International Journal of Technology

http://ijtech.eng.ui.ac.id

Research Article

Low-Cost Near-Infrared Spectroscopy for Rapid Prediction of Biodiesel Properties: Acid Value, Density, Viscosity, and Water Content

Kittisak Phetpan[1], **Chitwadee Thongphut**[1], **Thatchapol Chungcharoen**[1*], **Warunee Limmun**[1]

[1]Department of Engineering, King Mongkut's Institute of Technology Ladkrabang, Prince of Chumphon Campus, Chumphon 86160, Thailand

[*]Corresponding author: thatchapol.ch@kmitl.ac.th; Tel.: +66-970899785; Fax.: +66-77506433

Abstract: Partial least squares (PLS) regression, combined with various spectral pre-processing techniques, was employed to compare the performance of two diode array near-infrared (NIR) spectrometers in predicting key biodiesel quality parameters, including acidity, viscosity, density, and water content. An AvaSpec-Mini4096CL NIR spectrometer, operating within the 350–1100 nm wavelength range, was used as the representative shortwave near-infrared (SW-NIR) spectrometer, while a NIRQuest512 spectrometer, covering the 900–1700 nm range, was employed as the longwave near-infrared (LW-NIR) spectrometer. Both spectrometers were equipped with a transflection probe for spectral collection from oil palm-based biodiesel samples. The SW-NIR spectrometer outperformed the LW-NIR spectrometer. The optimal PLS models achieved root mean square errors of prediction (RMSEP) of 0.0037 mg KOH/g for acidity, 0.062 cSt (mm^2/s) for viscosity, 2.67 kg/m^3 for density, and 59.14 mg/kg for water content, highlighting the potential of compact SW-NIR spectrometers as effective, low-cost tools for rapid biodiesel quality monitoring.

Keywords: Biodiesel properties; Diode array spectrometer; Multivariate calibration; Near-infrared spectroscopy (NIRS)

1. Introduction

Increasing demand for diesel fuel has led to a growing awareness of alternative energy sources that are environmentally friendly and can reduce production costs (Oliveira et al., 2007). Biodiesel fuel has gained widespread attention as a renewable alternative to petroleum-derived diesel. Biodiesel can be produced from various sustainable feedstocks, including palm oil, vegetable oil, animal fat, algae, and used cooking oil (Cunha et al., 2020; Balabin and Smirnov, 2011; Balabin et al., 2011; Baptista et al., 2008; Felizardo et al., 2007). The production process involves transesterification, a chemical reaction in which biodiesel is synthesized by reacting oils or fats with alcohols such as methanol or ethanol, yielding esters that closely resemble diesel fuel in their properties (Wahyono et al., 2022; Sajjadi et al., 2017).

Biodiesel can be used effectively in diesel engines, but its composition differs significantly from that of conventional diesel. International standards, such as the American Society for Testing and Materials (ASTM D6751) and the European Standard (EN 14214) have been established to define the required characteristics and quality specifications for biodiesel. Quality control is essential to ensure that biodiesel meets these standards, as impurities can significantly affect engine performance (Balabin and Smirnov, 2011; Liu et al., 2008; Berchmans and Hirata, 2008; Rashid and Anwar, 2008; Sarin et al., 2007). Key quality parameters—including acid value, kinematic viscosity, density, and water content—strongly influence fuel stability, storage behavior, engine durability, and combustion performance. Various production technologies influence the final properties of biodiesel, making quality assessment a critical aspect of its commercial-

https://doi.org/10.14716/ijtech.v17i2.8331
Received December 2025; Revised January 2026; Accepted February 2026

ization (Meher et al., 2006; Bournay et al., 2005; Abreu et al., 2004). Conventional reference methods for determining these properties are well established and reliable; however, they are laboratory-based, time-consuming, reagent-intensive, and unsuitable for rapid decision-making during production (Bournay et al., 2005). As biodiesel manufacturing scales toward industrial production, the need for faster and more practical quality monitoring approaches is increasing.

Quality assurance and quality control (QA/QC) systems in modern biodiesel plants are progressively transitioning toward inline and at-line monitoring strategies to enable real-time process control. Rapid feedback during the transesterification, washing, and storage stages allows for the early detection of off-specification products and improves process efficiency. Near-infrared spectroscopy (NIRS) offers rapid, nondestructive, and reagent-free analysis and can be integrated into industrial environments to support continuous monitoring rather than merely serving as a faster laboratory test. Because near-infrared (NIR) spectra capture overtone and combination vibrations of molecular groups, such as O–H and C–H bonds, multivariate calibration models can extract spectral information correlated with biodiesel's physicochemical properties (Bukkarapu and Krishnasamy, 2022; Balabin et al., 2011; Balabin and Smirnov, 2011; Monteiro et al., 2009; Balabin and Safieva, 2007; Yang et al., 2003). Spectroscopic techniques integrated with multivariate modeling approaches have also been applied in engineering and environmental monitoring systems, demonstrating the broader applicability of optical absorption methods for quantitative analysis and process optimization (Naghipour et al., 2024; Kuriyama et al., 2011).

Numerous studies have demonstrated the feasibility of NIRS for biodiesel characterization (Cunha et al., 2020; Cunha et al., 2017; Balabin and Smirnov, 2011; Baptista et al., 2008; Balabin and Safieva, 2007; Felizardo et al., 2007). However, most previous investigations (Hradecká et al., 2023; Cunha et al., 2020; Balabin et al., 2011; Balabin and Smirnov, 2011) relied predominantly on Fourier-transform Near-infrared (FT-NIR) spectrometers operating over wide spectral ranges. While FT-NIR systems provide high spectral resolution and sensitivity, they are relatively expensive and unsuitable for widespread industrial deployment. Consequently, despite the demonstrated capability of NIRS for biodiesel analysis, its translation into cost-effective, production-oriented QA/QC systems remains limited.

Diode-array NIR spectrometers offer several practical advantages, including compact design, high acquisition speed, mechanical robustness, and low instrument cost. These characteristics make them attractive candidates for monitoring the quality of inline and at-line biodiesel. Nevertheless, systematic evaluations of diode-array NIR systems for comprehensive prediction of biodiesel properties remain scarce. In particular, there are limited comparative investigations between shortwave (SW, silicon detector, 350–1100 nm) and longwave (LW, InGaAs detector, 900–1700 nm) diode-array configurations. Furthermore, although previous studies have extensively examined properties such as viscosity and density—primarily using FT-NIR systems—the prediction of acid value using low-cost diode-array instruments has rarely been reported. Although NIRS is not new to biodiesel characterization, the practical potential of compact and affordable diode array systems for multi-property prediction remains underexplored.

To clarify the mechanistic basis of NIR-based property prediction, the translation of NIR spectra into physicochemical properties relies on molecular–spectral relationships. The NIR absorption bands primarily arise from the overtone and combination vibrations of the C–H and O–H bonds (Williams, 2019; Walsh et al., 2020). In biodiesel, the acid value is associated with free fatty acids containing carboxylic (–COOH) groups, whose O–H stretching overtones contribute to absorption features around 1400–1500 nm. Similarly, the water content exhibited strong O–H-related absorption in this region, enabling the sensitive detection of moisture variations.

Although viscosity and density are macroscopic bulk properties rather than direct spectral absorbers, they are governed by molecular weight distribution, carbon chain length, and degree of unsaturation (Kumbhar et al., 2022; Cunha et al., 2020; Knothe, 2005). These structural variations influence C–H stretching overtone regions within the measurable spectral win-

dows. Because molecular composition simultaneously determines both spectral fingerprints and bulk physicochemical behavior, multivariate calibration models, such as partial least squares regression (PLSR), can establish quantitative relationships between NIR spectra and laboratory reference measurements.

Therefore, this study does not introduce NIRS for biodiesel characterization for the first time; rather, it advances previous research by directly addressing the identified limitations. Specifically, most existing biodiesel NIR studies depend on high-cost FT-NIR instruments and provide limited evidence regarding the performance of low-cost diode array systems under industrial QA/QC conditions. Additionally, comparative evidence between SW and LW diode-array configurations is scarce, and the systematic evaluation of acid value prediction using such systems is limited. To fill these gaps, this work systematically compares SW (350–1100 nm) and LW (900–1700 nm) diode-array spectrometers using the same biodiesel sample set, standardized reference measurements, consistent preprocessing, and PLS regression modeling. This study provides the missing comparative evidence necessary for informed instrument selection by quantifying and contrasting predictive performance across two detector technologies and demonstrates the practical feasibility of implementing compact NIR systems for inline and at-line biodiesel quality monitoring.

