



OPEN ACCESS

EDITED BY

Lei Shu,
Nanjing Agricultural University, China

REVIEWED BY

Salim Heddami,
University of Skikda, Algeria
Agustami Sitorus,
National Research and Innovation Agency
(BRIN), Indonesia
Alireza Sanaeifar,
Zhejiang University, China
Xihui Bian,
Tiangong University, China

*CORRESPONDENCE

Zhiyong Pei
✉ ndtgzy@163.com

SPECIALTY SECTION

This article was submitted to
Technical Advances in Plant Science
a section of the journal
Frontiers in Plant Science

RECEIVED 11 December 2022

ACCEPTED 20 February 2023

PUBLISHED 10 March 2023

CITATION

Li Y, Via BK, Han F, Li Y and Pei Z (2023)
Comparison of various chemometric
methods on visible and near-infrared
spectral analysis for wood density
prediction among different tree species
and geographical origins.
Front. Plant Sci. 14:1121287.
doi: 10.3389/fpls.2023.1121287

COPYRIGHT

© 2023 Li, Via, Han, Li and Pei. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Comparison of various chemometric methods on visible and near-infrared spectral analysis for wood density prediction among different tree species and geographical origins

Ying Li¹, Brian K. Via², Feifei Han³, Yaoxiang Li⁴
and Zhiyong Pei^{1*}

¹College of Energy and Transportation Engineering, Inner Mongolia Agricultural University, Hohhot, China, ²Forest Products Development Center, School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL, United States, ³Laboratory Zhejiang Huadong Forestry Engineering Consulting and Design Corporation, Hangzhou, China, ⁴College of Engineering and Technology, Northeast Forestry University, Harbin, China

Visible and near-infrared (Vis-NIR) spectroscopy has been widely applied in many fields for the qualitative and quantitative analysis. Chemometric techniques including pre-processing, variable selection, and multivariate calibration models play an important role to better extract useful information from spectral data. In this study, a new de-noising method (lifting wavelet transform, LWT), four variable selection methods, as well as two non-linear machine learning models were simultaneously analyzed to compare the impact of chemometric approaches on wood density determination among various tree species and geographical locations. In addition, fruit fly optimization algorithm (FOA) and response surface methodology (RSM) were employed to optimize the parameters of generalized regression neural network (GRNN) and particle swarm optimization-support vector machine (PSO-SVM), respectively. As for various chemometric methods, the optimal chemometric method was different for the same tree species collected from different locations. FOA-GRNN model combined with LWT and CARS deliver the best performance for Chinese white poplar of Heilongjiang province. In contrast, PLS model showed a good performance for Chinese white poplar collected from Jilin province based on raw spectra. However, for other tree species, RSM-PSO-SVM models can improve the performance of wood density prediction compared to traditional linear and FOA-GRNN models. Especially for *Acer mono Maxim*, when compared to linear models, the coefficient of determination of prediction set (R_p^2) and relative prediction deviation (RPD) were increased by 47.70% and 44.48%, respectively. And the dimensionality of Vis-NIR spectral data was decreased from 2048 to 20. Therefore, the appropriate chemometric technique should be selected before building calibration models.

KEYWORDS

visible and near infrared spectroscopy, lifting wavelet transform, variable selection, response surface methodology, wood density

1 Introduction

Visible and near-infrared (Vis-NIR) spectroscopy, which contains the visible and NIR regions, has been widely applied in agriculture, petroleum, pharmaceuticals, and life sciences, such as soil particle size determination (Gozukara et al., 2022), propane content prediction of liquefied petroleum gas (Dantas et al., 2013), polymorphic forms of fluconazole identification (Mansouri et al., 2021), plant stress detection (Liang et al., 2018), and examination of Zika virus (Fernandes et al., 2018). The visible region (380–750 nm) contains the information of the pigments such as anthocyanin and chlorophyll based on their specific absorption bands (Zahir et al., 2022). Meanwhile, the near-infrared light records the vibration of the hydrogen bonds, for instance, C-H, N-H, and O-H, which are the main components of samples. Therefore, when the Vis-NIR light strikes samples, the light absorbed by samples includes the information of pigments and hydrogen bonds, which can be used to predict samples' components.

Wood density is an essential indicator for the assessment of wood qualities due to the relationship between the mechanical, optical, and chemical properties (Resquin et al., 2019). The traditional measurement of wood density is laboratory test based on the density formula ($\rho = m/v$, where ρ is wood density, m and v are the mass and volume of wood samples, respectively), such as China National Standards (GB/T 1933-2009), which is challenging because they require burdensome sample processings, meaning that it is a destructive behavior and difficult to measure numerous samples in a short time. In addition, wood properties are influenced by the tree species and geographical origins (climate, moisture, soil, etc.). Even within the same tree, there exist differences between juvenile and mature wood for density (Krajnc et al., 2021). These differences in wood properties have effect on end use and economic benefits. For example, the yellow rosewood (the raw material of classical Chinese furniture) grown in Hainan province is expensive than other locations due to its great quality (Huang et al., 2018). Therefore, it is necessary to analyze wood properties among different geographical locations and tree species, especially for native tree species and the Convention on International Trade in Endangered Species (CITES) listed species.

Many studies (Zhao et al., 2010; Zhao et al., 2012; Tigabu et al., 2020; Toscano et al., 2022) have demonstrated that Vis-NIR or NIR spectroscopy can be used to determine physical and chemical compositions, mechanical properties, wood microstructure, and seed quality over the years with the advantages of rapid, simple, and non-destructive detection for numerous samples. For the prediction of wood density using spectroscopy, various wood samples and chemometric techniques or the combination of these two sections are the main research directions. In terms of wood science, Schimleck et al. (2003) estimated air-dry density of green *Pinus taeda* radial samples with NIR spectroscopy, the coefficient of determination (R^2) are 0.85 and 0.87 for green and dry wood, respectively. Additionally, in another study, Schimleck et al. (2018) found that air-dry density of *Pinus taeda* L increased from pith to bark at all heights based on NIR spectroscopy technology. As for chemometric techniques, Zhang et al. (2022) proposed a deep transfer learning hybrid method with automatic calibration

capability (Resnet1D-SVR-TrAdaBoost.R2) to predict larch wood density in different moisture contents. Fernandes et al. (2013) compared the effect of two calibration methods (neural networks and partial least squares) on *Pinus pinea* density, the results demonstrated that neural networks was better than PLS technique. In addition, considering spectra quality and model accuracy, Li et al. (2022) analyzed various spectral pre-processing and multivariate calibration methods in the prediction of Chinese White Poplar density, the results showed that the best prediction was obtained by GRNN models combined with LWT and CARS method. These studies displayed that chemometric techniques are essential for NIR spectra analysis to better explore the relationship between spectra and properties.

The original spectra contain irrelevant information due to the interference of background and environment, therefore, chemometric methods are needed for Vis-NIR spectral analysis. The essence of Vis-NIR spectral data analysis is to extract useful information of components using the appropriate chemometric methods, which include pre-processing, feature variable selection, and multivariate calibration models. Pre-processing is an essential step for improving model prediction accuracy through converting raw spectral data into a new data set without interferences (Bian et al., 2020). The common used pre-processing techniques are multiplicative scatter correction (MSC), the first derivative, Savitzky-Golay (SG) filtering, detrending (DT), wavelet transform (WT) and standard normal variate (SNV) (Dotto et al., 2018; Li et al., 2020; Bian et al., 2022; Carvalho et al., 2022; Ling et al., 2022). Different results will be obtained using various pre-processing methods or their combinations due to the different mechanism and functions, thus, it is important to select the most useful method and prevent the phenomena of over-fitting.

A Vis-NIR spectrum of a sample with the region from 350 to 2500 nm includes 2151 spectral variables, the high-dimensional spectral data results in the "curse of dimensionality". Therefore, feature wavelengths selection techniques should be employed to address the problem. Uninformative variable elimination (UVE), competitive adaptive reweighted sampling (CARS), genetic algorithm (GA), iteratively retains informative variables (IRIV), and successive projections algorithm (SPA) are among the feature variable selection methods that have been analyzed in recent studies (Centner et al., 1996; Araújo et al., 2001; Ielpo et al., 2017). In addition to these mentioned techniques, a hybrid method that combing two or more methods were made to simplify high-dimensional spectral data. For example, Yun et al. proposed a hybrid strategy based on variable combination population analysis (VCPA), IRIV, and GA to optimize key variables (Yun et al., 2015; Yun et al., 2019).

Multivariate calibration models are employed to analyze the relationship between the selected feature wavelengths and targeted properties. Generally, modeling approaches are divided into linear and non-linear models such as partial least-squares (PLS), principal component regression (PCR), artificial neural network (ANN), and support vector machine (SVM). Regarding the performance of linear and non-linear models, there may exist differences in prediction accuracy for the same sample because of the different strategies. According to the Beer-Lambert's Law ($A = \epsilon \times l \times c$, where

A is absorbance, ϵ , l , and c are the molar absorption coefficient, path length, and concentration, respectively), there is a linear relationship between absorbance spectra and concentrations (Wang et al., 2011). However, the relationship between spectra and properties are complex or non-linearity due to the interference of environment, such as light scattering (Geladi et al., 1985). As for NIR spectra analysis, a linear method of PLS is usually used to model. If the residuals of the predicted model are normally distributed around zero then the PLS model is accurate. If the residuals will not be equally distributed around zero but will follow perhaps a "banana shaped" curve around zero or some unbalanced pattern, another calibration method or non-linear model will be employed to analyze the relationship between spectral data and properties. It is worth mentioning that non-linear calibration models have ability to perform linear analysis. What's more, non-linear relationships between spectra and concentrations can be handled by the linear calibration models, but at the cost of the increasing of the multivariate complexity. However, the linear models are not always effective in the spectral data analysis when the noises are multiplicative. Additionally, spectral differences as different unknown samples or growing environment are complex. In this case, it is not enough to deliver the optimal solution only using linear techniques for such complex problem.

