

# Determination of Geographical Origin and Tree Species Using Vis-NIR and Chemometric Methods

Ying Li  
Brian K. Via  
Yaoxiang Li  
Guozhong Wang

---

## Abstract

The variation of wood properties between different geographical origin and tree species has an important influence on end use applications. This study aimed to investigate the feasibility of wood origin and species classification based on visible and near infrared spectroscopy and chemometric methods. The influence of geographical origin on tree species identification also was analyzed. A total of 530 samples with 2 origins and 5 tree species were collected for analysis. The raw reflectance spectra were preprocessed by spectral transformation technique, and nonlinear discrimination models were built by support vector machine (SVM) using various spectral forms. Three algorithms—grid search (GS), genetic algorithm (GA), and particle swarm optimization (PSO)—were applied to optimize the parameters of SVM models, respectively. Regardless of spectral forms and optimization techniques, the prediction accuracy was lower than that of the calibration set for wood origin and tree species identification. Except for reflectance spectra, prediction accuracy of 100 percent was obtained based on SVM in combination with three algorithms for origin discrimination. However, SVM in combination with reflectance spectra and GS technique achieved the best prediction accuracy (93.18%) for tree species identification. These results demonstrated that visible and near infrared spectroscopy combined with chemometric techniques can be used for geographical origin and tree species determination.

---

Wood properties (e.g., physical, chemical, anatomical, and other mechanical parameters) are essential for forest cultivation and resource use (Vega et al. 2021, Villa et al. 2021). However, these property parameters are influenced by many factors. Two kinds of factors are of importance, one of which is genetics, that is, the differences in genotype that can produce different phenotypes or tree species (Hamilton and Potts 2007). The other factor is the environmental aspect including items such as origin and climate (Cordier et al. 2021). Therefore, different uses are employed for various tree species because of the differences in wood properties.

However, in order to maximize the economic benefits, some rare or expensive wood may be displaced by the cheap or illegal wood because of its similar appearance. For example, in the musical instruments industry, the best raw material for guitar soundboards and fingerboards is rosewood from Brazil (Phillips 2009), which may be displaced by North American rosewood by some companies. Likewise, in the furniture industry, the classical Chinese furniture made of yellow rosewood from Hainan

is expensive and rare because the wood is derived from one of the most endangered trees in China (Huang et al. 2018), whereas, rosewood products from other locations are relatively cheap. However, it is difficult to distinguish these two types of wood even with perceptiveness or experience. Therefore, a fast, accurate, and nondestructive technology is needed to classify tree species and geographical origin for large-scale samples.

---

The authors are, respectively, Lecturer (yingli@nefu.edu.cn) and Professor (xhk1512@163.com [corresponding author]), College of Energy and Transportation Engineering, Inner Mongolia Agric. Univ., Hohhot 010018, China; Professor, Forest Products Development Center, School of Forestry and Wildlife Sci., Auburn Univ., Auburn, Alabama (brianvia@auburn.edu); and Professor, College of Engineering and Technol., Northeast Forestry Univ., Harbin 150040, China (yaoxiangli@nefu.edu.cn). This paper was received for publication in February 2022. Article no. FPJ-D-22-00011R1.

©Forest Products Society 2022.  
Forest Prod. J. 72(3):147–154.  
doi:10.13073/FPJ-D-22-00011

Visible and near infrared (Vis-NIR) spectroscopy has been widely used for qualitative and quantitative analysis in many applications (van Kollenburg et al. 2021, Kapoor et al. 2022, Paltseva et al. 2022). Vis-NIR spectra include a visible region, which contains wavelengths from 380 nm to 780 nm; and a near infrared region, which contains short-wavelength (780 to 1,100 nm) and long-wavelength (1,100 to 2,526 nm) near infrared regions (Scotter 2005). Sample information can be extracted by Vis-NIR spectra and appropriate chemometric techniques. Thus, the selection and optimization of chemometric methods are essential for the spectral analysis. For the quantitative analysis of spectral data, it is not necessary to measure the concentration of samples (Lu 2007). Therefore, the prediction results are highly affected by spectral quality and chemometric methods. Some spectral preprocessing approaches are able to remove noise or reduce irrelevant variables. These techniques include multiplication scatter correction, standard normalized variate, successive projections algorithm, etc. The performance of a single method or combination of methods has been analyzed in many studies (Galvão et al. 2007, Cai et al. 2010, Jiang et al. 2019). Another crucial part of chemometric analysis is building the calibration model and including therein linear and nonlinear modelling techniques.