2. Materials and Methods

2.1 Acquisition of biodiesel sample and reference (biodiesel properties) data

A total of 108 palm oil-based biodiesel fuel samples were obtained from the washing process using biochar, based on 54 conditions with two replicates. Prior to spectral acquisition, all biodiesel samples were filtered to remove visible particulates and suspended impurities. The samples were obtained from the same washing process, ensuring comparable cleanliness and minimizing variation due to extraneous contaminants. The measurements were performed under controlled laboratory conditions using the same transflection probe and optical path length. Therefore, spectral variability was primarily associated with biodiesel compositional differences rather than external contamination factors.

All NIR spectral acquisitions and ASTM reference measurements were performed using the same biodiesel samples obtained from the same production batch. To minimize storage-related variability, spectral measurements and reference analyses were conducted within the same experimental period. Samples were stored in sealed, opaque containers at controlled room temperature (25 ± 1 °C) and analyzed within 48 h after collection. This procedure ensured that potential shifts in biodiesel properties due to oxidation, moisture absorption, or prolonged storage were minimized.

The reference data were determined according to the ASTM standards. The physicochemical properties of biodiesel, including acid value, kinematic viscosity, density, and water content, were determined using standard laboratory methods. Acid value was measured according to ASTM D664 via potentiometric titration, in which biodiesel samples were titrated with a standardized potassium hydroxide (KOH) solution to quantify free fatty acid content, expressed as mg KOH per gram of biodiesel. Kinematic viscosity was determined following ASTM D445 using a calibrated viscometer at 40 °C, ensuring temperature-controlled conditions throughout the measurement (Mishra et al., 2020). Density was measured at 15°C in accordance with ASTM D1298 using an auto pipette (Hoang, 2021). The water content was quantified using Karl Fischer titration according to ASTM D6304 (Delfino et al., 2018), which provides sensitive moisture detection in biodiesel samples. All measurements were performed in triplicate, and the average values were used as reference data for subsequent chemometric modeling.

2.2 Near-infrared spectroscopy (NIRS) instrumentation and spectral data collection

NIR instruments, such as spectrophotometers or spectrometers, are often used interchangeably. Although these instruments may differ in physical size, appearance, and scanning mode, their wavelength selection systems and detectors are the primary distinctions (Yuan et al., 2024). Several wavelength selection systems are available, including discrete narrow-band-pass filters, acousto-optic tunable filters (AOTF), NIR light-emitting diodes (NIR LEDs), NIR diode lasers, grating monochromators with diode array detectors, and interferometers. Among these, diode array and AOTF-driven instruments are the most suitable for applications requiring continuous analysis (online measurement) because of their high-speed capabilities and fast analysis. However, diode array instruments are inexpensive while maintaining comparable efficiency.

Commercial instruments typically use silicon, indium gallium arsenide (InGaAs), or lead sulfide (PbS), which cover wavelength ranges of 380–1100, 900–1700, and 1100–2500 nm, respectively. Some instruments combine silicon and PbS or use extended InGaAs detectors to cover the entire wavelength range of 380–2500 nm. However, the incorporation of multiple detector types significantly increases instrument costs. Therefore, diode array instruments equipped with InGaAs detectors (900–1700 nm) are highly attractive and widely used (Williams, 2019).

According to the section on NIR instrumentation, diode array spectrometers are suitable when very high-speed capabilities, fast analysis (real-time measurement), and low instrument cost are required. This study evaluated two diode array spectrometers operating in different wavelength ranges (350–1100 nm versus 900–1700 nm) to compare their performance in analyzing biodiesel properties. A transflection-based optical fiber probe, which combines transmission and reflection modes, was used for spectral measurements because it is particularly effective for liquid samples.

A NIR spectrometer (AvaSpec-Mini4096CL, Avantes BV, Netherlands), which operates in the wavelength range of 350–1100 nm, was used as a representative shortwave and near-infrared (SW-NIR) spectrometer in this research. This low-cost, high-speed instrument uses a diode array for wavelength selection. Before sample analysis, the spectrometer was configured by scanning the reference spectrum in deionized (DI) water using the transflection probe. The integration time was set to 2.07 ms, with an average of 100 scans per sample. For comparison, a longwave near-infrared (LW-NIR) spectrometer (NIRQuest 512, Ocean Optics, USA) covering the 900–1700 nm range was used. The same probe and DI water reference were employed with an integration time of 150 ms and five scans per sample to obtain the average spectrum.

The selection of the spectral ranges (350–1100 nm for SW-NIR and 900–1700 nm for LW-NIR) was determined by the detector types integrated within each diode-array spectrometer. The SW-NIR instrument uses a silicon detector, which is sensitive in the visible to shortwave near-infrared region, whereas the LW-NIR instrument employs an indium gallium arsenide (InGaAs) detector. These spectral regions encompass important overtone and combination absorption bands associated with O–H and C–H functional groups, which are directly related to biodiesel properties such as acid value and water content (O–H) and viscosity and density (C–H). The overlapping region (900–1100 nm) allows partial comparison between the two systems, whereas the extended coverage of the LW-NIR instrument captures higher-order overtones beyond 1100 nm. The comparative evaluation of these two detector-based spectral windows enables the assessment of whether the broader longwave region provides additional predictive advantage over the lower-cost shortwave configuration.

The NIR spectra were collected from the biodiesel samples at room temperature (25 ± 1 °C). Each sample was scanned in triplicate, yielding three spectra per sample. In total, 324 spectra were obtained from 108 samples. Throughout the study, a transflection probe with a 5 mm optical path length and 300 µm core-optic fiber was used for spectral collection.

2.3 Spectral pre-processing

Physical phenomena in spectra, such as noise, baseline deviation, and negligible overlapping areas not explained by the concentration of interest, often lead to suboptimal models. Thus, NIR spectral data preprocessing can enhance interpretation, reduce model complexity, and improve robustness against undesirable variations in new samples (Ruttanadech et al., 2023; Cunha et al., 2020; Phetpan et al., 2018; Balabin and Smirnov, 2011). It has become an integral part of chemometric modeling based on NIR.

The moving-average (MA) filter is the simplest smoothing technique, employing a predefined window that moves successively along equally spaced spectral data. All data points are equally weighted. The application of the MA filter reduces noise in the spectral data. However, the window size (filter width) must be optimized to achieve effective smoothing. The original height of a spectral peak decreases if the filter width is too large, leading to simultaneous broadening (Otto, 2017).

Mean centering (MC) is a pre-processing technique that eliminates constant offsets in spectral data. Each x-variable at each wavelength is centered by subtracting the column mean. As a result, each variable has a mean of zero, making the differences in relative intensities more noticeable (Cunha et al., 2020; Otto, 2017; Naes et al., 2002).

The orthogonal signal correction (OSC) technique is applied to separate variances that are not correlated with the variation of the y-variable (concentration of interest). This technique suppresses non-relevant information in the y-response within matrix X (spectral data). As a result, the information orthogonal to the response is removed from the original spectral data, allowing the valuable calibration information to remain largely (Laref et al., 2017; Fearn, 2000).

A derivative is defined as the slope of a line at any given point. The first derivative can remove an additive baseline shift in the NIR spectra, which tends to have linear baseline increases, while the second derivative can further eliminate such trends. However, taking derivatives increases noise, so averaging over segments during computation can help mitigate this issue. Derivative analysis is a powerful technique for revealing overlapping peaks in a spectra. Additionally, the derivatives can be computed using the Savitzky–Golay method, which fits a curve through a small spectrum section. Then, the slope of the tangent to this curve at the central point is determined. This process is applied as a moving window across the entire spectrum (Davies, 2007).

Standard normal variate (SNV) is a scaling-based preprocessing approach that processes each spectrum individually without requiring a reference spectrum. SNV is commonly used to eliminate scattering effects in NIR spectra, thereby improving prediction accuracy. However, it does not reduce systematic interference in spectral data (Cunha et al., 2020; Phetpan et al., 2018).