Recently, some new algorithms, such as random forest (RF) (Marta et al., 2022) and fruit fly optimization algorithm (FOA) (Li et al., 2019), were applied in the analysis of spectral data. Modeling methods are the same as pre-processing and feature variables selection techniques, they can be combined by two or three approaches to obtain the most suitable model. The limitations of modeling methods (linear and non-linear models) are the optimization of modeling parameters. For instance, the principal component number (PCs) and the selection of radial basis function (RBF) are the key step of PLS and SVM models, respectively.

In this study, various chemometric methods including a new spectral data de-noising (Lifting wavelet transform, LWT), four feature variables selection techniques (SPA, UVE, IRIV, and CARS), and two hybrid multivariate calibration models, i.e.,

generalized regression neural network (GRNN) optimized by FOA (GRNN-FOA), and PSO-SVM model optimized by response surface methodology (RSM-PSO-SVM), were compared simultaneously to obtain the most suitable chemometric technique for wood density prediction among different tree species. In addition to, the effect of geographical location on wood Vis-NIR spectra was investigated to predict wood density based on Vis-NIR spectroscopy.

2 Materials and methods

2.1 Study area and wood sampling

A total of 37 trees were collected from two physiographic regions of China (Figure 1). The main tree species comprising *Populus davidiana*, *Ulmus pumila* L., *Acer mono Maxim.*, and *Tilia tuan Szyszyl.* were obtained from Jilin province. *Populus davidiana* and *Larixgmelinii* were simultaneously made from Heilongjiang province to analyze the influence of locations on wood density determination. The study area have east Asian monsoon climate and temperate continental monsoon climate, respectively. Wood disks with five-centimeter-thick were made from the stump up to the top at 1 m intervals. In total, 530 wood samples with the dimensions of $2 \times 2 \times 2 \text{ cm}^3$ (tangential, radial, and longitudinal) were generated and then air-dried in laboratory for three months.

2.2 Vis-NIR spectra collection and wood density measurement

The reflectance spectra of wood were measured from the cross-section using a portable spectrometer (LabSpec, Analytical Spectral Devices, Inc., Boulder, USA). The wavelength range and spectral resolution are 350-2500 nm and 3 nm @700 nm, 10 nm @1400/2100 nm, respectively. A white panel was used for instrumental calibration every fifteen min. Three random spectra were collected from each sample and the average spectrum was regarded as the raw

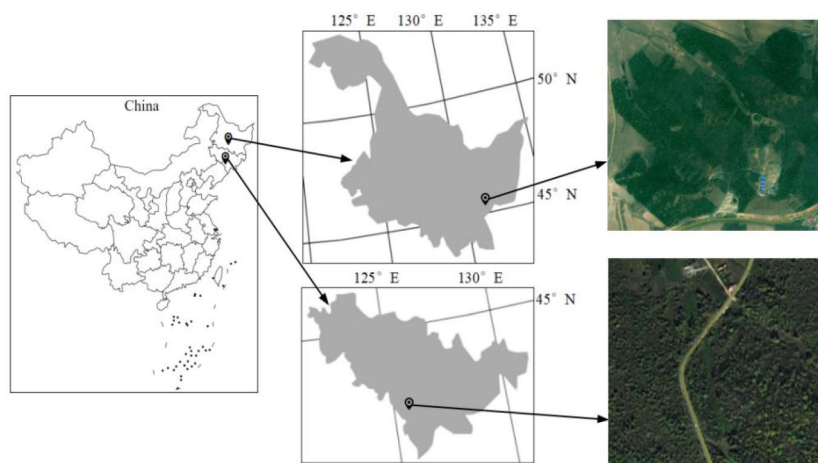


FIGURE 1
Geographical location of wood samples.

spectrum. Wood density was measured according to China National Standards (GB/T 1933-2009).

2.3 Chemometric techniques

2.3.1 Spectral pre-processing

Wavelet transform (WT) is a powerful signal analysis technique in data compression and spectral de-noising (Mojsilovic et al., 2000). Lifting wavelet transform (LWT), the second-generation wavelet, can conquer the weakness of traditional WT. The inconspicuous information are magnified by LWT with the advantages of high computation speed and small memory. Wavelet function, wavelet order, and decomposition level (k) are essential for spectral de-noising using LWT and WT. Wood spectra with the bands ranging from 350 to 2397 nm (No. of wavelengths = 2048) were used to analyze in this study. Assuming that decomposition level and wavelet order are 8 and 3, respectively, four wavelet function including Haar, sym3, bior1.3 (biorNrNd, Nr: the wavelet order in decomposition process, Nd: the wavelet order of reconstruction wavelet), and db3 were compared in this study. The optimal wavelet function was determined by the performance of partial least squares (PLS) models. Then the optimal wavelet order and decomposition level were obtained based on the best wavelet function.

In order to better analyze the suitability of LWT, wood spectra were processed by three traditional pre-processing methods including WT, SNV, and MSC. The main parameters of WT were the same as LWT. LWT and WT were implemented in Matlab R2010b (MathWorks, Natick, USA). MSC, SNV, and PLS were performed with Unscrambler V10.4 (CAMO Software AS, Oslo, Norway).

2.3.2 Feature variables selection

After the optimal de-noising technique was determined, four regular variable selection algorithms (i.e., UVE, CARS, IRIV, and

SPA) were used to select the feature wavelengths of wood density. The selection strategy of UVE and CARS are filter-based and MPA-based, respectively. SPA is based on extreme value search and forward selection. In contrast with SPA, backward selection and MPA-based are employed by IRIV (Yun et al., 2014). These four variables selection methods were implemented with Matlab R2010b. The performance of these methods was analyzed by the PLS models according to the values of the coefficient of determination (R^2), root mean square error (RMSE), relative prediction deviation (RPD), and relative standard deviation (RSD). Generally, a higher R^2 , RPD and lower RMSE, RSD value indicates a good predictive ability (Yan et al., 2013).

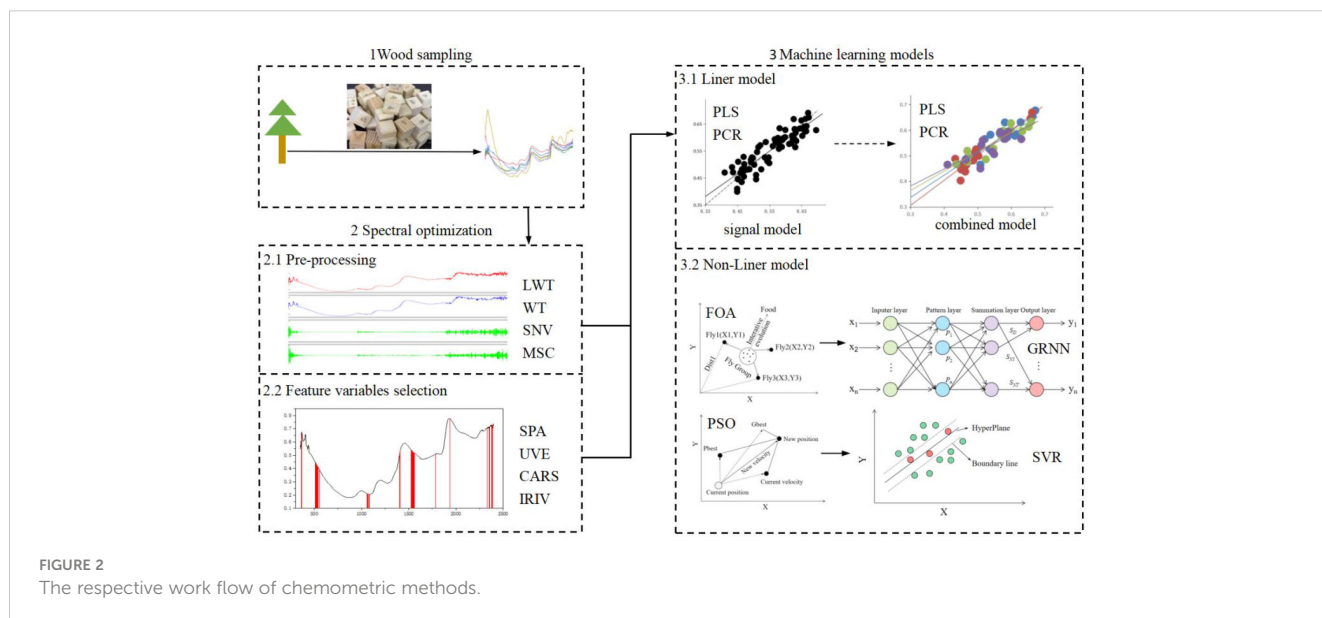
2.3.3 Machine learning models

Linear modeling method, namely PLS model, was used for determining the optimal pre-processing and variable selection method. Additionally, non-linear deep learning techniques including GRNN and PSO-SVM were performed using Matlab based on the selected variables. The parameter of Spread is of great importance for GRNN model. As FOA was applied to select the optimal Spread value. As for PSO-SVM model, PSO was used to optimize the parameters of penalty factor (C) and kernel function (g) in the radial basis function (RBF) kernel. Furthermore, in order to improve the accuracy of PSO-SVM model, three parameters of PSO-SVM, namely cross-validation number, maximum generation, and population, were optimized by Box-Behnken design of RSM method. GRNN and PSO-SVM models were established in Matlab. RSM was performed in Design-Expert Software 11 (Stat-Ease, Minneapolis, Minnesota, USA) (Figure 2).