In the forestry field, the traditional techniques of geographical origin and tree species discrimination are laboratory detection methods, or the sensory means performed by the relevant working personnel, which have the disadvantages of being time-consuming, labor-intensive, and destructive. *Populus davidiana*, *Tilia tuan* Szyszyl., *Ulmus pumila* L., *Acer mono* Maxim., and *Larix gmelinii* have been widely planted as the ecological public welfare forest with high value in terms of ecological, economic, and social benefits. For instance, the stems of *Ulmus pumila* L. are used as materials in furniture and vehicles; and additionally, it is a medicinal plant with bark, leaves, and winged fruits that could play an important role in treating anxiety and diuretics (Wang et al. 2020). Additionally, *Tilia tuan* Szyszyl. is a medicinal and honey plant that is essential for reducing the risk for heart disease, high blood pressure, constipation, and insomnia (Wu and Zhang 2022). This study aimed to investigate the feasibility of origin and tree species classification using Vis-NIR spectra in combination with chemometric methods. Spectral transformation algorithm was employed to analyze the performance of various spectral data. A support vector machine (SVM) was used to classify wood geographical origin and tree species. Three algorithms (i.e., grid search [GS], genetic algorithm [GA], and particle swarm optimization [PSO]), were adopted to optimize the parameters of SVM models, respectively.

## Materials and Methods

### Wood sample origin information

In order to better analyze the performance of Vis-NIR spectra in combination with chemometric methods for wood origin and tree species identification, 354 samples with 4 species were collected from the Jinsha forest of Heilongjiang Province (131°08' to 131°21'E, 45°44' to 45°53'N). One hundred seventy six samples with two species came from the Jingshan forest of Jilin Province (126°30' to 127°16'E, 42°06' to 42°48'N), China (Fig. 1). Taking into account the influence of origin on tree species discrimina-

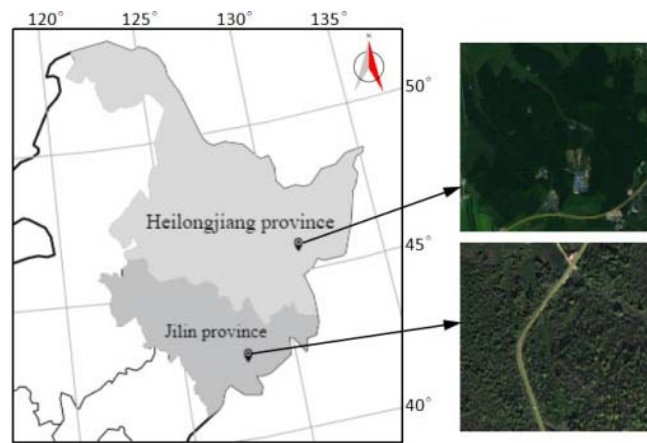


Figure 1.—Geographical location of samples.

tion, the same tree species of Chinese white poplar samples of the same age-class were harvested from two regions, respectively. The first location (Heilongjiang Province) has temperate continental monsoon climate and dark brown soil, with a total area of 2,765 hm<sup>2</sup>. Average annual relative humidity is between 65 and 75 percent (Jiang 2014). The second region (Jilin Province) is East Asian monsoon climate and average elevation is 560 m (Tan 2012). The main soil types are black soil and loess soil with soil thickness of 35 to 50 cm (Liu et al. 2012).

### Samples preparation and spectral measurement

Twenty seven standard woods, including eight *Acer mono* Maxim., four *Populus davidiana*, eight *Ulmus pumila* L., and seven *Tilia tuan* Szyszyl., were randomly harvested from the first location. Five *Populus davidiana* and five *Larix gmelinii* were collected from the second location at the same time. These tree species were represented by A, B, C, D, E, and F, respectively. Five-centimeter disks were made from each tree and 530 wood samples with the dimensions of 2 by 2 by 2 cm<sup>3</sup> were collected from these two regions for spectral measurement (Table 1).

The reflectance spectra of the wood samples were measured on the cross-section using a spectrometer with the wavelength range from 350 to 2,500 nm (LabSpec Pro, Analytical Spectral Devices, Boulder, Colorado, USA). The spectral resolution was 3 nm@700 nm, 10 nm@1,400/2,100 nm. Two spectra of each sample were measured and the average spectrum was treated as the raw spectrum after preheating the spectrometer for half an hour. A white panel, primarily consisting of everypolytetrafluoroethylene

Table 1.—Wood sample information for two regions. A–D are *Acer mono* Maxim., *Populus davidiana*, *Ulmus pumila* L., and *Tilia tuan* Szyszyl. from the Heilongjiang Province, respectively; E and F are *Populus davidiana* and *Larix gmelinii* of the Jilin Province, respectively.

Region	Location 1				Location 2		Total
	A	B	C	D	E	F	
Training set	70	61	74	61	65	67	398
Test set	23	20	25	20	22	22	132

(PTFE), was used to calibrate every three wood samples (Davari et al. 2022).