2.4 Data analysis and development of calibration model

Simple models were constructed to describe the relationship between the y-response and an independent x-variable. Multiple independent variables are required for calibration in multivariate modeling. Among the various regression methods, partial least squares regression (PLSR) is widely applied, particularly in NIRS applications (Udompetaikul et al., 2021; Mishra et al., 2021; Barra et al., 2020; Phetpan et al., 2018; Feng et al., 2018; Laref et al., 2017;Moreira et al., 2015 Davies, 2007). PLSR is a linear-based calibration technique that uses biased parameter estimation while simultaneously reducing the regression problem's dimensionality (Otto, 2017). The concept behind PLSR is to decompose both matrix X and the y-response by considering each other's information (Puttipipatkajorn et al., 2023; Moghaddam et al., 2022). Thus, the information from the spectral data and the biodiesel properties are simultaneously employed in the calibration phase. Modeling using the PLSR approach can also address collinearity problems while maintaining satisfactory predictive ability (Cunha et al., 2020).

Figure 1 presents the algorithmic workflow for data analysis and model development.

Python 3.11 and relevant scientific computing libraries were used for data processing and model construction. First, exploratory data analysis was conducted using box plots to visualize the distribution and variability of biodiesel properties. Principal component analysis (PCA) was then carried out to evaluate the spectral structure and detect potential outliers. The PCA was implemented using the scikit-learn decomposition module.

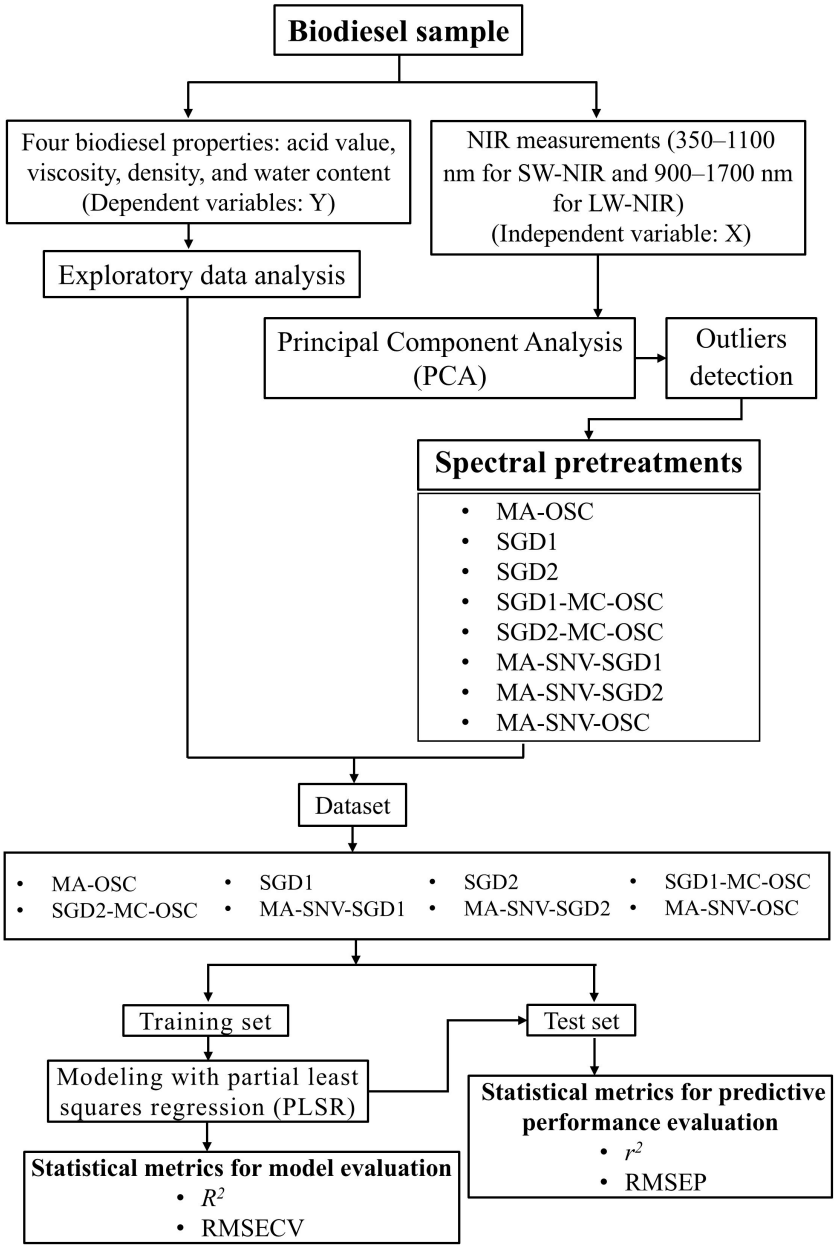

Figure 1 Algorithmic workflow for data analysis and model development

Outlier detection was conducted independently for the SW-NIR and LW-NIR datasets based on the PCA results. Spectra exceeding the 95% confidence limits according to Hotelling's T^2 and Q-residual statistics were considered anomalous and removed before the development of the model. Six spectra originating from two biodiesel samples in the LW-NIR dataset were identified as outliers and excluded due to abnormal spectral behavior. No comparable outliers were detected in the SW-NIR dataset. To avoid bias in model evaluation, outlier removal was performed prior to training–test splitting.

Six spectral pre-processing techniques were applied in this study: moving-average (MA) smoothing, mean centering (MC), orthogonal signal correction (OSC), derivatives (1^{st} derivative: D1 and 2^{nd} derivative: D2), and standard normal variate (SNV). Python was used for

the MA and MC approaches, and the OSC was performed using the orthogonal projections to latent structures (O-PLS) algorithm from the pyopls library. The Savitzky-Golay "savgol_filter" function from the signal processing module in the SciPy library (scipy.signal) was employed for the first and second derivatives, SGD1 and SGD2, respectively. The preprocessing module (sklearn.preprocessing) from the scikit-learn library was used for the SNV-based spectral pretreatment.

MA smoothing was performed using a window size of 5 spectral points. The Savitzky–Golay derivatives were computed with a window length of 11 points and a polynomial order of 2. The number of predictive components in the O-PLS model was set to one, as determined through cross-validation by minimizing the prediction error within the training set.

Based on these preprocessing techniques, eight different applications were utilized for manipulating the spectral data, including MA followed by OSC (MA – OSC), SGD1, SGD2, SGD1 followed by MC and OSC (SGD1 – MC – OSC), SGD2 followed by MC and OSC (SGD2 – MC – OSC), MA followed by SNV and SGD1 (MA – SNV – SGD1), MA followed by SNV and SGD2 (MA – SNV – SGD2), and MA followed by SNV and OSC (MA – SNV – OSC).

The dataset comprised 324 observations (3 spectra per sample) from 108 biodiesel samples, including NIR spectra (X variables) and corresponding ASTM reference values (Y variables) for four biodiesel properties. For each property-specific dataset, four-fifths of the samples were allocated to the training set (88 samples × 3 spectra = 264 observations), whereas one-fifth was assigned to the test set (20 samples × 3 spectra = 60 observations). The dataset was split at the biodiesel sample level using sample IDs to ensure that all replicate spectra from a given sample were assigned exclusively to either the training or test set, thereby preventing information leakage. Both subsets contained paired NIR spectra and their corresponding ASTM reference values; no dataset was constructed solely from spectral or reference measurements.

The training set for each property dataset was used for modeling using the PLSR approach. The "PLSRegression" function in the scikit-learn package's cross-decomposition module was applied to fit the model. The optimal number of latent variables (LVs) was determined via grid search using GridSearchCV combined with grouped 10-fold cross-validation ("GroupKFold"), where sample IDs were used as grouping labels to ensure that replicate spectra from the same biodiesel sample were not split across folds. Model performance during cross-validation was evaluated using the negative mean squared error scoring metric ("neg_mean_squared_error"). Model performance was evaluated using the coefficient of determination for calibration (R^2) and root mean square error of cross-validation (RMSECV). After tuning, the PLSR model was applied to the independent test set (60 observations from 20 samples), and predictive performance was assessed using the coefficient of determination for predictions (r^2) and root mean square error of prediction (RMSEP).