3 Results

3.1 Wood density analysis

For evaluation the performance of various chemometric methods on different tree species, the calibration set and



prediction set for each tree species were divided using random sampling method. The statistical descriptive of wood density is demonstrated in Figure 3. Wood density values ranged from 0.576 to 1.124 g/cm³ among these tree species. The mean density value was different. Japanese elm from Jilin and Heilongjiang province presented a large mean value (1.047 and 1.058 g/cm³, respectively) and a higher standard deviation. In terms of data set, the range of density in the prediction set was within the corresponding calibration set. Additionally, regardless of the type of tree species, a similar mean value and standard deviation were obtained for a determined tree species between calibration and prediction set.

3.2 Comparison of various spectral de-noising methods

The selection of wavelet function is the first step of de-noising based on LWT or WT. Table 1 depicts the PLS models accuracy of four wavelet functions when the wavelet order and decomposition level were assumed to be 3 and 8, respectively. The overall accuracy of calibration set is higher than 0.7, regardless of the wavelet function used. For the precision of a determined tree species, various performance were obtained among these four wavelet functions. For example, different values of cross-validation set were achieved for *Tilia tuan* Szyszyl. when sym3 and bior1.3 functions were used. In contrast, db3 performs good with respect to the high R², RPD value and lower RMSE, RSD value for *Tilia tuan* Szyszyl., *Acer mono Maxim*, Japanese elm, and Dahurian larch. However, the same tree species of Chinese white poplar from Jilin and Heilongjiang province are not identical, with the optimal wavelet function of sym3 and bior1.3, respectively. In addition, the parameters of PLS models are different among these two locations. These demonstrated that geographical origin has impact on the prediction of density and spectral pre-processing based on Vis-NIR spectroscopy and LWT.

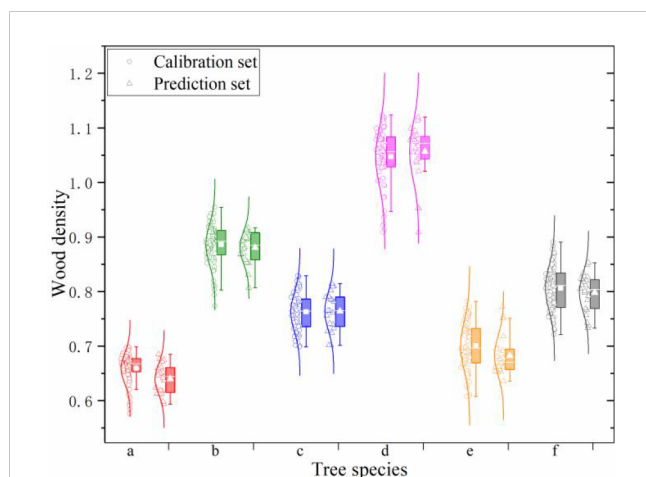


FIGURE 3
Statistical descriptive of wood density. (a–f) are *Tilia tuan* Szyszyl., *Acer mono Maxim*, Chinese white poplar (Jilin), Japanese elm, Chinese white poplar (Heilongjiang), and Dahurian larch, respectively.

Figure 4 displays the variation of PLS models using the optimal wavelet function in relation to the increasing wavelet orders from 2 to 8. The results show that a non-obvious trend was received with the enhance of wavelet orders. This is consistent with the results of Zhang et al. (Zhang et al., 2009). In contrast, the performance is relatively good for *Tilia tuan* Szyszyl., *Acer mono Maxim*, Japanese elm, and Dahurian larch when the wavelet order equals 4. In terms of Chinese white poplar from Jilin and Heilongjiang province, similar to the results of wavelet function, the optimal wavelet order is different with the values of 5 and 6, respectively. Additionally, for *Tilia tuan* Szyszyl., *Acer mono Maxim*, and Dahurian larch, the RMSE values of calibration and cross-validation set are similar, but the R² and RPD values slightly outperform the other orders when the order is 4.

When the optimal wavelet function and order were determined for these tree species, the performance of PLS models using various decomposition level (1–8) was analyzed. Unlike the results of wavelet order, Figure 5 illustrates that despite the similar RMSE values for a determined tree among various decomposition levels, the R² and RPD reach the largest value when the decomposition level increases from 1 to 8. This can be explained that the noise were removed from spectra with the increasing of decomposition level, while the useful information were regarded as noise when the decomposition level is too large. Considering the performance of calibration set and cross-validation set, the optimal decomposition level is 4 for Chinese white poplar (Jilin province), Japanese elm, and Dahurian larch. *Acer mono Maxim* and Chinese white poplar performed well when the decomposition level equals 5. For *Tilia tuan* Szyszyl., the optimal decomposition level is 6. As for Chinese white poplar harvested from two locations, the results of the optimal decomposition levels were similar to wavelet orders.

Considering the feasibility of LWT, the performance of the three traditional pre-processing techniques was further compared. Only the data of Chinese white poplar from Heilongjiang was shown in Figure 6. The results show that the LWT performs better with respect to the R² and RPD values in the calibration and cross-validation dataset. The RMSE and RSD values were smaller than that of corresponding raw model. In contrast, the performance of WT, MSC, and SNV was worse than LWT with a lower R² and RPD of cross-validation model, especially for the latter two methods. These results demonstrated that the PLS models using MSC and SNV approaches involve the overfitting problem in the prediction of wood density.

3.3 Feature variable selection of wood density

Figure 7 illustrates two dimensional (2D) correlation spectroscopy between wavelengths of various wood spectra. Regardless of tree species, the high correlation values (r) indicate that more redundant information or collinearity were exhibited, especially for Chinese white poplar (Jilin) (Figure 7C), Japanese elm (Figure 7D), and Dahurian larch (Figure 7F). Additionally, the correlation of adjacent spectral variables was higher than other

TABLE 1 The PLS model results of wood density with different wavelet functions.

Location	Tree species	Models	Indicators	Wavelet function			
				Haar	db3	sym3	bior1.3
Jilin province	<i>Tilia tuan</i> Szyszyl.	Calibration	R ²	0.812	0.823	0.814	0.814
			RMSE	0.012	0.011	0.012	0.012
			RPD	2.306	2.377	2.319	2.319
			RSD	1.820%	1.668%	1.820%	1.820%
		Cross-validation	R ²	0.800	0.801	0.803	0.812
			RMSE	0.012	0.012	0.012	0.012
			RPD	2.236	2.242	2.253	2.306
			RSD	1.820%	1.820%	1.820%	1.820%
	Acer mono Maxim	Calibration	R ²	0.755	0.740	0.759	0.748
			RMSE	0.018	0.019	0.018	0.018
			RPD	2.02	1.961	2.037	1.992
			RSD	2.031%	2.144%	2.031%	2.031%
		Cross-validation	R ²	0.639	0.654	0.705	0.665
			RMSE	0.022	0.022	0.021	0.021
			RPD	1.664	1.7	1.841	1.728
			RSD	2.483%	2.483%	2.370%	2.370%
	Chinese white poplar	Calibration	R ²	0.763	0.780	0.761	0.769
			RMSE	0.017	0.017	0.017	0.017
			RPD	2.054	2.132	2.046	2.081
			RSD	2.228%	2.228%	2.228%	2.228%
		Cross-validation	R ²	0.652	0.650	0.679	0.657
			RMSE	0.021	0.021	0.020	0.021
			RPD	1.695	1.69	1.765	1.707
			RSD	2.753%	2.753%	2.622%	2.753%
Japanese elm	Calibration	R ²	0.843	0.857	0.854	0.825	
		RMSE	0.020	0.020	0.020	0.022	
		RPD	2.524	2.644	2.617	2.39	
		RSD	1.910%	1.910%	1.910%	2.101%	
	Cross-validation	R ²	0.826	0.829	0.823	0.807	
		RMSE	0.022	0.022	0.022	0.023	
		RPD	2.397	2.418	2.377	2.276	
		RSD	2.101%	2.101%	2.101%	2.196%	
Heilongjiang province	Chinese white poplar	Calibration	R ²	0.760	0.749	0.749	0.758
			RMSE	0.022	0.022	0.022	0.022
			RPD	2.041	1.996	1.996	2.033
			RSD	3.140%	3.140%	3.140%	3.140%
		Cross-validation	R ²	0.705	0.689	0.726	0.730
			RMSE	0.024	0.025	0.024	0.024

(Continued)

TABLE 1 Continued

Location	Tree species	Models	Indicators	Wavelet function			
				Haar	db3	sym3	bior1.3
	Dahurian larch	Calibration	RPD	1.841	1.793	1.91	1.925
			RSD	3.425%	3.568%	3.425%	3.425%
			R ²	0.720	0.734	0.735	0.716
			RMSE	0.021	0.021	0.021	0.022
		Cross-validation	RPD	1.89	1.939	1.943	1.876
			RSD	2.605%	2.605%	2.605%	2.729%
			R ²	0.684	0.684	0.653	0.659
			RMSE	0.023	0.023	0.024	0.024
			RPD	1.779	1.779	1.698	1.712
			RSD	2.853%	2.853%	2.977%	2.977%

regions, which increases the computation time and complexity in modeling. Therefore, four variable selection approaches, SPA, UVE, CARS, and IRIV, were employed to select the feature variable related to wood density and decrease the redundant information.