### Chemometric analysis

*Spectral preprocessing.* The quantitative analysis of wood origin and tree species identification is mainly affected by the spectral quality and discrimination models. In terms of spectral data, regardless of the influence of environment factors and instrumental background, the various spectral forms are essential for classification results. For this reason, the raw reflectance spectra (R) were subjected to the reciprocal (1/R) and logarithm reflectance [ $\log(1/R)$ ] based on spectral transformation technology to reduce the nonlinear relationship and multiplicative light-scattering effects (Zhuang et al. 2017). The statistics formulas of spectral transformation were determined by the following equations:

$$R = X_{mn} \quad (1)$$

$$1/R = 1/R_{mn} \quad (2)$$

$$\log(1/R) = \log(1/R_{mn}) \quad (3)$$

where  $X_{mn}$  represents raw reflectance spectral matrix,

$$X_{mn} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$$

and  $m$  and  $n$  are the number of samples and wavelength variables, respectively. Transformation technology was implemented in Matlab R2010b (MathWorks, Natick, Massachusetts, USA).

*Development of classification models.*—The SVM was employed to establish classification models. The basic idea of SVM is to determine the optimal hyperplane that can separate data samples with the largest isolation edge. The kernel function is used to map spectral data onto the feature space. Spectral data can be fitted using the following equation:

$$f(x) = w \cdot g(x) + b \quad (4)$$

where  $(x_i, y_i)$  is spectral data of training set,  $i = 1, 2, \dots, n$ ,  $y_i \in \mathbb{R}$ ,  $n$  is the number of samples,  $w$  is weight vector,  $g(x)$  is nonlinear mapping function, and  $b$  is a threshold value. According to the structural risk minimization principle, the optimal separating hyperplane should satisfy

$$y_i[w \cdot g(x_i) + b] \geq 1 \quad (5)$$

The optimization problem is transferred into a classification problem by adding the relaxation factor (Yang 2021). It can be expressed as follows:

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i \\ s.t. y_i \{ [w \cdot g(x_i) + b] \} \geq 1 - \xi_i \end{cases} \quad (6)$$

Here,  $c$  is penalty function,  $\xi_i$  is relaxation factor ( $\xi_i > 0$ ). According to the Lagrange algorithm, Equation 6 can be transferred into the following equation:

$$\begin{cases} \min \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j a_i a_j K(x_i, x_j) - \sum_{i=1}^n a_i \\ s.t. \sum_{i=1}^n y_i a_i = 0, 0 \leq a_i \leq c \end{cases} \quad (7)$$

where  $K(x_i, x_j)$  represents kernel function. There are four types of kernel function—linear, radial basis function (RBF), sigmoid, and polynomial kernel. RBF was applied in this study because of the advantage of good prediction ability. The SVM was conducted using Matlab R2010b.

*The optimization of models.*—SVM in combination with kernel function can solve nonlinear identification problem. However, the optimization of parameters is essential for achieving good prediction accuracy (Tian et al. 2012). Three algorithms, including grid search (GS), genetic algorithm (GA), and particle swarm optimization (PSO), were compared to optimize the parameters of penalty factor ( $c$ ) and kernel function ( $g$ ) in the discrimination of wood origin and tree species based on various spectral data.

As for the GS method, the grid was established from  $2^{-8}$  to  $2^8$  with the moving step of  $2^{0.5}$ . The fitness extreme value can be updated according to the grid parameters until the best  $c$  and  $g$  are achieved. For the GA and PSO techniques, initial sizepop and maxgen are 20 and 100, respectively. Five-fold cross-validation was performed for the SVM model optimized by these three algorithms. The model with the largest accuracy rate of cross-validation was used to predict after five runs in this study.

*Overview of classification analysis.*—According to the spectral transformation technology, the raw reflectance spectra (R) of each tree species were subjected to the reciprocal (1/R) and logarithm reflectance [ $\log(1/R)$ ], SVM models were applied to classify origins and tree species based on three types of wood spectra. The GS, GA, and PSO methods were employed to optimize the parameters of the SVM model and RBF kernel function. The optimal classification model can be obtained based on the accuracy rate of cross-validation. The category labels of samples are shown in Table 2.

## Results and Discussion

### Wood spectral analysis

The average spectrum in various spectral forms based on spectral transformation technology is shown in Figure 2. Spectral curves of each tree sample show the similar pattern tendency for a specific spectral form. In addition, the feature wavelength variables, namely crests and troughs, are almost at the same position. Except for reflectance spectra (R), good spectral consistency is achieved according to spectral

Table 2.—Numbers and category labels of samples.