3. Results and Discussion

3.1 Explanatory Data Analysis

Following an initial spectral assessment using the PCA approach, six outlier spectra from two biodiesel samples scanned by the LW-NIR spectrometer were identified and subsequently removed. Table 1 presents the statistical summary of biodiesel properties for both the training and test datasets across the SW-NIR and LW-NIR wavelength ranges, providing key insights into variability and distribution. The mean values of acid value, viscosity, density, and water content are consistent between the training and test sets, indicating the reliability of the dataset. However, slight variations in the standard deviation (SD) and coefficient of variation (%CV) suggest different levels of dispersion. The water content exhibited substantial variability, with a %CV exceeding 26% in some cases, underscoring its sensitivity to measurement and sample heterogeneity. This was followed by the acid value, which exhibited moderate variation (%CV ∼24-25%). The density has the least variation (%CV ∼0.6-0.7%). Overall, the statistical summary indicates that the datasets are well-structured for modeling and prediction, with rea-

sonable consistency between the training and test sets, as illustrated in the box plot in figure 2. Based on the manual splitting of the dataset, the training set exhibits greater variability than the test set for all biodiesel properties. In addition, the four centers, indicated by the medians for each property, overlap. In the case of water content in biodiesel samples, three outliers are observed above the whiskers of the box plots for the training set. However, these outliers were retained during the modeling step.

Table 1 Statistical summary of the biodiesel properties used in the training and test datasets

Wavelength range	Dataset	Statistical terms	Biodiesel properties			
			Acid value (mg KOH/g)	Viscosity (cSt)	Density (kg/m^3)	Water content (mg/kg)
SW-NIR	Training set	N	264	264	264	264
		Max	0.0823	4.457	893.01	848.14
		Min	0.0358	3.964	865.28	166.33
		Mean	0.0559	4.262	878.88	395.81
		SD	0.0140	0.128	6.27	144.16
		% CV	24.9571	3.006	0.71	36.42
	Test set	N	60	60	60	60
		Max	0.0789	4.419	888.85	552.25
		Min	0.0394	4.102	870.88	204.45
		Mean	0.0550	4.264	878.69	375.88
		SD	0.0125	0.109	5.12	97.97
		% CV	22.7230	2.554	0.58	26.06
LW-NIR	Training set	N	258	258	258	258
		Max	0.0823	4.457	893.01	848.14
		Min	0.0358	3.964	865.28	166.33
		Mean	0.0560	4.266	878.97	396.52
		SD	0.0139	0.128	6.23	141.67
		% CV	24.8899	3.008	0.71	35.73
	Test set	N	60	60	60	60
		Max	0.0800	4.426	890.15	634.95
		Min	0.0394	4.102	870.88	204.45
		Mean	0.0560	4.267	879.06	389.53
		SD	0.0128	0.109	5.42	115.30
		% CV	22.8876	2.550	0.62	29.60

3.2 Comparative investigation of regression models

This section considers the effectiveness of PLSR models developed from eight different spectral preprocessing approaches for predicting each biodiesel property. The performance of models built using NIR spectra from two instruments (SW-NIR and LW-NIR) is also compared to assess their relative potential. Tables 2 and 3 present the results of the PLS method to biodiesel NIR spectra obtained from the SW-NIR and LW-NIR spectrometers, respectively. A comparative analysis of the models for each property follows.

3.2.1 Biodiesel acidity

The best result for the acidity prediction model was obtained using the SW-NIR spectral data with MA–SNV–SGD1 pretreatments. This model, employing only three latent variables (LVs), achieved an R^2 of 0.90 and RMSECV of 0.0044 mg KOH/g. When the model was validated

with an independent test set, the prediction performance regarding RMSEP and R^2 were 0.91 and 0.0037 mg KOH/g, respectively. The model using LW-NIR spectra as input variables had slightly higher errors. The best model in this case gave prediction performance with an R^2 of 0.80 and RMSEP of 0.0057 mg KOH/g. Based on the best-performing model for each spectral type, figure 3 shows scatter plots of predicted versus reference acid values and box plots of the corresponding prediction errors. The scatter plots revealed strong linear relationships between the predicted and reference acid values for biodiesel samples across both NIR instruments. Examination of the box plots indicated that the prediction errors for both instruments largely overlapped in the training and test sets, with a median error of approximately 0.000 mg KOH/g. However, in this study, the SW-NIR model exhibited lower variability in its prediction errors, demonstrating superior performance compared to the LW-NIR model for acid value prediction of biodiesel.

3.2.2 Biodiesel viscosity

The best-performing models for viscosity prediction showed similar results for both SW-NIR (Table 2) and LW-NIR (Table 3). For the SW-NIR spectral dataset, the model using spectra pretreated by MA followed by the OSC method yielded the best results, with R^2 and RMSECV values of 0.75 and 0.064 cSt, respectively. Using 5 LVs for the PLS modeling, this model achieved comparable results when predicting biodiesel viscosity in the test dataset, with R^2 and RMSEP values of 0.67 and 0.062 cSt, respectively. In contrast, the model built from the LW-NIR spectra, based on the SGD1–MC–OSC pretreatment, required only 2 LVs. This model produced the same RMSEP of 0.063 cSt when applied to the test dataset. The linear relationships between the predicted and actual biodiesel viscosity values were observed for both NIR instruments (Figure 3). However, the SW-NIR spectra-based model demonstrated lower variability in prediction errors, as shown in the box plots in Figure 3.

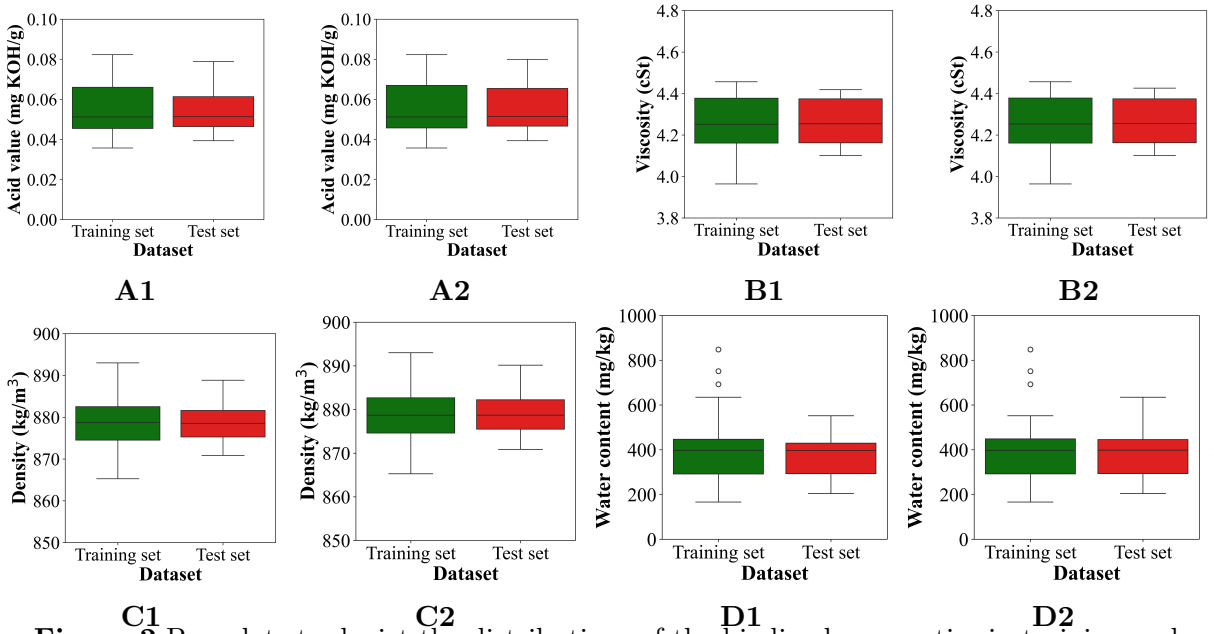

Figure 2 Box plots to depict the distributions of the biodiesel proper ties in training and independent test sets for acid value (A1: SW-NIR and A2: LW-NIR), viscosity (B1: SW-NIR and B2: LW-NIR), density (C1: SW-NIR and C2: LW-NIR), and water content (D1: SW-NIR and D2: LW-NIR)

3.2.3 Biodiesel density

Similar to the previous two biodiesel properties, the SW-NIR spectra-based model outperformed the LW-NIR-based model, although it required more LVs. Using 5 LVs and SGD1-

pretreated spectra, the SW-NIR model achieved an R^2 of 0.89 and RMSECV of 2.07 kg/m³. When applied to the test dataset, it predicted biodiesel density with an R^2 of 0.72 and RMSEP of 2.67 kg/m³. In contrast, the LW-NIR-based model, which was built using MA–OSC-pretreated spectra and fewer LVs, showed lower performance, with an R^2 of 0.76 and RMSECV of 3.05 kg/m^3. As expected, this model had a higher prediction error than the SW-NIR-based model. Linear patterns were observed in the data points for both model types (Figure 3). However, the LW-NIR-based model exhibited higher variability in prediction errors, indicating lower precision in predicting biodiesel density (Figure 3).