Table 2 displays the prediction accuracy of four variable selection approaches on wood density using PLS models. For the number of selected variables, different numbers were obtained for UVE, CARS, and IRIV among these tree species, except for SPA. When using the SPA, 60 wavelengths were selected for *Tilia tuan* Szyszyl. and Chinese white poplar (Jilin), Chinese white poplar (Heilongjiang) and Dahurian larch also have the same numbers of selected variables (64). However, the indicators of PLS models were different, for example, the R² values of calibration set are 0.839 and 0.775 for *Tilia tuan* Szyszyl. and Chinese white poplar (Jilin), respectively. In terms of precision, the performance of cross-validation set was worse than that of calibration set for SPA. This demonstrated that despite a dimensionality reduction in the spectral matrix, SPA method has a overfitting problem for the

density prediction among these tree species. In contrast, CARS and IRIV have stable and better results for calibration and validation set. The optimal variable selection method of *Tilia tuan* Szyszyl. and Chinese white poplar (Jilin and Heilongjiang) is CARS, and IRIV performed better than other methods for *Acer mono Maxim*, Japanese elm, and Dahurian larch.

The distributions of selected variables for these tree species using the optimal method are shown in Figure 8. The selected bands of CARS and IRIV, 1157, 1171, 1370, 1597, 1811, 1830, 2200, and 2353 nm, are associated with hemicellulose, cellulose, and lignin (Ali et al., 2001; Sandak et al., 2011; Schwanninger et al., 2011; Yonobu and Tsuchikawa, 2003). This results are consistent with our previous studies (Li et al., 2020), indicating that the determination of wood density are related to chemical compounds. Additionally, comparing of the raw spectra and denoising spectra with LWT for these tree species (Figure 8), it can be found that the LWT makes the wood spectra smooth and has a similar trend with the corresponding raw spectra.

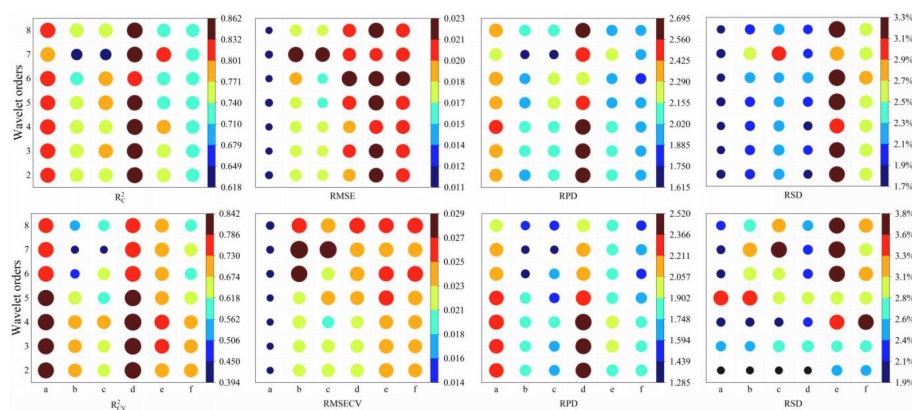
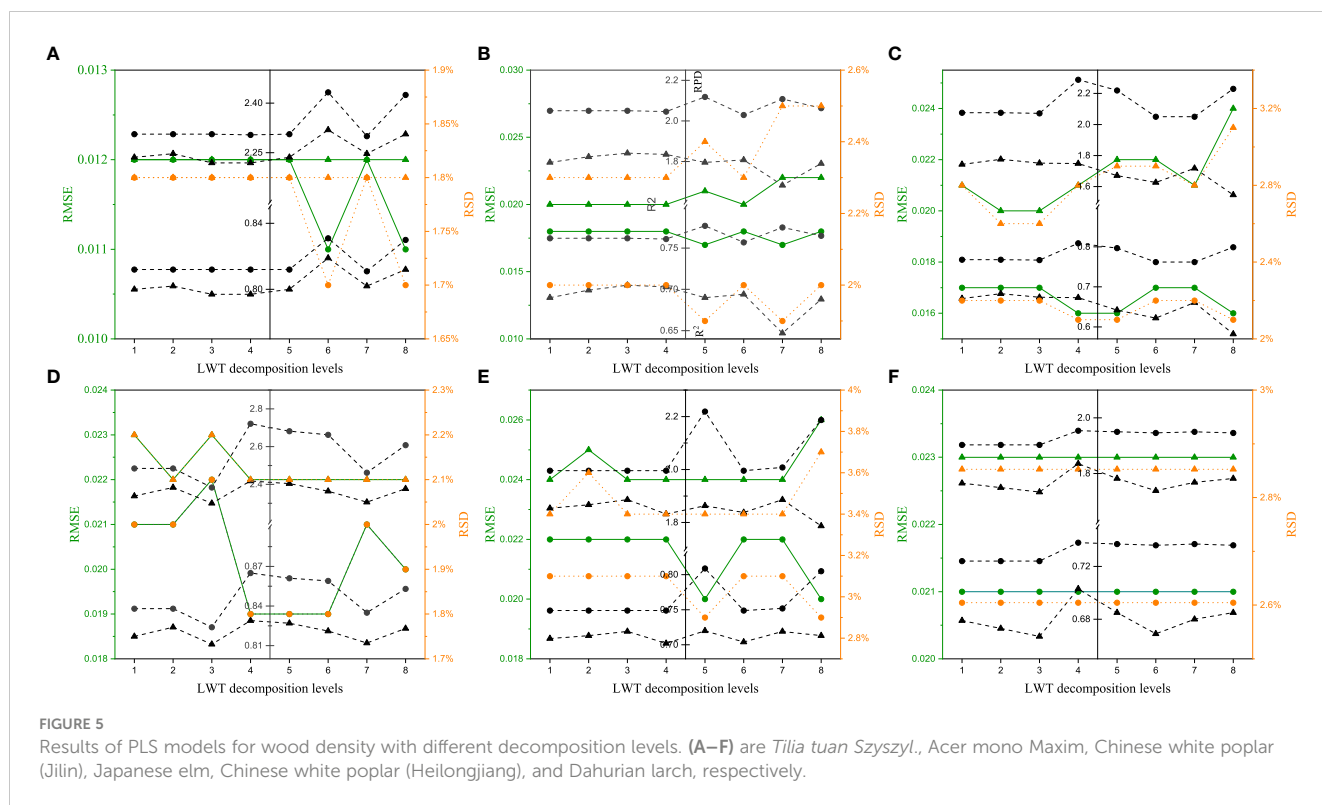


FIGURE 4 Results of PLS models for wood density with different wavelet orders. (a–f) are *Tilia tuan* Szyszyl., *Acer mono Maxim*, Chinese white poplar (Jilin), Japanese elm, Chinese white poplar (Heilongjiang), and Dahurian larch, respectively.

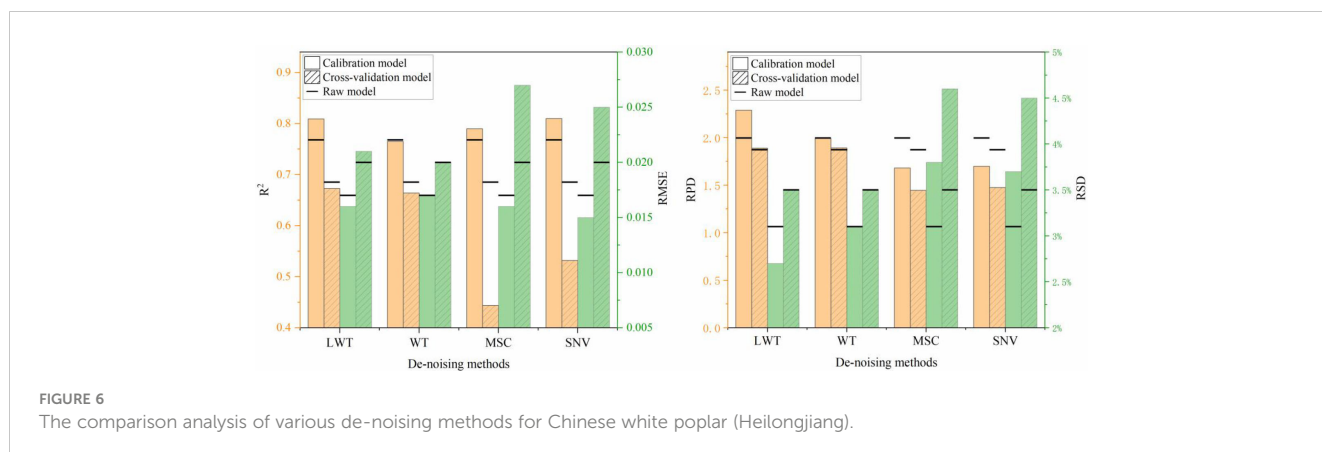


3.3.1 The optimization of non-linear calibration models

When the wood spectra were pre-processed by the optimal de-noising and variable selection methods, GRNN and SVM were used to build non-linear models based on these selected variables. In addition, FOA and PSO were employed to optimize the parameter of GRNN (Spread value) and SVM (penalty factor and kernel function), respectively. Regardless of tree species, the prediction accuracy resulting from LWT dataset was higher than that of raw prediction set. Considering four indicators, Chinese white poplar (Heilongjiang) delivers the best performance based on LWT dataset with R^2 , RMSE, RPD, and RSD value of 0.867, 0.013, 2.742 and 1.904%, respectively (Figure 9). Compared to FOA-GRNN models, four indicators are better, regardless of the pretreatment of

prediction set, when using the PSO-SVM, apart from Japanese elm and Chinese white poplar (Heilongjiang) (Table 3). These results show that tree species have affect on the selection of the optimal chemometric method.