Regions	Tree species	No. of samples		Label	
		Training set	Test set	Origins	Tree species
Location 1	A	1–70	1–23	0	2
	B	71–131	24–43	0	3
	C	132–205	44–68	0	4
	D	206–266	69–88	0	5
Location 2	E	267–331	89–110	1	6
	F	332–398	111–132	1	7

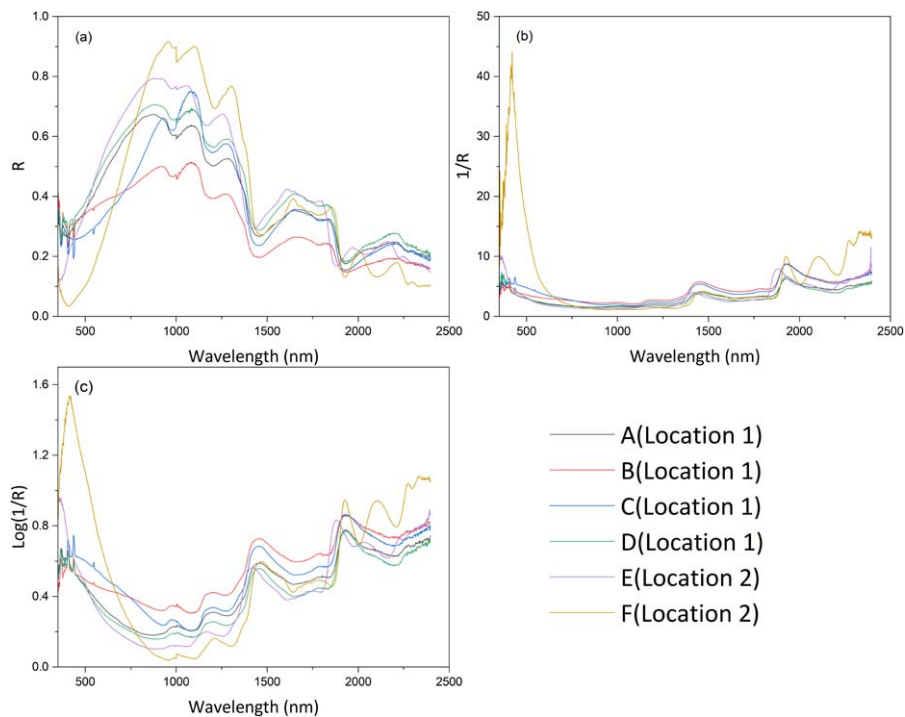


Figure 2.—The average spectrum for different spectral transformations: (a)  $R$ ; (b)  $1/R$ ; (c)  $\text{Log}(1/R)$ .

distance in the range of 750 to 1,750 nm. However, a large difference is obtained in the visible light spectrum for each tree. This was mainly because the pigment absorption is located in the visible region (Moberg et al. 2002), and there exist the color difference between heartwood and sapwood. In addition, regardless of various spectral forms, the spectra of *Larix gmelinii* (softwood) are different from other trees (hardwood). For example, the maximum peak value was generated by *L. gmelinii* wood with reflectance spectra. The reason may be due to the difference of physicochemical and anatomical properties between hardwood and softwood.

It can be seen in Figure 2a and c that six trees showed the obvious peaks for  $R$  and  $\text{Log}(1/R)$  spectra. As for  $\text{Log}(1/R)$  spectra, the absorption peak at 1,448 nm was associated with the first overtone of O-H stretching in lignin, and 1,940 nm was related to O-H of water (Bassett et al. 1963). In terms of  $R$  spectra, the feature peaks at 1,212 nm, 1,447 nm, 1,929 nm, and 2,200 nm were attributed to the C-H and O-H bands of cellulose, lignin, and water (Mitsui et al. 2008). However, it is difficult to classify the geographical origin and tree species through observing these spectra because of spectral overlap and bandwidth. Therefore, the SVM method was employed to build discrimination models.

### Geographical origin classification models

The parameters of penalty factor ( $c$ ) and kernel function ( $g$ ) are essential for the SVM model, which determine the number of support vectors and the behavior of kernel. In order to select the best hyperparameters of the SVM models, the GS, GA, and PSO algorithms were employed to optimize these parameters of geographical origin classification models based on various spectral data, respectively. The parameter optimization processes of  $\text{Log}(1/R)$  spectra are shown in Figure 3.

As can be seen in Figure 3, as for the GS method of  $\text{Log}(1/R)$  spectra, the accuracy increases gradually and obtains the maximum value with the increasing  $c$  value. For the GA method, the fitness value increases rapidly before 41 evolution iterations. In contrast, the fitness value tends to stabilize at approximately 100 evolution iterations for the PSO method. The parameter optimization results for each algorithm are shown in Table 3.

As seen in Table 3, in terms of penalty factor, different parameter values were achieved with various algorithms based on three spectral data. Regardless of these algorithms, a small  $g$  value was obtained for these three spectral forms. This demonstrated that the linear limit is one of the major impact factors in this study. The performance of SVM models based on the optimal hyperparameters for geographical origin classification is shown in Figure 4.

Performance of origin classification with these three algorithms (GS, GA, and PSO) for a specific spectral form was the same, and the prediction results of SVM models optimized by the GS algorithm with various spectral data were shown in Table 4. As can be seen, the SVM model has demonstrated good performance with an accuracy rate of

Table 3.—The parameter optimization results of origin classification according to support vector machine (SVM) models with different spectral forms.