Table 2 Result of the PLS model built from the SW-NIR spectral data. The pink-highlighted cells represent the optimal model for each biodiesel property

Biodiesel properties	Pre-processing methods	No. of latent variables	Training set		Prediction set	
			R^2	RMSECV	r^2	RMSEP
Acid value (mg KOH/g)	MA–OSC	3	0.83	0.0058	0.89	0.0041
	SGD1	5	0.92	0.0038	0.86	0.0046
	SGD2	4	0.88	0.0045	0.79	0.0057
	SGD1–MC–OSC	3	0.93	0.0037	0.87	0.0045
	SGD2–MC–OSC	2	0.91	0.0042	0.85	0.0048
	MA–SNV–SGD1	3	0.90	0.0044	0.91	0.0037
	MA–SNV–SGD2	4	0.88	0.0049	0.78	0.0058
	MA–SNV–OSC	4	0.88	0.0048	0.88	0.0043
Viscosity (cSt)	MA–OSC	5	0.75	0.064	0.67	0.062
	SGD1	3	0.65	0.076	0.39	0.084
	SGD2	1	0.41	0.098	0.29	0.091
	SGD1–MC–OSC	2	0.74	0.066	0.47	0.078
	SGD2–MC–OSC	2	0.70	0.070	0.36	0.086
	MA–SNV–SGD1	2	0.65	0.076	0.46	0.079
	MA–SNV–SGD2	1	0.43	0.097	0.31	0.090
	MA–SNV–OSC	3	0.73	0.067	0.31	0.089
Density (kg/m^3)	MA–OSC	5	0.87	2.24	0.67	2.89
	SGD1	5	0.89	2.07	0.72	2.67
	SGD2	1	0.64	3.76	0.42	3.87
	SGD1–MC–OSC	2	0.88	2.16	0.66	2.94
	SGD2–MC–OSC	2	0.85	2.40	0.51	3.57
	MA–SNV–SGD1	4	0.89	2.08	0.70	2.78
	MA–SNV–SGD2	1	0.65	3.72	0.46	3.71
	MA–SNV–OSC	5	0.89	2.06	0.61	3.15
Water content (mg/kg)	MA–OSC	4	0.73	74.10	0.66	56.37
	SGD1	5	0.87	51.88	0.54	66.08
	SGD2	1	0.63	87.78	0.13	90.85
	SGD1–MC–OSC	2	0.85	55.60	0.63	59.14
	SGD2–MC–OSC	2	0.84	56.90	0.23	85.35
	MA–SNV–SGD1	1	0.68	81.93	0.33	79.33
	MA–SNV–SGD2	1	0.64	85.84	0.21	86.30
	MA–SNV–OSC	3	0.76	70.45	0.49	69.16

3.2.4 Water content of biodiesel

Regarding water content prediction in biodiesel, the best model from each spectral type used the same number of LVs. The SW-NIR-based model outperformed the LW-NIR-based model, achieving an R^2 of 0.85 and RMSECV of 55.60 mg/kg. The spectra were preprocessed using SGD1 followed by MC and OSC techniques. With only 2 LVs used for PLS modeling, the SW-NIR-based model predicted the biodiesel water content in the independent test set with a RMSEP of 59.14 mg/kg. In contrast, the LW-NIR-based model demonstrated significantly lower prediction performance, with an RMSEP of 70.34 mg/kg. The higher variability of the

LW-NIR-based model is evident in the prediction error box plot, as shown in Figure 3.

The observed differences among preprocessing combinations in Tables 2 and 3 can be interpreted based on their influence on spectral variance and chemically relevant absorption features. Moving-average (MA) smoothing reduces high-frequency instrumental noise, improving stability in broad C–H overtone regions associated with molecular structure and viscosity-related information. Derivative preprocessing (SGD1 and SGD2) enhances spectral resolution by emphasizing slope and curvature changes, which is particularly beneficial for acid value prediction, where subtle variations in O–H-related absorption bands reflect changes in FFA concentration.

Table 3 Result of the proposed PLS model built from LW-NIR spectral data. The pink-highlighted cells represent the optimal model for each biodiesel property

Biodiesel properties	Pre-processing methods	No. of latent variables	Training set		Prediction set	
			R^2	RMSECV	r^2	RMSEP
Acid value (mg KOH/g)	MA–OSC	3	0.82	0.0060	0.80	0.0057
	SGD1	5	0.91	0.0041	0.53	0.0087
	SGD2	4	0.89	0.0045	0.56	0.0085
	SGD1–MC–OSC	2	0.89	0.0046	0.64	0.0076
	SGD2–MC–OSC	1	0.84	0.0056	0.76	0.0062
	MA–SNV–SGD1	5	0.90	0.0044	0.53	0.0087
	MA–SNV–SGD2	3	0.85	0.0054	0.68	0.0072
	MA–SNV–OSC	4	0.87	0.0051	0.53	0.0087
Viscosity (cSt)	MA–OSC	2	0.60	0.081	0.66	0.063
	SGD1	3	0.70	0.070	0.56	0.072
	SGD2	3	0.74	0.066	0.31	0.090
	SGD1–MC–OSC	2	0.70	0.070	0.66	0.063
	SGD2–MC–OSC	2	0.74	0.066	0.38	0.085
	MA–SNV–SGD1	3	0.66	0.075	0.72	0.057
	MA–SNV–SGD2	3	0.73	0.067	0.52	0.075
	MA–SNV–OSC	3	0.66	0.075	0.73	0.056
Density (kg/m^3)	MA–OSC	3	0.76	3.05	0.71	2.87
	SGD1	3	0.86	2.33	0.50	3.81
	SGD2	2	0.66	3.62	0.59	3.43
	SGD1–MC–OSC	1	0.77	2.98	0.68	3.04
	SGD2–MC–OSC	1	0.72	3.26	0.65	3.19
	MA–SNV–SGD1	3	0.78	2.92	0.65	3.16
	MA–SNV–SGD2	2	0.69	3.45	0.66	3.15
	MA–SNV–OSC	2	0.68	3.52	0.51	3.77
Water content (mg/kg)	MA–OSC	2	0.72	74.28	0.62	70.34
	SGD1	3	0.78	65.83	0.26	98.68
	SGD2	2	0.65	84.19	0.45	84.69
	SGD1–MC–OSC	1	0.75	71.01	0.43	86.36
	SGD2–MC–OSC	1	0.72	75.43	0.45	84.52
	MA–SNV–SGD1	3	0.78	66.24	0.45	84.54
	MA–SNV–SGD2	3	0.78	66.98	0.36	91.239
	MA–SNV–OSC	2	0.67	81.71	0.56	75.69

Standard normal variate (SNV) correction minimizes multiplicative scattering and path-length effects, which is especially important for water prediction due to the strong sensitivity of O–H absorption intensity near 1400 nm. Orthogonal signal correction (OSC) improves model robustness by removing spectral components unrelated to the response variable, thereby enhancing the extraction of chemically meaningful variance during partial least squares (PLS) modeling.