As for PSO-SVM model, although the PSO was used to optimize the parameters of penalty factor and kernel function, three parameters, including population size, maximum generation, and the No. of cross-validation, also have influence on the performance of modeling. This may be the reason of low accuracy of Japanese elm and Chinese white poplar (Heilongjiang). Therefore, the Box-Behnken design with a three-factor and three-level of RSM was used to analyze the relationship between these three parameters and model performance. According to the experimental values and coded levels of these three factors (Table 4), the relationship between three



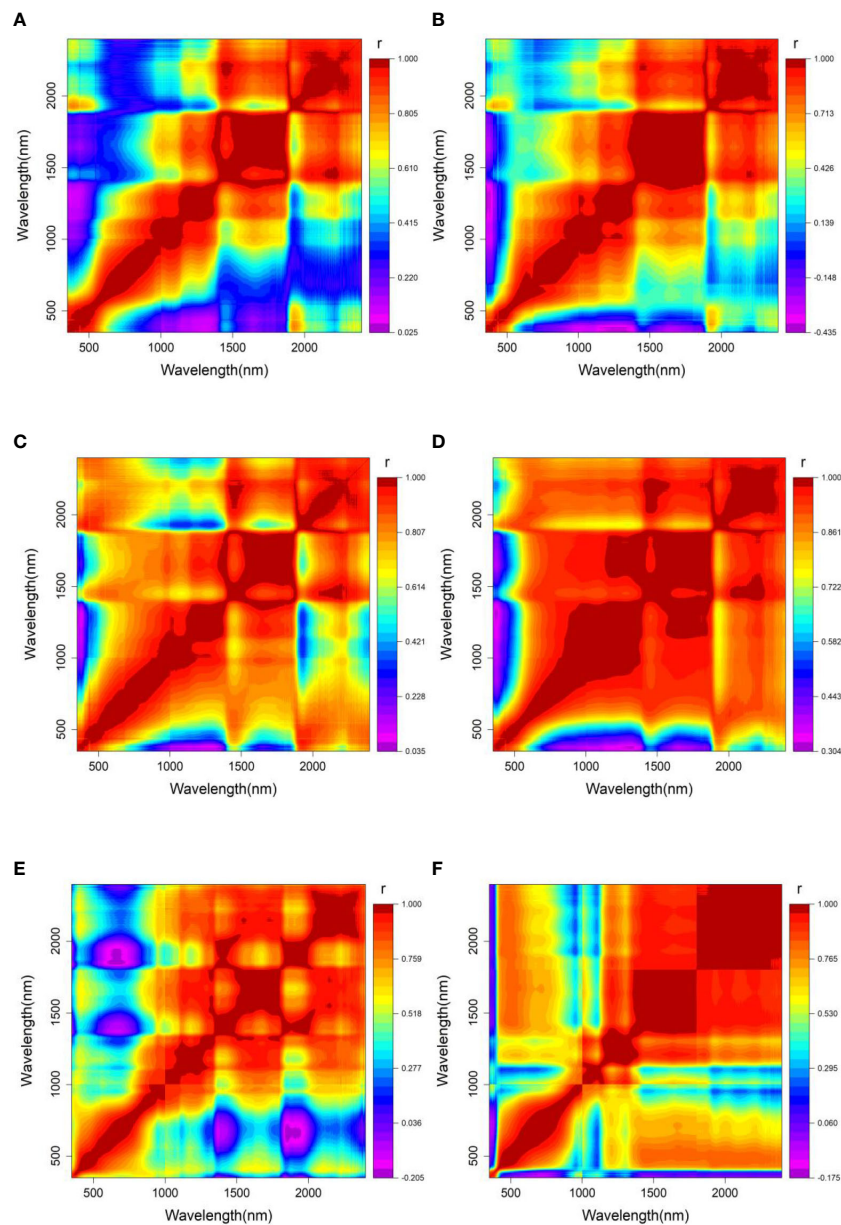


FIGURE 7

2-D correlation spectra of wavelength variables for each tree species with $\log(1/R)$ spectra. (A–F) are *Tilia tuan* Szyszyl., *Acer mono* Maxim, Chinese white poplar (Jilin), Japanese elm, Chinese white poplar (Heilongjiang), and Dahurian larch.

parameters and mean squared error of cross-validation (CVmse) for Dahurian larch are shown in Figure 10.

Figure 10 displays the response surface plot between three parameters and CVmse. The optimal parameter was determined when the CVmse has the lowest value. In terms of Dahurian larch, the CVmse value first reduced and then increased with the cross-validation number increasing from 5 to 15, when the maximum generation is a certain value. The lowest value of CVmse was achieved with the No. of cross-validation, maximum generation, and population size at 10, 75, and 40, respectively. According to the ANOVA results of Dahurian larch (data not shown), the CVmse was significantly influenced by cross-validation number ($p < 0.01$)

than by population size and maximum generation. For *Tilia tuan* Szyszyl. (data not shown), the minimum CVmse was obtained when the No. of cross-validation, maximum generation, and population size were 5, 50, and 20, respectively.

Figure 11 shows the wood density accuracy of RSM-PSO-SVM models based on the Box-Behnken design. In terms of accuracy, Japanese elm delivers the best performance with the R^2 of 0.955 and 0.862 for calibration and prediction set, respectively. For *Acer mono* Maxim, the improvement of R_p^2 value slightly outperforms corresponding R_c^2 value. This indicates that RSM-PSO-SVM models are more stable than that of PSO-SVM models. However, RSM-PSO-SVM had a poor performance for Chinese

TABLE 2 The comparison of various variable selection methods for each tree species.

Tree species	No. of Variables	R _c ²	RMSE	RSD	RPD	No. of Variables	R _c ²	RMSE	RSD	RPD
Full-PLS						LWT-PLS				
<i>Tilia tuan</i> Szyszyl.	2048	0.812	0.012	1.820%	2.306	2048	0.831	0.011	1.668%	2.433
<i>Acer mono Maxim</i>	2048	0.762	0.018	2.031%	2.05	2048	0.777	0.017	1.918%	2.118
<i>Chinese white poplar</i>	2048	0.768	0.017	2.228%	2.076	2048	0.809	0.016	2.097%	2.288
<i>Japanese elm</i>	2048	0.838	0.021	2.005%	2.485	2048	0.865	0.019	1.184%	2.722
<i>Chinese white poplar</i>	2048	0.749	0.022	3.140%	1.996	2048	0.809	0.019	2.711%	2.288
<i>Dahurian larch</i>	2048	0.724	0.021	2.605%	1.903	2048	0.738	0.021	2.605%	1.954
SPA-PLS						UVE-PLS				
<i>Tilia tuan</i> Szyszyl.	60	0.839	0.011	1.668%	2.492	1249	0.825	0.011	1.668%	2.39
<i>Acer mono Maxim</i>	69	0.782	0.018	2.301%	2.142	800	0.745	0.018	2.031%	1.98
<i>Chinese white poplar</i>	60	0.775	0.017	2.228%	2.108	733	0.679	0.020	2.622%	1.765
<i>Japanese elm</i>	73	0.840	0.021	2.005%	2.5	1139	0.841	0.021	2.005%	2.508
<i>Chinese white poplar</i>	64	0.765	0.022	3.140%	2.063	410	0.679	0.026	3.710%	1.765
<i>Dahurian larch</i>	64	0.682	0.023	2.853%	1.773	418	0.729	0.021	2.605%	1.921
CARS-PLS						IRIV-PLS				
<i>Tilia tuan</i> Szyszyl.	992	0.828	0.011	1.668%	2.411	5	0.806	0.012	1.820%	2.27
<i>Acer mono Maxim</i>	62	0.777	0.017	1.918%	2.118	20	0.819	0.016	1.805%	2.351
<i>Chinese white poplar</i>	35	0.800	0.016	2.097%	2.236	10	0.779	0.021	2.753%	2.127
<i>Japanese elm</i>	47	0.874	0.018	1.719%	2.817	25	0.901	0.012	1.146%	3.178
<i>Chinese white poplar</i>	23	0.783	0.021	2.997%	2.147	16	0.755	0.022	3.140%	2.02
<i>Dahurian larch</i>	13	0.767	0.020	2.481%	2.072	15	0.800	0.018	2.232%	2.236

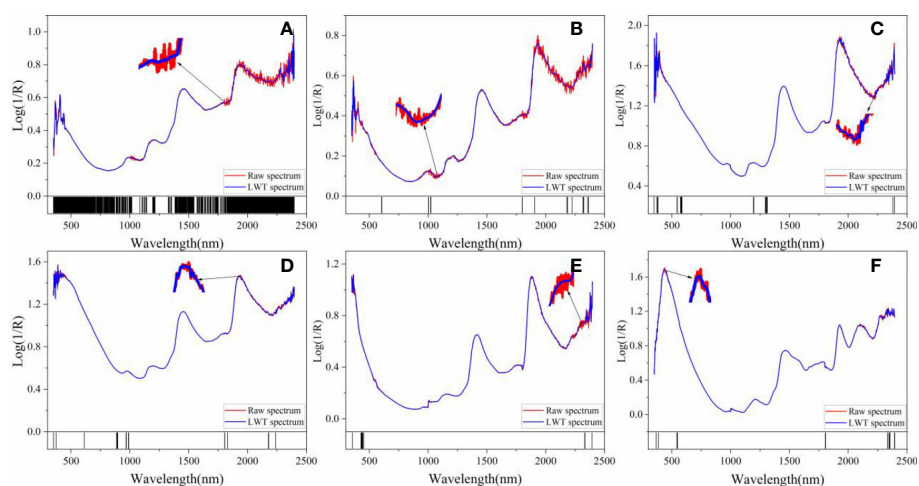


FIGURE 8

The distributions of selected variables by the optimal method for each tree species. (A–F) are *Tilia tuan* Szyszyl., *Acer mono Maxim*, Chinese white poplar (Jilin), Japanese elm, Chinese white poplar (Heilongjiang), and Dahurian larch, respectively.

