Asia West scholarship, and was enrolled at Lucian Blaga University of Sibiu, Romania, "Hermann Oberth" Faculty of Engineering, for 6 months (September 2011 – February 2012), within the frame of Erasmus Mundus External Cooperation Windows Program.

Oral presentation

- Shrestha, B.; Posom, J.; Shrestha, B.P.; Sirisomboon, P.; Saechua, W. Precision Test of NIR Spectroscopy and Bomb Calorimeter for Evaluation of Higher Heating Value on Different Varieties of Nepal Biomass Pellets. 15th Thai Society of Agricultural Engineering International Conference (TSAE 2022) at Thailand by online system. August 18-19, 2022.
- Shrestha, B.; Posom, J.; Shrestha, B.P.; Sirisomboon, P. Upgrading the Model Performance for Rapid Prediction of the Ash Content of Biomass Using NIR spectroscopy. The 8th Asian NIR Symposium at South Korea by online system. November 28-30, 2022.
- Shrestha, B.; Shrestha, Z.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Assessment of Limit of Detection and Quantification in Predictive Modeling of Higher Heating Value and Ultimate Analysis for Nepal Biomass Using Near-Infrared Spectroscopy. 16th Thai Society of Agricultural Engineering International Conference (TSAE 2023) at Oakwood hotel and residence Sri Racha, Chonburi, Thailand. July 26-27, 2023.
- Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Shrestha, Z.; Pornchaloempong, P. A Rapid and Nondestructive Prediction of Biomass Higher Heating Value Using FT-NIR spectroscopy and Partial Least Squares Regression. 17th Thai Society of Agricultural Engineering International Conference (TSAE 2024) at Bangkok International Trade & Exhibition Centre (BITEC), Thailand May 22-24, 2024.

International Published papers

- Shrestha, B.; Posom, J.; Pornchaloempong, P.; Sirisomboon, P.; Shrestha, B.P.; Ariffin, H. Near-infrared Spectroscopy Modeling of Combustion Characteristics in Chip and Ground Biomass from Fast-Growing Trees and Agricultural Residue. *Energies* 2024, 17,1338. 800. <https://doi.org/10.3390/en17061338>
- Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Pornchaloempong, P.; Funke, A. NIR Spectroscopy as an Alternative to Thermogravimetric Analyzer for Biomass Proximate Analysis: Comparison of Chip and Ground Biomass Models. *Energies* 2024, 17, 800. <https://doi.org/10.3390/en17040800>
- Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P.; Funke, A. Effect of Combined Non-Wood and Wood Spectra of Biomass Chips on Rapid Prediction of Ultimate Analysis Parameters Using Near Infrared Spectroscopy. *Energies* 2024, 17, 439. <https://doi.org/10.3390/en17020439>
- Shrestha, B.; Shrestha, Z.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Evaluating limit of detection and quantification for higher heating value and ultimate analysis of fast-growing trees and agricultural residues biomass using NIRS. *Engineering and Applied Science Research* 2023, 50 (6), 612-618. <https://ph01.tci-thaijo.org/index.php/easr/article/view/253649>
- Shrestha, B.; Posom, J.; Sirisomboon, P.; Shrestha, B.P. Comprehensive Assessment of Biomass Properties for Energy Usage Using Near-Infrared Spectroscopy and Spectral Multi-Preprocessing Techniques. *Energies* 2023, 16, 5351. <https://doi.org/10.3390/en16145351>

Work experience

- Researcher (September 2010 – September 2013) at Department of Mechanical Engineering, School of Engineering, Kathmandu University, Dhulikhel, Nepal.
- Lecturer (10 September 2013 – 11 July 2019) at Department of Mechanical Engineering, School of Engineering, Kathmandu University, Dhulikhel, Nepal.

- Assistant Professor (12 July 2019 – till date) at Department of Mechanical Engineering, School of Engineering, Kathmandu University, Dhulikhel, Nepal.

Areas of expertise and interest

- Application of near infrared spectroscopy
- Renewable Energy (biomass)
- Waste to energy technology

Reference persons

- Prof. Dr. Bim Prasad Shrestha, Department of Mechanical Engineering, School of Engineering, Kathmandu University, Nepal
Email: shrestha@ku.edu.np
- Prof. Dr. Panmanas Sirisomboon, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology, Thailand
Email: panmanas.si@kmitl.ac.th
- Associate Prof. Dr. Jetsada Posom, Department of Agricultural Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
Email: jetspo@ku.edu.np