The differences in model performance between the SW and LW systems may be attributed to their spectral coverage. The LW region (900–1700 nm) captures stronger overtone features of O–H and C–H bonds, which can enhance sensitivity for properties such as acid value and water content. In contrast, the SW region includes shorter wavelengths that may be more affected by

scattering and baseline variation, influencing the effectiveness of preprocessing. These findings demonstrate that preprocessing selection is physically and chemically linked to the influence of molecular composition on NIR spectral behavior.

The differences in the optimal number of latent variables (LVs) between the SW and LW models reflect a trade-off between capturing relevant compositional variance and avoiding noise. In some cases, the SW models required more LVs than the LW models. This can occur because SW spectra (350–1100 nm) include shorter wavelengths where baseline variation and scattering effects can be more pronounced, and chemically informative bands may be weaker or more overlapped compared with the LW region (900–1700 nm). Consequently, additional LVs may be required to represent the multivariate structure linking spectra to reference properties, but excessive LVs may also increase the risk of fitting noise. Therefore, the selected LV counts represent the point where cross-validation minimized prediction error while balancing information capture and robustness.

Water content showed the most challenging prediction behavior, which is consistent with its relatively high variability and heterogeneity across samples. As summarized in Table 1, the percentage coefficient of variation (%CV) for water content is the highest among the measured properties (reaching ∼36%), indicating substantial dispersion. In biodiesel, moisture can vary due to washing efficiency, residual emulsified water, and moisture uptake during handling and storage. Such heterogeneity increases within-class spectral variability and reduces the consistency of the spectral–property relationship, thereby increasing prediction uncertainty even when strong O–H absorption features are present.

In contrast, density was generally easier to model, exhibiting low prediction errors, because it is strongly constrained by overall composition and typically varies over a narrower absolute range than other properties. However, the prediction r^2 for density can decrease even when RMSEP remains small. This occurs because r^2 is sensitive to the variance of the reference values: when the test-set density range is narrow, the total variance in the denominator of r^2 is small, so even small absolute prediction errors can reduce r^2 noticeably. Therefore, density models may show a low RMSEP but a comparatively lower r^2 under narrow-range external testing.

3.3 Comparison with available literature data

In 2007, Felizardo et al., 2007 applied a near-infrared (NIR) spectroscopic technique combined with a multivariate calibration approach to predict the methanol and water content of biodiesel. An ABB Bomem MB-160 FT-NIR spectrometer with a transflection probe was used to collect the NIR spectra. Biodiesel samples were produced from various feedstocks, such as soybean, soybean-palm mixtures, and waste frying oils. Their study explored different spectral preprocessing methods and calibration techniques, such as principal component regression (PCR) and partial least square regression (PLSR).

In 2008, Baptista et al., 2008 used NIR spectroscopy to determine four biodiesel properties: iodine value, cold filter plugging point, kinematic viscosity, and density. NIR spectra were collected using biodiesel samples from various feedstocks with an ABB Bomem MB-160 FT-NIR spectrometer equipped with an FTLA ACC131 liquid accessory. This study applied PLSR-based calibration combined with multiple spectral pre-processing techniques.

Two notable studies on the prediction of biodiesel properties using NIR spectroscopy were published in 2011. Balabin and Smirnov, 2011 used an MPA Multi-Purpose FT-NIR Analyzer to collect spectra from biodiesel samples. Industrial-based and used frying oil-based biodiesel were used as samples, and four properties were considered: density, viscosity, water content, and methanol content. This study compared the performance of linear and nonlinear calibration approaches and investigated the influence of different preprocessing methods. Balabin and Smirnov [5] expanded their work in a follow-up study by focusing on full-wavelength-based calibration and incorporating variable (spectral wavelength) selection techniques combined with multiple linear regression (MLR) and PLSR-based calibrations.

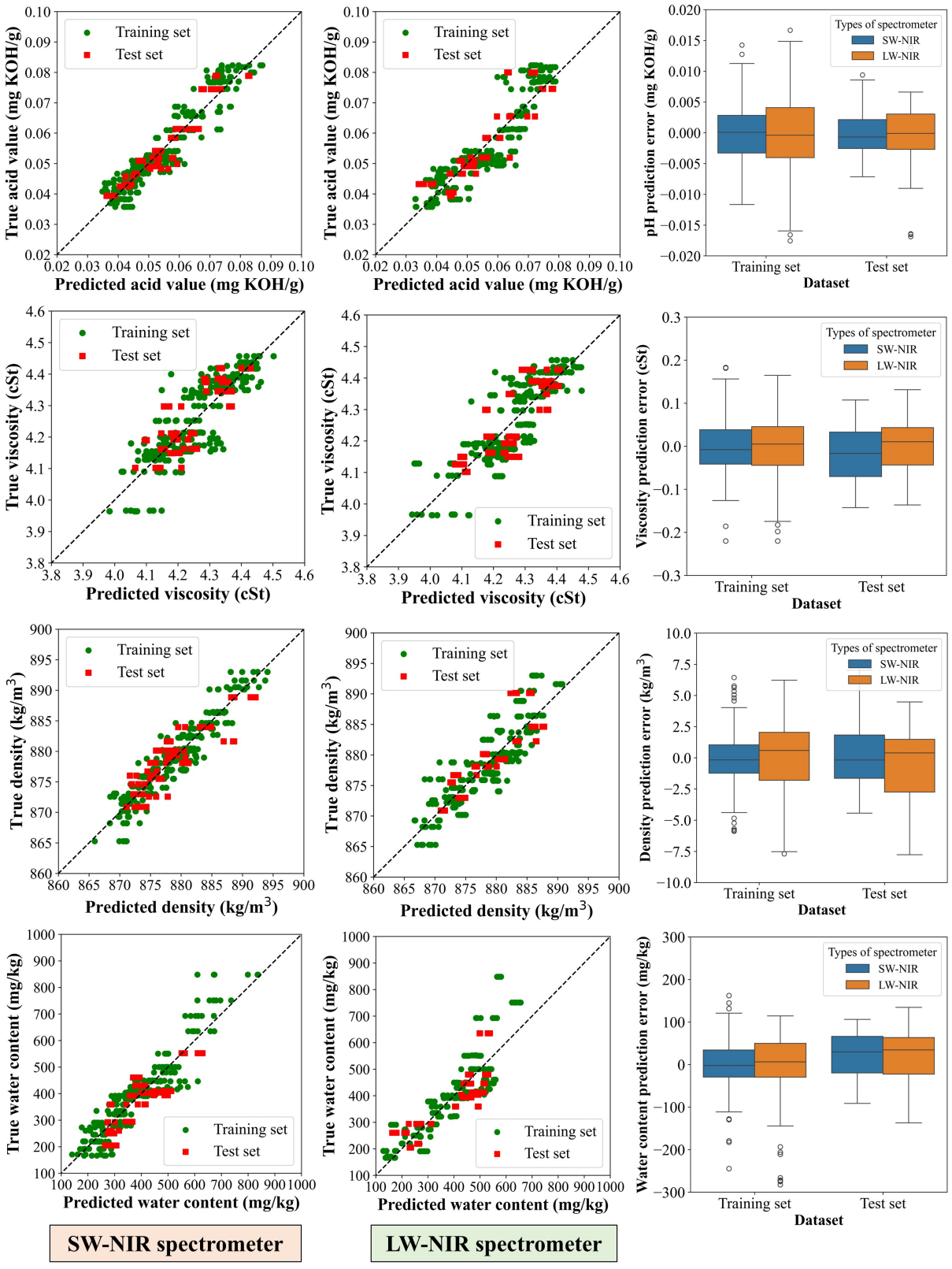

Figure 3 Scatter plots of predicted and true biodiesel properties and box plots of the prediction error for comparing the performance of the SW-NIR and LW-NIR spectrometers

A recent study by Cunha et al., 2020 in 2020 explored NIR spectroscopic approach for the prediction of biodiesel properties. This study employed multivariate calibrations based on PLS, RF, and SVM methods, along with several variable selection techniques, to relate the NIR spectral data of biodiesel fuel to its physical-chemical properties. Biodiesel samples synthesized

from different refined oils, such as soybean, canola, corn, and sunflower, were analyzed using a transflection probe and Perkin-Elmer FT-NIR spectrometer.