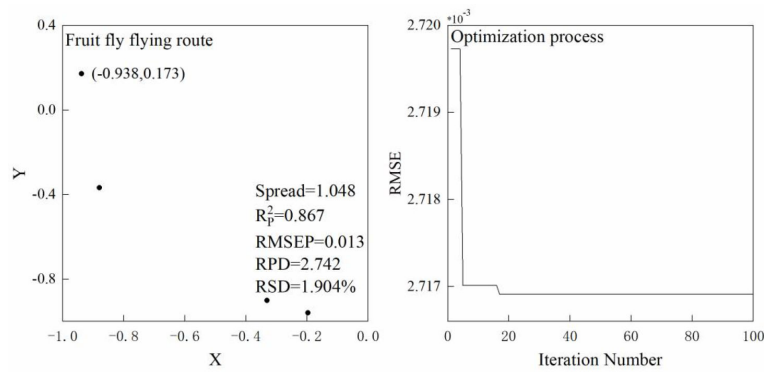


FIGURE 9 The optimization process of FOA-GRNN models for Chinese white poplar (Heilongjiang).

TABLE 3 The results of PSO-SVM models for prediction sets with different pretreatment.

Dataset	Location	Tree species	Variables	Calibration model				Prediction model			
				R ²	RMSE	RPD	RSD	R ²	RMSE	RPD	RSD
Prediction set with raw spectra	Jilin province	<i>Tilia tuan</i> Szyszyl.	992	0.908	0.008	3.297	1.213%	0.785	0.012	2.157	1.875%
		<i>Acer mono</i> Maxim	20	0.815	0.016	2.325	1.805%	0.725	0.015	1.907	1.703%
		Chinese white poplar	35	0.767	0.017	2.072	2.228%	0.721	0.018	1.893	2.355%
		Japanese elm	250	0.876	0.018	2.840	1.719%	0.815	0.022	2.325	2.082%
	Heilongjiang province	Chinese white poplar	23	0.784	0.021	2.152	2.997%	0.738	0.018	1.954	2.637%
		Dahurian larch	15	0.802	0.018	2.247	2.232%	0.733	0.017	1.935	2.130%
Prediction set with LWT de-noising	Jilin province	<i>Tilia tuan</i> Szyszyl.	992	0.915	0.008	3.430	1.213%	0.787	0.012	2.167	1.875%
		<i>Acer mono</i> Maxim	20	0.817	0.016	2.338	1.805%	0.727	0.015	1.914	1.703%
		Chinese white poplar	35	0.773	0.017	2.099	2.228%	0.734	0.018	1.939	2.355%
		Japanese elm	250	0.889	0.017	3.002	1.623%	0.830	0.021	2.425	1.987%
	Heilongjiang province	Chinese white poplar	23	0.784	0.021	2.152	2.997%	0.746	0.018	1.984	2.637%
		Dahurian larch	15	0.811	0.018	2.300	2.232%	0.747	0.017	1.988	2.130%

TABLE 4 Experimental values and coded levels of variable using Box–Behnken design.

Factor Levels	Variable		
	No. of cross-validation (A)	Maximum generation (B)	Population size (C)
-1	5	50	20
0	10	75	40
1	15	100	60

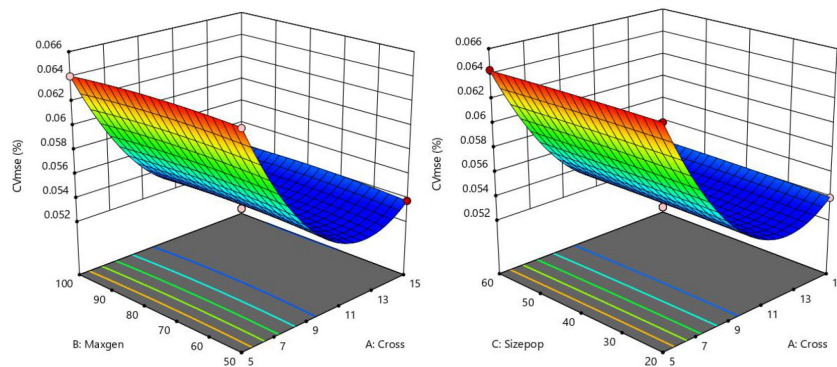


FIGURE 10 Response surface plot for interactions between three variables on CVmse for PSO-SVM models of Dahurian larch.

white poplar (Heilongjiang) ($R_p^2=0.752$, RMSEP=0.018, RPD=2.008, RSD=2.637%) when comparing with FOA-GRNN model ($R_p^2=0.867$, RMSEP=0.013, RPD=2.742, RSD=1.904%). These results demonstrated that there is not a universal chemometric method that works for all scenarios.

4 Discussions

Researchers pay more attention to spectral de-noising or feature variable selection to reduce the influence of irrelevant or interference information in the spectral analysis. A comparison of various chemometric methods including spectral pre-processing, feature variable selection, and the optimization of non-linear calibration models in the wood density prediction among

different tree species and geographical origins simultaneously were first analyzed using Vis-NIR spectroscopy in this study. The results demonstrated that LWT outperform WT, MSC, SNV, and raw spectra among these tree species for wood spectra optimization. There are few studies on LWT and WT de-noising in NIR spectra analysis. Abasi et al. (2019) employed WT to optimize Gala apple Vis-NIR spectra in the determination of quality parameters. The R^2 values were higher than 0.85 for soluble solids content, moisture content, and pH. In the forestry field, WT (Daubechies-5, db5) and LWT (db2) were used to optimize Populus davidiana and larch spectra, respectively in our previous studies (Li et al., 2018a; Li et al., 2018b). In this study, LWT with four wavelet functions including Haar, sym3, db3, and bior1.3 were compared simultaneously among various tree species from two locations. The results demonstrated that there has differences in the optimal de-noising

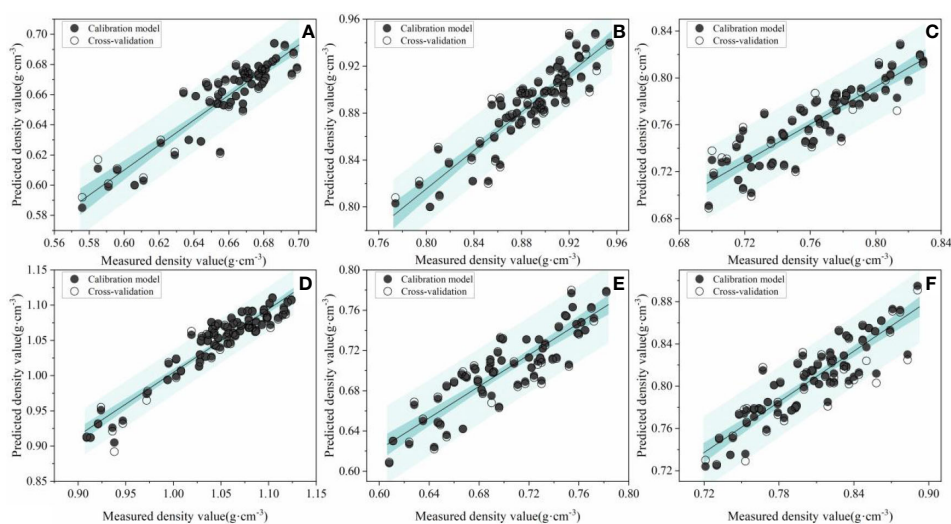


FIGURE 11 The scatter plots of predicted and measured values for RSM-PSO-SVM models. (A–F) are *Tilia tuan* Szyszyl, *Acer mono* Maxim, Chinese white poplar (Jilin), Japanese elm, Chinese white poplar (Heilongjiang), and Dahurian larch, respectively.

parameters of LWT among these tree species. Therefore, an appropriate pre-processing technique should be selected before building models.

As for feature variable selection, compared to full spectra de-noised by LWT, CARS and IRIV achieved the best results, and the spectral dimensionality was reduced by 51.56%, 99.02%, 98.29%, 98.78%, 98.88%, and 99.27% for *Tilia tuan* Szyszyl., Acer mono Maxim, Chinese white poplar (Jilin), Japanese elm, Chinese white poplar (Heilongjiang), and Dahurian larch, respectively. In terms of non-linear models, RSM-PSO-SVM delivers the best performance than corresponding FOA-GRNN models, except for Chinese white poplar harvest from Heilongjiang province. In order to better analyze the performance of these chemometric methods, two traditional linear models, i.e., PLS and PCR, were employed to build signal and combined models using raw spectra, respectively. Signal model is a model that includes one tree species. And combined model is a model that includes all tree species and geographical locations simultaneously. The results of these two kinds of models are shown in Table 5, respectively.