Algorithms <sup>a</sup>	R		1/R		Log(1/R)	
	$c$	$g$	$c$	$g$	$c$	$g$
GS	1.414	0.008	8.000	0.004	0.500	0.008
GA	1.307	0.062	18.370	0.004	12.532	0.001
PSO	1.055	0.010	3.768	0.010	0.364	0.010

<sup>a</sup> Grid search (GS), genetic algorithm (GA), and particle swarm optimization (PSO).

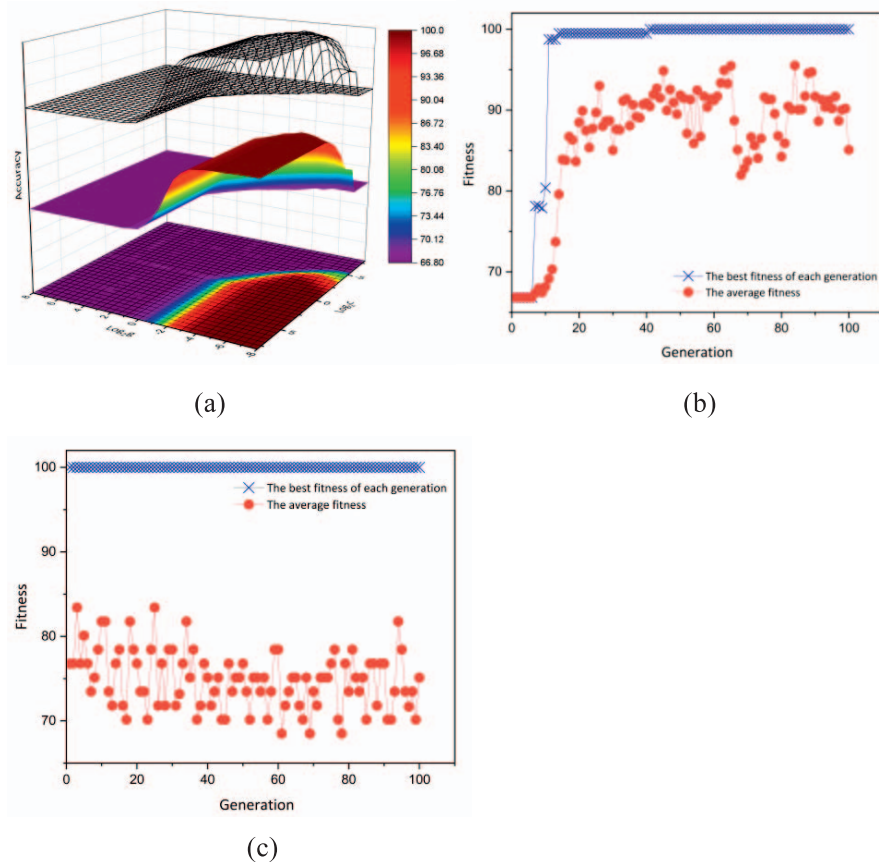


Figure 3.—The parameter optimization process of wood origin classification models with  $\text{Log}(1/R)$  spectra: (a) grid search [GS]; (b) genetic algorithm [GA]; (c) particle swarm optimization [PSO].

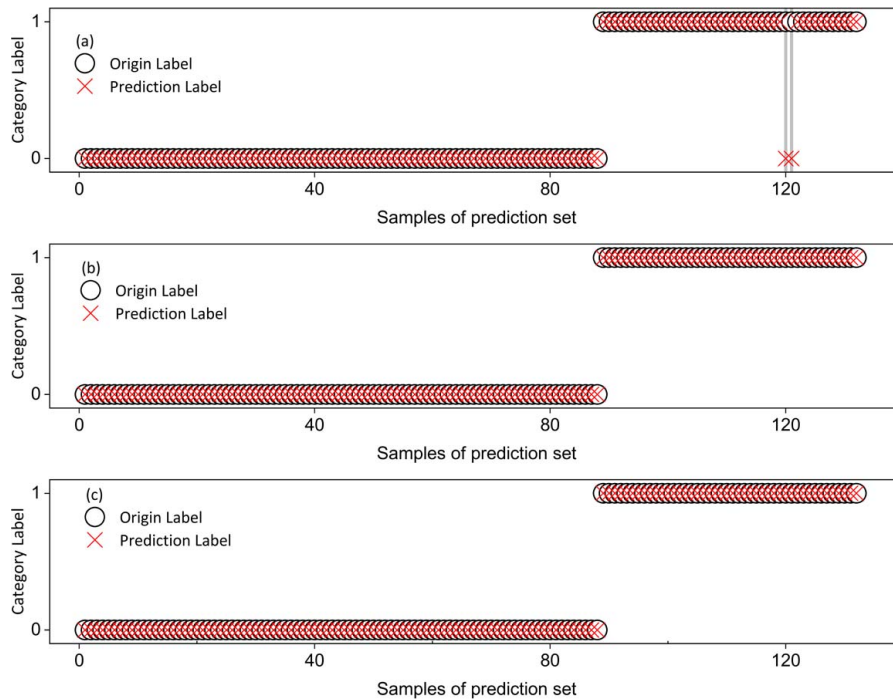


Figure 4.—The classification results for geographical origin with various spectral data for the grid search [GS] algorithm: (a)  $R$ ; (b)  $1/R$ ; (c)  $\text{Log}(1/R)$ .