A comparison of the current study with earlier works is provided in Table 4. Notably, none of the previous studies have reported NIR-based acidity prediction in biodiesel. Except for density, the calibration models proposed in our study were superior to the models reported in Baptista et al. [3], Balabin et al., 2011 and Balabin and Smirnov, 2011, particularly for viscosity prediction.s. However, our model developed based on full wavelength data yielded worse results than those of Cunha et al., 2020, whose best model was developed using a variable selection tool called uninformative variable elimination (UVE). Regarding water content prediction in biodiesel, the model developed from the SW-NIR spectral data achieved the best accuracy with the lowest RMSEP value (59 mg/kg). The LW-NIR-based model yielded results comparable to those of previous studiesd. However, our models exhibited higher prediction errors for density prediction compared with those of earlier works.

Our results suggest that diode array NIR instruments have significant potential for predicting biodiesel properties. Additionally, they exhibit comparable efficiency to FT-NIR spectrometers, which are interferometer-based and were used in all the aforementioned studies. Notably, our findings highlight the advantage of diode array NIR spectrometers, which with a silicon detector operating in the 380–1100 nm (SW-NIR) spectral range, offer a lower-cost alternative with promising potential for real-time biodiesel property measurement.

Note that NIR spectroscopy is sensitive to molecular vibrational overtones primarily associated with organic functional groups (e.g., O–H and C–H bonds). Trace metal content typically does not exhibit strong NIR absorption, and pre-measurement filtration minimized particulates.Because all samples were derived from the same process and measured under consistent optical conditions, the observed spectral variations are predominantly attributed to differences in physicochemical properties rather than external cleanliness or metallic interferences. Nevertheless, future studies involving more heterogeneous industrial samples would be valuable to further assess model robustness under varying contamination scenarios.

Table 4 Comparison of PLS results with available literature data: model prediction errors (RMSEP)

	Spectral region	Biodiesel properties prediction errors (RMSEP)			
		Acid value (mg KOH/g)	Viscosity (cSt)	Density (kg/m^3)	Water content (mg/kg)
Felizardo et al., 2007	1110–2220 nm (9000–4500 cm^{-1})	–	–	–	75[a]
Baptista et al., 2008	1110–2220 nm (9000–4500 cm^{-1})	–	0.09[a]	0.90[a]	–
Balabin et al., 2011	1110–2220 nm (9000–4500 cm^{-1})	–	0.09[a]	0.83[a]	69[a]
Balabin and Smirnov, 2011	1110–2220 nm (9000–4500 cm^{-1})	–	0.14[b]	0.64[b]	70[c]
Cunha et al., 2020	1000–2500 nm (10000–4000 cm^{-1})	–	0.01[d]	–	–
Chungcharoen et al.	350–1100 nm	0.0037[a]	0.06[a]	2.67[a]	59[a]
(Current study)	900–1700 nm	0.0057[a]	0.06[a]	2.87[a]	70[a]

Note: [a]Full-PLSR, [b]Kohonen artificial neural network (K-ANN) based PLSR, [c]Genetic algorithm (GA)-PLSR, [d]uninformative variable elimination (UVE)-PLSR.

4. Conclusions

This study compared the performance of two diode array NIR spectrometers (SW-NIR versus LW-NIR) for predicting biodiesel properties. The SW-NIR spectrometer outperformed

the LW-NIR spectrometer in predicting all biodiesel properties evaluated. The optimal spectral preprocessing methods varied by property: MA–SNV–SGD1 was the most effective for acidity prediction, while MA–OSC produced the best viscosity model. For density and water content prediction, the best results were achieved using SGD1 and SGD1–MC–OSC. Further research is required to assess the efficiency of the model in predicting additional biodiesel properties and handling more complex samples from diverse feedstocks, as this study focused exclusively on oil palm-based biodiesel.

Acknowledgements

King Mongkut's Institute of Technology Ladkrabang provided funding support for this research [2567-02-08-003].

Author Contributions

Kittisak Phetpan: Conceptualization, methodology, writing, original draft, review and editing, formal analysis, visualization, validation, funding acquisition, supervision. Chitwadee Thongphut: Conceptualization, methodology, data curation, formal analysis, writing, review, and editing, visualization, and investigation. Thatchapol Chungcharoen: Conceptualization, methodology, writing, review, and editing, funding acquisition, visualization, and investigation. Warunee Limmun: Conceptualization, Methodology, Review and Editing, Validation.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

Abreu, F., Lima, D., Hamú, E., Wolf, C., & Suarez, P. (2004). Utilization of metal complexes as catalysts in the transesterification of brazilian vegetable oils with different alcohols. *Journal of Molecular Catalysis A: Chemical*, *209*, 29–33. https://doi.org/10.1016/j.molcata.2003.08.003

Balabin, R., Lomakina, E., & Safieva, R. (2011). Neural network (ann) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (nir) spectroscopy. *Fuel*, *90*, 2007–2015. https://doi.org/10.1016/j.fuel.2010.11.038

Balabin, R., & Safieva, R. (2007). Capabilities of near infrared spectroscopy for the determination of petroleum macromolecule content in aromatic solutions. *Journal of Near Infrared Spectroscopy*, *15*, 343–349. https://doi.org/10.1255/jnirs.749

Balabin, R., & Smirnov, S. (2011). Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. *Analytica Chimica Acta*, *692*, 63–72. https://doi.org/10.1016/j.aca.2011.03.006

Baptista, P., Felizardo, P., Menezes, J., & Correia, M. (2008). Multivariate near infrared spectroscopy models for predicting the iodine value, cfpp, kinematic viscosity at 40 °c and density at 15 °c of biodiesel. *Talanta*, *77*, 144–151. https://doi.org/10.1016/j.talanta.2008.06.001

Barra, I., Kharbach, M., Qannari, E., Hanafi, M., Cherrah, Y., & Bouklouze, A. (2020). Predicting cetane number in diesel fuels using ftir spectroscopy and pls regression. *Vibrational Spectroscopy*, *111*, 103157. https://doi.org/10.1016/j.vibspec.2020.103157

Berchmans, H., & Hirata, S. (2008). Biodiesel production from crude jatropha curcas l. seed oil with a high content of free fatty acids. *Bioresource Technology*, *99*, 1716–1721. https://doi.org/10.1016/j.biortech.2007.03.051

Bournay, L., Casanave, D., Delfort, B., Hillion, G., & Chodorge, J. (2005). New heterogeneous process for biodiesel production: A way to improve the quality and the value of the crude glycerin produced by biodiesel plants. *Catalysis Today*, *106*, 190–192. https://doi.org/10.1016/j.cattod.2005.07.181

Bukkarapu, K. R., & Krishnasamy, A. (2022). Predicting engine fuel properties of biodiesel and biodiesel-diesel blends using spectroscopy based approach. *Fuel Processing Technology*, *230*, 107227. https://doi.org/10.1016/j.fuproc.2022.107227

Cunha, C., Luna, A., Oliveira, R., Xavier, G., Paredes, M., & Torres, A. (2017). Predicting the properties of biodiesel and its blends using mid-ft-ir spectroscopy and first-order multivariate calibration. *Fuel*, *204*, 185–194. https://doi.org/10.1016/j.fuel.2017.05.057

Cunha, C., Torres, A., & Luna, A. (2020). Multivariate regression models obtained from near-infrared spectroscopy data for prediction of the physical properties of biodiesel and its blends. *Fuel*, *261*, 116344. https://doi.org/10.1016/j.fuel.2019.116344

Davies, A. (2007). Back to basics: Spectral pre-treatments, derivatives. *Tony Davies Column*, *19*, 32–33.