Comparison of the accuracy of PLS and PCR method based on raw spectra (Table 5), the calibration set outperforms the cross-validation set with a higher R^2 , RPD and smaller RMSE, RSD value, regardless of signal and combined models. In addition, PLS model

provides high accuracy on various wood density prediction than that of PCR model. However, the performance of combined models was worse than signal model, especially for PCR method.

Table 6, 7 show the prediction accuracy of signal models and combined models, respectively. Similar to the results of calibration set, the PLS method achieved a higher prediction accuracy than PCR model. Additionally, the combined model delivers the worst performance on prediction dataset. Compared to the optimal model of RSM-PSO-SVM, the PLS approach is relatively bad except the Chinese white poplar from Jilin province.

5 Conclusions

This study demonstrates the feasibility of using Vis-NIR spectra combined with various chemometric methods including spectral de-noising, feature variables selection, and the optimization of modeling parameters, to predict wood density. LWT is excellent for spectral de-noising when comparing traditional methods. CARS and IRIV outperforms in the feature variables selection among these tree species. In terms of linear and non-linear calibration models, RSM-PSO-SVM delivers the best performance except for Chinese white poplar. The optimal model of Chinese white poplar from Jilin

TABLE 5 The results of PLS and PCR signal and combined models for various tree species.

Models	Indicators	Signal models						Combined models
		Jilin province			Heilongjiang province			/
		<i>Tilia tuan</i> Szyszyl.	Acer mono Maxim	Chinese white poplar	Japanese elm	Chinese white poplar	Dahurian larch	/
PLS	R_c^2	0.812	0.762	0.768	0.838	0.749	0.724	0.660
	RMSEC	0.012	0.018	0.017	0.021	0.022	0.021	0.080
	RPD	2.306	2.05	2.076	2.485	1.996	1.903	1.715
	RSD	1.820%	2.031%	2.228%	2.005%	3.140%	2.605%	9.771%
	R_{cv}^2	0.799	0.722	0.685	0.825	0.715	0.690	0.561
	RMSECV	0.012	0.020	0.020	0.022	0.024	0.023	0.091
	RPD	2.23	1.897	1.782	2.39	1.873	1.796	1.509
	RSD	1.820%	2.257%	2.622%	2.101%	3.425%	2.853%	11.115%
PCR	R_c^2	0.785	0.751	0.752	0.836	0.725	0.703	0.546
	RMSEC	0.013	0.018	0.018	0.021	0.023	0.022	0.092
	RPD	2.157	2.004	2.008	2.469	1.907	1.835	1.484
	RSD	1.972%	2.031%	2.360%	2.005%	3.282%	2.729%	11.237%
	R_{cv}^2	0.770	0.712	0.668	0.813	0.699	0.685	0.507
	RMSECV	0.013	0.020	0.021	0.022	0.024	0.023	0.097
	RPD	2.085	1.863	1.736	2.312	1.823	1.782	1.424
	RSD	1.972%	2.257%	2.753%	2.101%	3.425%	2.853%	11.847%

TABLE 6 The prediction results of signal models for each tree species.

Models	Indicators	Jilin province				Heilongjiang province	
		<i>Tilia tuan</i> Szyszyl.	Acer mono Maxim	Chinese white poplar	Japanese elm	Chinese white poplar	Dahurian larch
PLS	R_p^2	0.771	0.522	0.758	0.855	0.797	0.714
	RMSEP	0.107	0.144	0.110	0.076	0.096	0.114
	RPD	2.090	1.446	2.033	2.626	2.219	1.870
	RSD	16.722%	16.352%	14.395%	7.191%	14.062%	14.282%
PCR	R_p^2	0.648	0.520	0.736	0.838	0.714	0.644
	RMSEP	0.133	0.144	0.115	0.081	0.114	0.127
	RPD	1.685	1.443	1.946	2.485	1.870	1.676
	RSD	20.785%	16.352%	15.049%	7.664%	16.699%	15.911%

TABLE 7 The prediction results of combined models for each tree species.

Location	Tree species	Indicators	Combined models with PLS				Combined models with PCR			
			All tree species	Jilin	Heilong jiang	Chinese white poplar	All tree species	Jilin	Heilong jiang	Chinese white poplar
Jilin province	<i>Tilia tuan</i> Szyszyl.	R_p^2	-	-			-	-		
		RMSEP	-	-			-	-		
		RPD	-							
		RSD	-							
	Acer mono Maxim	R_p^2	-	-			-	-		
		RMSEP	-	-			-	-		
		RPD	-							
		RSD	-							
	Chinese white poplar	R_p^2	-	-		0.496	-	-		0.445
		RMSEP	-	-		0.159	-	-		0.167
		RPD	-			1.409				1.342
		RSD	-			20.807%				21.854%
	Japanese elm	R_p^2	-	-			-	-		
		RMSEP	-	-			-	-		
		RPD	-							
		RSD	-							
Heilongjiang province	Chinese white poplar	R_p^2	-		0.402	0.714	-		0.386	0.463
		RMSEP	-		0.165	0.114	-		0.167	0.156
		RPD	-		1.293	1.870			1.276	1.365
		RSD	-		24.170%	16.699%			24.463%	22.851%
	Dahurian larch	R_p^2	-		-		-		-	
		RMSEP	-		-		-		-	
		RPD	-							
		RSD	-							

Blank indicates that no experiment is scheduled, “-” indicates that the prediction result is negative.

and Heilongjiang province are PLS and FOA-GRNN model, respectively. These results indicate that the geographical location has an effect on the selection of chemometric methods in wood density determination. Additionally, in order to overcome the difference of the optimal model, model transfer will be used to predict wood properties of different locations in future studies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YiL analyzed the data and wrote the original draft. BV contributed to editing. FH performed formal analysis. ZP performed supervision. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Outstanding Doctoral Introduction Fund of School, grant number NDYB2020-11; the Natural Science Foundation of Inner Mongolia Autonomous Region, grant number 2021BS03019; and the Science and Technology Project of Inner Mongolia, grant number 2020GG0078.

References

- Abasi, S., Minaei, S., Jamshidi, B., Fathi, D., and Khoshnaghza, M. H. (2019). Rapid measurement of apple quality parameters using wavelet de-noising transform with Vis/NIR analysis. *Sci. Hortic.* 252, 7–13. doi: 10.1016/j.scienta.2019.02.085
- Ali, M., Emsley, A. M., Herman, H., and Heywooda, R. J. (2001). Spectroscopic studies of the ageing of cellulosic paper. *Polymer* 42, 2893–2900. doi: 10.1016/S0032-3861(00)00691-1
- Araújo, M. C. U., Saldanha, T. C. B., Galvão, R. K. H., Yoneyama, T., Chame, H. C., and Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics Intelligent Lab. Systems* 57, 65–73. doi: 10.1016/S0169-7439(01)00119-8
- Bian, X. H., Ling, M. X., Chu, Y. Y., Liu, P., and Tan, X. Y. (2022). Spectral denoising based on Hilbert-Huang transform combined with f-test. *Front. Chem.* 10. doi: 10.3389/fchem.2022.949461
- Bian, X. H., Wang, K. Y., Tan, E., Diwu, P. Y., Zhang, F., and Guo, Y. G. (2020). A selective ensemble preprocessing strategy for near-infrared spectral quantitative analysis of complex samples. *Chemometr. Intell. Lab. Syst.* 197, 103916. doi: 10.1016/j.chemolab.2019.103916
- Carvalho, J. K., Moura-Bueno, J. M., Ramon, R., Almeida, T. F., Naibo, G., Martins, A. P., et al. (2022). Combining different pre-processing and multivariate methods for prediction of soil organic matter by near infrared spectroscopy (NIRS) in southern Brazil. *Geoderma Regional* 29, e00530. doi: 10.1016/j.geodrs.2022.e00530
- Centner, V., Massart, D. L., De Noord, O. E., Jong, S., Vandeginste, B. M., and Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Anal. Chem.* 68, 3851–3858. doi: 10.1021/ac960321m
- Dantas, H. V., Barbosa, M. F., Nascimento, E. C. L., Moreira, P. N. T., Galvão, R. K. H., and Araújo, M. C. U. (2013). An automatic flow system for NIR screening analysis of liquefied petroleum gas with respect to propane content. *Talanta* 106, 158–162. doi: 10.1016/j.talanta.2012.12.024
- Dotto, A. C., Dalmolin, R. S. D., Caten, A. T., and Grunwald, S. (2018). A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by vis-NIR spectra. *Geoderma* 314, 262–274. doi: 10.1016/j.geoderma.2017.11.006
- Fernandes, J. N., Dos Santos, L. M. B., Thais, C. C., Pavan, M. G., Garcia, G. A., David, M. R., et al. (2018). Rapid, noninvasive detection of zika virus in aedes aegypti mosquitoes by near-infrared spectroscopy. *Sci. Adv.* 4 (5), eaat0496. doi: 10.1126/sciadv.aat0496
- Fernandes, A., Lousada, J., Morais, J., Xavier, J., Pereira, J., and Pedro, M. P. (2013). Comparison between neural networks and partial least squares for intra-growth ring wood density measurement with hyperspectral imaging. *Comput. Electron Agric.* 94, 71–81. doi: 10.1016/j.compag.2013.03.010
- Geladi, P., Macdougall, D., and Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 39, 491–500. doi: 10.1366/0003702854248656
- Gozukara, G., Akça, E., Dengiz, O., Kapur, S., and Adak, A. (2022). Soil particle size prediction using vis-NIR and pXRF spectra in a semiarid agricultural ecosystem in central Anatolia of türkiye. *Catena* 217, 106514. doi: 10.1016/j.catena.2022.106514
- Huang, H. L., Mai, X. J., and Huang, H. D. (2018). Resources and development research of hainan yellow pear in China. *Chin. J. Agric. Resour. Reg. Plan.* 39, 123–129. doi: 10.7621 / cjarrp. 1005-9121. 20180916
- Ielpo, P., Leardi, R., Pappagallo, G., and Uricchio, V. F. (2017). Tools based on multivariate statistical analysis for classification of soil and groundwater in apulian agricultural sites. *Environ. Sci. Pollut. Res.* 24, 13967–13978. doi: 10.1007/s11356-016-7944-y
- Krajnc, L., Hafner, P., and Gričar, J. (2021). The effect of bedrock and species mixture on wood density and radial wood increment in pubescent oak and black pine. *For. Ecol. Manage.* 481, 118753. doi: 10.1016/j.foreco.2020.118753
- Li, Y., Li, Y. X., Li, W. B., and Jiang, L. C. (2018b). Model optimization of wood property and quality tracing based on wavelet transform and NIR spectroscopy. *Spectrosc Spect Anal.* 38, 1384–1392. doi: 10.3964/j.issn.1000-0593(2018)05-1384-09
- Li, Y., Via, B. K., Cheng, Q. Z., and Li, Y. X. (2018a). Lifting wavelet transform denoising for model optimization of vis-NIR spectroscopy to predict wood tracheid length in trees. *Sensors* 18, 4306. doi: 10.3390/s18124306
- Li, Y., Via, B. K., Cheng, Q. Z., and Li, Y. X. (2019). Prediction of microfibril angle in dahurian larch wood using visible-near infrared spectroscopy and chemometric techniques. *J. Near Infrared Spectroscopy*. 0, 1–8. doi: 10.1177/0967033519849725