Table 4.—Clustering results of tree species with various spectral forms.

Algorithms <sup>a</sup>	Location	Tree species	Training set						Accuracy rate (%)	Test set						Accuracy rate (%)
			A	B	C	D	E	F		A	B	C	D	E	F	
Spectral Form-R																
GS	1	A	68	1	0	1	0	0	98.49	22	0	0	1	0	0	93.18
		B	1	58	2	0	0	0		3	16	1	0	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	0	0	61	0	0	0	1	1	18	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
GA	1	A	68	1	0	1	0	0	98.49	22	0	0	1	0	0	91.67
		B	1	58	2	0	0	0		2	14	2	2	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	0	0	61	0	0	0	1	0	18	1	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
PSO	1	A	68	1	0	1	0	0	98.49	22	0	0	1	0	0	92.42
		B	1	58	2	0	0	0		1	15	1	3	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	0	0	61	0	0	0	1	1	18	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
Spectral Form-1/R																
GS	1	A	69	0	0	1	0	0	97.99	22	0	1	0	0	0	90.15
		B	1	57	0	3	0	0		1	14	2	3	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	1	1	59	0	0	1	3	0	16	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
GA	1	A	69	0	0	1	0	0	97.74	22	0	0	1	0	0	89.39
		B	1	56	0	4	0	0		1	13	2	4	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	1	1	59	0	0	1	3	0	16	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
PSO	1	A	70	0	0	0	0	0	98.49	23	0	0	0	0	0	88.64
		B	1	57	0	3	0	0		1	11	3	5	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	0	1	60	0	0	1	3	0	16	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
Spectral Form-Log(1/R)																
GS	1	A	69	1	0	0	0	0	98.99	23	0	0	0	0	0	93.18
		B	1	59	0	1	0	0		1	15	1	3	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	0	0	61	0	0	0	1	1	18	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
GA	1	A	68	1	1	0	0	0	97.74	21	0	1	1	0	0	92.42
		B	2	57	0	2	0	0		1	15	2	2	0	0	
		C	1	0	73	0	0	0		2	0	23	0	0	0	
	2	D	0	1	1	59	0	0	0	1	0	19	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		
PSO	1	A	69	1	0	0	0	0	99.25	21	0	0	2	0	0	92.42
		B	0	60	0	1	0	0		1	15	1	3	0	0	
		C	0	0	73	1	0	0		2	0	23	0	0	0	
	2	D	0	0	0	61	0	0	0	0	1	19	0	0		
		E	0	0	0	0	65	0	0	0	0	0	22	0		
		F	0	0	0	0	0	67	0	0	0	0	0	22		

<sup>a</sup> Grid search (GS), genetic algorithm (GA), and particle swarm optimization (PSO).

100 percent for calibration set (data not shown) and prediction set, except for R spectra. As for the performance of R spectra, two samples from Heilongjiang Province (location 2) of the prediction set were misjudged by the samples of Jilin Province (location 1), and the same accuracy rate of 98.48 percent was generated for various algorithms. This indicated that the performance of Log(1/R) and 1/R spectra were better than that of R spectra in the discrimination of geographical origin.

### Tree species discrimination models

In order to better compare the ability of spectral transformation and these algorithms for tree species discrimination, and to take the impact of geographical origin on tree species classification into account, the same tree species was collected from two locations in this study. The clustering results of tree species with various spectral forms and algorithms are shown in Table 4.

As the table shows, regarding R spectra, the calibration set achieved the same classification results for three algorithms with an accuracy rate of 98.49 percent, and the misjudgment occurred only in three types of tree species (A, B, and C). However, different accuracies were obtained for the prediction set. The performance of the GS method was better than that of the GA and PSO techniques. This may be due to the fact that the number of misjudgments of tree species B was smaller than other two methods. In addition, the distributions of tree species B are different between the three algorithms. For example, 2 of 20 trees in the prediction set of tree species B with the GA method were misclassified as tree species A, C, and D, respectively. However, 1 of 20 trees were misjudged as A and C, respectively, and 3 trees were misclassified as D for the PSO technique. Additionally, for the GA mean, although the distributions of misjudgment of tree species D were different from other two methods, the number of misjudgments was the same as the others. The best discrimination results were demonstrated by the GS method for R spectra with the prediction accuracy rate of 93.18 percent.