Delfino, J., Pereira, T., Viegas, H., Marques, E., Ferreira, A., Zhang, L., Zhang, J., & Marques, A. (2018). A simple and fast method to determine water content in biodiesel by electrochemical impedance spectroscopy. *Talanta*, *179*, 753–759. https://doi.org/10.1016/j.talanta.2017.11.053

Fearn, T. (2000). On orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems*, *50*, 47–52. https://doi.org/10.1016/S0169-7439(99)00045-3

Felizardo, P., Baptista, P., Menezes, J., & Correia, M. (2007). Multivariate near infrared spectroscopy models for predicting methanol and water content in biodiesel. *Analytica Chimica Acta*, *595*, 107–113. https://doi.org/10.1016/j.aca.2007.02.050

Feng, X., Yu, C., Shu, Z., Liu, X., Yan, W., Zheng, Q., Sheng, K., & He, Y. (2018). Rapid and non-destructive measurement of biofuel pellet quality indices based on two-dimensional near infrared spectroscopic imaging. *Fuel*, *228*, 197–205. https://doi.org/10.1016/j.fuel.2018.04.149

Hoang, A. (2021). Prediction of the density and viscosity of biodiesel and the influence of biodiesel properties on a diesel engine fuel supply system. *Journal of Marine Engineering and Technology*, *20*, 299–311. https://doi.org/10.1080/20464177.2018.1532734

Hradecká, I., Vráblík, A., Fratczak, J., Sharkov, N., Černý, R., & Hönig, V. (2023). Near-infrared spectroscopy as a tool for simultaneous determination of diesel fuel improvers. *ACS Omega*, *8*, 4038–4045. https://doi.org/10.1021/acsomega.2c06845

Knothe, G. (2005). Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. *Fuel Processing Technology*, *86*(10), 1059–1070. https://doi.org/10.1016/j.fuproc.2004.11.002

Kumbhar, V., Pandey, A., Sonawane, R., El-Shafay, S., Panchal, H., & Chamkha, A. (2022). Statistical analysis on prediction of biodiesel properties from its fatty acid composition. *Case Studies in Thermal Engineering*, *30*, 101775. https://doi.org/10.1016/j.csite.2022.101775

Kuriyama, K., Kaba, Y., Saitoh, H., Bannu, M., Manago, N., Harayama, Y., Osa, K., Yamamoto, M., & Kuze, H. (2011). Visible and near-infrared differential optical absorption spectroscopy (doas) for the measurement of nitrogen dioxide, carbon dioxide and water vapor. *International Journal of Technology*, *2*(2), 94–101. https://doi.org/10.14716/ijtech.v2i2.54

Laref, R., Ahmadou, D., Losson, E., & Siadat, M. (2017). Orthogonal signal correction to improve stability regression model in gas sensor systems. *Journal of Sensors*, 1–8. https://doi.org/10.1155/2017/9851406

Liu, X., Piao, X., Wang, Y., Zhu, S., & He, H. (2008). Calcium methoxide as a solid base catalyst for the transesterification of soybean oil to biodiesel with methanol. *Fuel*, *87*, 1076–1082. https://doi.org/10.1016/j.fuel.2007.05.059

Meher, L., Sagar, D., & Naik, S. (2006). Technical aspects of biodiesel production by transesterification—a review. *Renewable and Sustainable Energy Reviews*, *10*, 248–268. https://doi.org/10.1016/j.rser.2004.09.002

Mishra, P., Gupta, A., Kumar, A., & Bose, A. (2020). Methanol and petrol blended alternate fuel for future sustainable engine: A performance and emission analysis. *Measurement*, *155*, 107519. https://doi.org/10.1016/j.measurement.2020.107519

Mishra, P., Marini, F., Biancolillo, A., & Roger, J. (2021). Improved prediction of fuel properties with near-infrared spectroscopy using a complementary sequential fusion of scatter correction techniques. *Talanta*, *223*, 121693. https://doi.org/10.1016/j.talanta.2020.121693

Moghaddam, H., Tamiji, Z., Lakeh, M., Khoshayand, M., & Mahmoodi, M. (2022). Multivariate analysis of food fraud: A review of nir based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis*, *107*, 104343. https://doi.org/10.1016/j.jfca.2021.104343

Monteiro, M., Ambrozin, A., Santos, M., Boffo, E., Pereira-Filho, E., Lião, L., & Ferreira, A. (2009). Evaluation of biodiesel–diesel blends quality using 1h nmr and chemometrics. *Talanta*, *78*, 660–664. https://doi.org/10.1016/j.talanta.2008.12.026

Moreira, S., Sarraguça, J., Saraiva, D., Carvalho, R., & Lopes, J. (2015). Optimization of nir spectroscopy based plsr models for critical properties of vegetable oils used in biodiesel production. *Fuel*, *150*, 697–704. https://doi.org/10.1016/j.fuel.2015.02.082

Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user-friendly guide to multivariate calibration and classification*. NIR Publications.

Naghipour, M., Ling, L. S., & Connie, T. (2024). A review of ai techniques in fruit detection and classification: Analyzing data, features and ai models used in agricultural industry. *International Journal of Technology*, *15*(3), 585–596. https://doi.org/10.14716/ijtech.v15i3.6404

Oliveira, F., Brandao, C., Ramalho, H., Costa, L., Suarez, P., & Rubim, J. (2007). Adulteration of diesel/biodiesel blends by vegetable oil as determined by fourier transform (ft) near infrared spectrometry and ft-raman spectroscopy. *Analytica Chimica Acta*, *587*, 194–199. https://doi.org/10.1016/j.aca.2007.01.045

Otto, M. (2017). *Chemometrics: Statistics and computer application in analytical chemistry* (3rd ed.). Wiley-VCH.

Phetpan, K., Udompetaikul, V., & Sirisomboon, P. (2018). An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor. *Computers and Electronics in Agriculture*, *154*, 406–466. https://doi.org/10.1016/j.compag.2018.09.033

Puttipipatkajorn, A., Terdwongworakul, A., Puttipipatkajorn, A., Kulmutiwat, S., Sangwanangkul, P., & Cheepsomsong, T. (2023). Indirect prediction of dry matter in durian pulp with combined features using miniature nir spectrophotometer. *IEEE Access*, *11*, 84810–84821. https://doi.org/10.1109/ACCESS.2023.3303020

Rashid, U., & Anwar, F. (2008). Production of biodiesel through optimized alkaline-catalyzed transesterification of rapeseed oil. *Fuel*, *87*, 265–273. https://doi.org/10.1016/j.fuel.2007.05.003

Ruttanadech, N., Phetpan, K., Srisang, N., Srisang, S., Chungcharoen, T., Limmun, W., Youryon, P., & Kongtragoul, P. (2023). Rapid and accurate classification of aspergillus ochraceous contamination in robusta green coffee bean through near-infrared spectral analysis using machine learning. *Food Control*, *145*, 109446. https://doi.org/10.1016/j.foodcont.2022.109446

Sajjadi, B., Asaithambi, P., Raman, A., & Ibrahim, S. (2017). Hybrid neuro-fuzzy methods for estimation of ultrasound and mechanically stirring influences on biodiesel synthesis through transesterification. *Measurement*, *103*, 62–76. https://doi.org/10.1016/j.measurement.2017.01.044

Sarin, R., Sharma, M., Sinharay, S., & Malhotra, R. (2007). Jatropha–palm biodiesel blends: An optimum mix for asia. *Fuel*, *86*, 1365–1371. https://doi.org/10.1016/j.fuel.2006.11.040

Udompetaikul, V., Phetpan, K., & Sirisomboon, P. (2021). Development of the partial least-squares model to determine the soluble solids content of sugarcane billets on an elevator conveyor. *Measurement*, *167*, 107898. https://doi.org/10.1016/j.measurement.2020.107898

Wahyono, Y., Hadiyanto, Budihardjo, M. A., Hariyono, Y., & Baihaqi, R. A. (2022). Multi-feedstock biodiesel production from a blend of five oils through transesterification with variation of moles ratio of oil: Methanol. *International Journal of Technology*, *13*(3), 606–618. https://doi.org/10.14716/ijtech.v13i3.4804

Walsh, K. B., Blasco, J., Zude-Sasse, M., & Sun, X. (2020). Visible-nir point spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. *Postharvest Biology and Technology*, *168*, 111246. https://doi.org/10.1016/j.postharvbio.2020.111246

Williams, P. (2019). *Near-infrared technology: Getting the best out of light*. African Sun Media.

Yang, H., Griffiths, P., & Tate, J. (2003). Comparison of partial least squares regression and multi-layer neural networks for quantification of nonlinear systems and application to gas phase fourier transform infrared spectra. *Analytica Chimica Acta*, *498*, 125–136. https://doi.org/10.1016/S0003-2670(03)00726-8