Acknowledgments

The authors would like to thank the editor and reviewers for their constructive comments.

Conflict of interest

Author FH is employed by Laboratory Zhejiang Huadong Forestry Engineering Consulting and Design corporation, Hangzhou, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1121287/full#supplementary-material>

- Li, Y., Via, B. K., and Li, Y. X. (2020). Lifting wavelet transform for vis-NIR spectral data optimization to predict wood density. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 240, 118566. doi: 10.1016/j.saa.2020.118566
- Li, Y., Wang, G. Z., Guo, G. S., Li, Y. X., Via, B. K., and Pei, Z. Y. (2022). Spectral pre-processing and multivariate calibration methods for the prediction of wood density in Chinese white poplar by visible and near infrared spectroscopy. *Forests* 13, 62–62. doi: 10.3390/F13010062
- Liang, S., Haff, R. P., Hua, S. S. T., Munyaneza, J. E., Mustafa, T., and Sarreal, S. B. L. (2018). Nondestructive detection of zebra chip disease in potatoes using near-infrared spectroscopy. *Biosyst. Eng.* 166, 161–169. doi: 10.1016/j.biosystemseng.2017.11.019
- Ling, M. X., Bian, X. H., Wang, S. S., Huang, T., Liu, P., Wang, S. Y., et al. (2022). A piecewise mirror extension local mean decomposition method for denoising of near-infrared spectra with uneven noise. *Chemometr. Intell. Lab. Syst.* 230, 104655. doi: 10.1016/j.chemolab.2022.104655
- Mansouri, M. A., Ziemons, E., Sacré, P. Y., Kharbach, M., Barra, I., Cherrah, Y., et al. (2021). Classification of polymorphic forms of fluconazole in pharmaceuticals by FT-IR and FT-NIR spectroscopy. *J. Pharm. Biomed. Analysis.* 196, 113922. doi: 10.1016/j.jpba.2021.113922
- Marta, B. S., Marta, F. G., Calle, J. L. P., Barbero, G. F., Ayuso, J., and Palma, M. (2022). Comparison of different processing approaches by SVM and RF on HS-MS eNose and NIR spectrometry data for the discrimination of gasoline samples. *Microchemical J.* 172, 106893. doi: 10.1016/j.microc.2021.106893
- Mojilovic, A., Popovic, M. V., and Rackov, D. M. (2000). On the selection of an optimal wavelet basis for texture characterization. *IEEE Trans. Image Process.* 9, 2043–2050. doi: 10.1109/83.887972
- Resquin, F., Navarro-Cerrillo, R. M., Carrasco-Letelier, L., and Casnati, C. R. (2019). Influence of contrasting stocking densities on the dynamics of above-ground biomass and wood density of eucalyptus benthamii, eucalyptus dunii, and eucalyptus grandis for bioenergy in Uruguay. *For. Ecol. Manage.* 438, 63–74. doi: 10.1016/j.foreco.2019.02.007
- Sandak, A., Sandak, J., and Negri, M. (2011). Relationship between near-infrared (NIR) spectra and the geographical provenance of timber. *Wood Sci. Technol.* 45, 35–48. doi: 10.1007/s00226-010-0313-y
- Schimleck, L. R., Antony, F., Mora, C., and Dahlen, J. (2018). Comparison of whole-tree wood property maps for 13- and 22-year-old loblolly pine. *Forests* 9, 287. doi: 10.3390/f9060287
- Schimleck, L. R., Mora, C., and Daniels, R. F. (2003). Estimation of the physical wood properties of green pinus taeda radial samples by near infrared spectroscopy. *Can. J. For. Res.* 33, 2297–2305. doi: 10.1139/x03-173
- Schwanninger, M., Rodrigues, J. C., and Fackler, K. (2011). A review of band assignments in near infrared spectra of wood and wood components. *J. Near Infrared Spectroscopy.* 19 (5), 287–308. doi: 10.1255/jnirs.955
- Tigabu, M., Daneshvar, A., Wu, P. F., Ma, X. Q., and Odén, P. C. (2020). Rapid and non-destructive evaluation of seed quality of Chinese fir by near infrared spectroscopy and multivariate discriminant analysis. *New Forests.* 51, 395–408. doi: 10.1007/s11056-019-09735-8
- Toscano, G., Leoni, E., Gasperini, T., and Picchi, G. (2022). Performance of a portable NIR spectrometer for the determination of moisture content of industrial wood chips fuel. *Fuel* 320, 123948. doi: 10.1016/j.fuel.2022.123948
- Wang, K., Chi, G., Lau, R., and Chen, T. (2011). Multivariate calibration of near infrared spectroscopy in the presence of light scattering effect: a comparative study. *Analytical Letters.* 44, 824–836. doi: 10.1080/00032711003789967
- Yan, Y. L., Chen, B., Zhu, D. Z., and Zhang, L. D. (2013). *Near infrared spectroscopy principle, technology and application* (Beijing: China Light Industry Press).
- Yonenobu, H., and Tsuchikawa, S. (2003). Near-infrared spectroscopic comparison of antique and modern wood. *Appl. spectroscopy.* 57, 1451–1453. doi: 10.1366/000370203322554635
- Yun, Y. H., Bin, J., Liu, D. L., Xu, L., Yan, T. L., Cao, D. S., et al. (2019). A hybrid variable selection strategy based on continuous shrinkage of variable space in multivariate calibration. *Analytica Chimica Acta* 1058, 58–69. doi: 10.1016/j.aca.2019.01.022
- Yun, Y. H., Wang, W. T., Deng, B. C., Lai, G. B., Liu, X. B., Ren, D. B., et al. (2015). Using variable combination population analysis for variable selection in multivariate calibration. *Anal. Chim. Acta* 862, 14–23. doi: 10.1016/j.aca.2014.12.048
- Yun, Y. H., Wang, W. T., Tan, M. L., Liang, Y. Z., Li, H. D., Cao, D. S., et al. (2014). A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Analytica Chimica Acta* 807, 36–43. doi: 10.1016/j.aca.2013.11.032
- Zahir, S. A. D. M., Omar, A. F., Jamlos, M. F., Azmi, M. A. M., and Muncan, J. (2022). A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. *Sensors Actuators A:Physical.* 338, 113468. doi: 10.1016/j.sna.2022.113468
- Zhang, Z. Y., Li, Y. X., and Li, Y. (2022). Prediction approach of larch wood density from visible-near-infrared spectroscopy based on parameter calibrating and transfer learning. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1006292
- Zhang, G. J., Li, L. N., Li, Q. B., and Xu, Y. P. (2009). Application of denoising and background elimination based on wavelet transform to blood glucose noninvasive measurement of near infrared spectroscopy. *J. Infrared Millimeter Waves.* 28, 107–110. doi: 10.3724/SP.J.1010.2009.00107
- Zhao, R. J., Xing, X. T., Lü, J. X., and Zhang, J. Z. (2012). Estimation of wood mechanical properties of eucalyptus pellita by near infrared spectroscopy. *Scientia Silvae sincae.* 48 (6), 106–111. doi: 1001-7488(2012)06-0106-06
- Zhao, R. J., Zhang, L., Huo, X. M., and Ren, H. Q. (2010). Microfibril angle prediction of eucalyptus pellita wood samples based on radial and tangential section by near infrared spectroscopy. *Spectrosc. Spectral Analysis.* 30 (9), 2355–2359. doi: 10.3964/jissn-1000-0593(2010)09-2355-05