In terms of 1/R spectra, different results were generated among three algorithms. Additionally, compared with R spectra, the accuracy rate was decreased for calibration and prediction set except for the calibration set of the PSO method. The best accuracy rate was achieved with the GS technique for 1/R spectra. As for Log(1/R) spectra, the performance of the calibration set was improved except for the model optimized by the GA method. However, the prediction accuracy of GA was the same as the PSO technique because of the same number of misjudgments. The best results of the prediction set were achieved by the GS method with the accuracy rate of 93.18 percent. Additionally, regardless of various algorithms and spectral forms, an accuracy rate of 100 percent was obtained for the tree species from the second location. This indicated that the performance of discrimination of tree species from the second location was better than that of trees from the first location for these algorithms. Compared with other studies (Korpela et al. 2011, Lang et al. 2017, Zulfa et al. 2020), although an accuracy rate of 100 percent was achieved for the classification of tree species (The second location), there existed a large number of misjudgments for A, B, and C species from the first location. This may be due to the single level of tree species spectra or the spectral noise. Therefore,

spectral data at different scales and some spectral denoising techniques should be analyzed in future scopes.

### Conclusions

The combination of Vis-NIR and chemometric techniques have been shown to be suitable methods for the determination of geographical origin and tree species. For origin classification, except for reflectance spectra, the prediction accuracy of 100 percent was achieved with three algorithms. However, reflectance spectra together with the GS technique demonstrated the best performance of tree species discrimination. Furthermore, these spectral forms and algorithms displayed good performance of tree species identification for the second location according to the discrimination results of *Populus davidiana* wood. Therefore, an appropriate combination of various algorithms can be an alternative for geographical origin and tree species classification. Although the best results were achieved by Vis-NIR spectra and chemometric methods, the limitation of this study is that the models can be used for the determination of these five tree species (*Populus davidiana*, *Tilia tuan* Szyszyl., *Ulmus pumila* L., *Acer mono* Maxim., and *Larix gmelinii*). Therefore, the spectral data of other tree species, especially for the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) -listed tree species, should be added in future studies.

### Acknowledgments

This research was funded by the Natural Science Foundation of Inner Mongolia Autonomous Region, grant number 2021BS03019; the Outstanding Doctoral Introduction Fund of School, grant number NDYB2020-11; and the Outstanding Doctoral Introduction Fund of Inner Mongolia Autonomous Region, grant number DC2100001417.

### Literature Cited

- Bassett, K. H., C. Y. Liang, and R. H. Marchessault. 1963. The infrared spectrum of crystalline polysaccharides. IX. The near infrared spectrum of cellulose. *J. Polymer Sci. Part A: Gen. Pap.* 1:1687–1692.
- Cai, C., L. Xu, Q. Han, H. Wu, J. Nie, H. Fu, and R. Yu. 2010. Combining the least correlation design, wavelet packet transform and correlation coefficient test to reduce the size of calibration set for NIR quantitative analysis in multi-component systems. *Talanta* 81:799–804.
- Cordier, S., F. Robichaud, P. Blanchet, and B. Amor. 2021. Regional environmental life cycle consequences of material substitutions: The case of increasing wood structures for non-residential buildings. *J. Clean Prod.* 328:129671.
- Davari, M., S. Fahmideh, and M. R. Mosaddeghi. 2022. Rapid assessment of soil water repellency indices using Vis-NIR spectroscopy and pedo-transfer functions. *Geoderma* 406:115486. <https://doi.org/10.1016/j.geoderma.2021.115486>
- Galvão, R. K. H., H. A. D. Filho, M. N. Martins, M. C. U. Araújo, and C. Pasquini. 2007. Sub-optimal wavelet denoising of coaveraged spectra employing statistics from individual scans. *Anal. Chim. Acta.* 581:159–167.
- Hamilton, M. and B. Potts. 2007. Review of *Eucalyptus nitens* genetic parameters. *N. Z. J. Forestry Sci.* 38:102–119.
- Huang, H. L., X. J. Mai, and H. D. Huang. 2018. Resources and development research of Hainan Yellow Pear in China. *Chinese J. Agric. Resour. Reg. Plan.* 39(9):7.
- Jiang, H., W. Xu, and Q. Chen. 2019. Comparison of algorithms for wavelength variables selection from near-infrared (NIR) spectra for quantitative monitoring of yeast (*Saccharomyces cerevisiae*) cultivations. *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* 214:366–371.
- Jiang, Y. H. 2014. Simulation on stem taper of Dahurian Larch based on

- crown characteristics. Master's thesis. Northeast Forestry University, Harbin, Heilongjiang, China.
- Kapoor, R., A. Malvandi, H. Feng, and M. Kamruzzaman. 2022. Real-time moisture monitoring of edible coated apple chips during hot air drying using miniature NIR spectroscopy and chemometrics. *LWT* 154:112602. <https://doi.org/10.1016/j.lwt.2021.112602>
- Korpela, I., V. Heikkinen, E. Honkavaara, F. Rohrbach, and T. Tokola. 2011. Variation and directional anisotropy of reflectance at the crown scale—Implications for tree species classification in digital aerial images. *Remote Sens. Environ.* 115:2062–2074.
- Lang, C., D. R. A. Almeida, and F. R. C. Costa. 2017. Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). *Forest Ecol. Manag.* 406:219–227.
- Liu, Z. B., Y. L. Sun, S. H. Li, Q. Zhang, H. W. Zhu, and H. W. Xue. 2012. Preliminary study on construction technology of artificial larch and *Fraxinus mandshurica* mixed forest in Changbai Mountain Area. *Sci. Technol. Inf.* 7:149.
- Lu, W. Z. 2007. Modern Near Infrared Spectroscopy Analytical Technology. 2nd ed. China Petrochemical Press, Beijing. pp. 59–61.
- Mitsui, K., T. Inagaki, and S. Tsuchikawa. 2008. Monitoring of hydroxyl groups in wood during heat treatment using NIR spectroscopy. *Biomacromolecules* 9:286–288.
- Moberg, L., B. Karlberg, K. Srensen, and T. Källqvist. 2002. Assessment of phytoplankton class abundance using absorption spectra and chemometrics. *Talanta* 56:153–160.
- Paltseva, A. A., M. Deeb, E. Di Iorio, L. Circelli, Z. Cheng, and C. Colombo. 2022. Prediction of bioaccessible lead in urban and suburban soils with Vis-NIR diffuse reflectance spectroscopy. *Sci. Total Environ.* 809:151107. <https://doi.org/10.1016/j.scitotenv.2021.151107>
- Phillips, S. 2009. Bio-composite aterial applications to musical instruments. Master's thesis. McGill University, Montreal, Quebec, Canada.
- Scotter, C. 2005. Infrared spectroscopy|near-infrared. In: Encyclopedia of Analytical Science. 2nd edition. P. Worsfold, A. Townshend, C. Poole (Eds.). Elsevier, Boston and Amsterdam. pp. 415–426.
- Tan, M. 2012. Natural forest protection project social impact assessment in Sanchazi forestry bureau. Master's thesis. Northeast Forestry University, Harbin, Heilongjiang, China.
- Tian, F., Y. Jia, X. Shan, J. Feng, and C. Kadri. 2012. Classification of electronic nose data on wound infection detection using support vector machine combined GA. *JCIS* 8:3349–3357.
- van Kollenburg, G. H., H. van Manen, N. Admiraal, J. Gerretzen, and J. J. Jansen. 2021. Low-cost handheld NIR spectroscopy for identification of organic solvents and low-level quantification of water contamination. *Talanta* 223:121865. <https://doi.org/10.1016/j.talanta.2020.121865>
- Vega, M., P. Harrison, M. Hamilton, R. Musk, P. Adams, and B. Potts. 2021. Modelling wood property variation among Tasmanian *Eucalyptus nitens* plantations. *Forest Ecol. Manag.* 491:1–11. <https://doi.org/10.1016/j.foreco.2021.119203>
- Villa, P. M., S. V. Martins, É. S. S. Nolasco De Oliveira Neto, Diniz, A. V. Neri, H. Pinto-Junior, J. A. Nunes, M. L. Bueno, and A. Ali. 2021. Taxonomic and functional beta diversity of woody communities along Amazon forest succession: The relative importance of stand age, soil properties and spatial factor. *Forest Ecol. Manag.* 482:1–11. <https://doi.org/10.1016/j.foreco.2020.118885>
- Wang, Z. M., J. Zhou, L. L. Wang, Y. T. Li, D. X. Liu, and Q. S. Yang. 2020. Comprehensive evaluation of 15 elm trees and varieties introduced in Yellow River Delta area. *Forestry Ecolog. Sci.* 35:419423.
- Wu, H. T. and Y. W. Zhang. 2022. Comparative study on processing characteristics of Linden honey polysaccharide and multifloral honey polysaccharide. *Sci. Technol. Food Ind.* 43:323–328.
- Yang, Y. G. 2021. Research on rolling bearing fault diagnosis method based on neural network and support vector machine. Master's thesis. Lanzhou Jiaotong University, Lanzhou, Gansu, China.
- Zhuang, X., L. Wang, Q. Chen, X. Wu, and J. Fang. 2017. Identification of green tea origins by near-infrared (NIR) spectroscopy and different regression tools. *Sci. China Technol. Sci.* 60:84–90.
- Zulfa, A. W., K. Norizah, O. Hamdan, S. Zulkifly, I. Faridah-Hanum, and P. P. Rhyma. 2020. Discriminating trees species from the relationship between spectral reflectance and chlorophyll contents of mangrove forest in Malaysia. *Ecol. Indicat.* 111:106024. <https://doi.org/10.1016/j.ecolind.2019.106024>