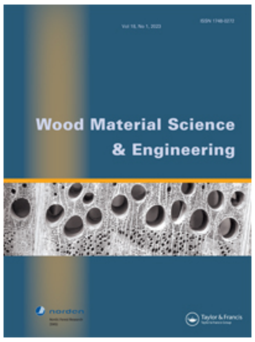


Wood Material Science & Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/swoo20

Efficiency of preprocessing methods for discrimination of anatomically similar pine species by NIR spectroscopy

F. Digdem Tuncer, Dilek Dogu & Esra Akdeniz

To cite this article: F. Digdem Tuncer, Dilek Dogu & Esra Akdeniz (2023) Efficiency of preprocessing methods for discrimination of anatomically similar pine species by NIR spectroscopy, Wood Material Science & Engineering, 18:1, 212-221, DOI: 10.1080/17480272.2021.2012821

To link to this article: https://doi.org/10.1080/17480272.2021.2012821

WOOD MATERIAL SCIENCE & ENGINEERING
2023, VOL. 18, NO. 1, 212–221
https://doi.org/10.1080/17480272.2021.2012821

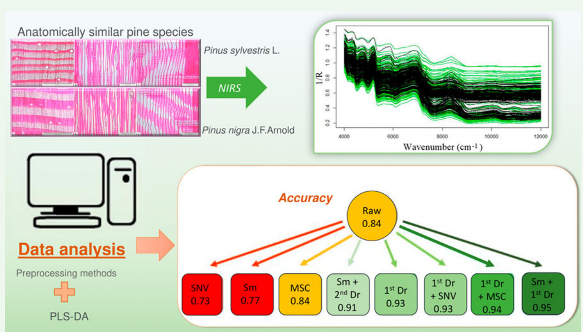
ORIGINAL ARTICLE

Efficiency of preprocessing methods for discrimination of anatomically similar pine species by NIR spectroscopy

F. Digdem Tuncer [a], Dilek Dogu [a] and Esra Akdeniz [b]

[a]Department of Forest Biology and Wood Protection Technology, Faculty of Forestry, Istanbul University-Cerrahpasa, Istanbul, Turkey; [b]Department of Medical Education, School of Medicine, Marmara University, Istanbul, Turkey

ABSTRACT
Identification of wood species with fast, reliable and non-destructive methods is highly important for forestry and wood-related industries. Near-infrared spectra of anatomically similar pine species (*Pinus sylvestris* L. and *Pinus nigra* J.F. Arnold) were taken and analysed by partial least squared discriminant analysis (PLS-DA) for comparing the efficiency of preprocessing methods. Raw data were subjected to multiple scatter correction (MSC), standard normal variate (SNV), Savitzky–Golay for derivatives (1st and 2nd Dr) and smoothing (Sm) and combination of these preprocessing methods (1st Dr, 1st Dr + SNV, 1st Dr + MSC, Sm + 1st Dr and Sm + 2nd Dr). The success of the models was determined by the accuracies of test sets that did not participate in the calibration phase. In this study, it was determined that not all the preprocessing methods improve the model performance. Smoothing with 1st derivatives (Sm + 1st Dr) enhanced 14.3% improvement and have the best performance (95%) for classification of pine species. For understanding modelled relationship, mean spectra and selectivity ratio were used. It was found that discrimination was held by the differences at their absorption, and the most important variables for wood classification were noted around 4000–7000 cm^{-1}.



ARTICLE HISTORY
Received 6 August 2021
Revised 27 November 2021
Accepted 28 November 2021

KEYWORDS
Near-infrared spectroscopy; wood identification; discrimination; preprocessing methods; classification; PLS-DA

Introduction

Identification of wood species is generally carried out by wood anatomy through microscopic examinations. This identification method is the most established and reliable one (Wheeler and Baas 1998), but beyond that, it also requires expertise, needs relatively long preparation and inspection time and is quite labour-intensive. The development of new technologies in the world and our expectation of fast access to information are becoming a necessity for forest industries, as in every stage of life. Consequently, it leads up to a search for substitution/different diagnostic methods (Dormontt *et al.* 2015). Some of the developed methods for the identification of the species with high success rates are as follows: application of multivariate data analysis to wood anatomy (Clark 2003, Hellberg and Carcaillet 2003, Evans *et al.* 2008, Gasson and MacLachlan 2010, Turhan and Serdar 2013, Giménez *et al.* 2014, Marques *et al.* 2015, Esteban *et al.* 2017); machine vision, which allows wood anatomy experts to teach the necessary information to computer-based camera systems to distinguish tree species with similar characteristics (Hermanson and Wiedenhoeft 2011, Ravindran *et al.* 2018, de Andrade *et al.* 2020); DNA barcoding, which allows us to work on very small samples but can take up to several days (Jiao *et al.* 2014); and application of mass spectroscopy based on extractives content of wood (Espinoza *et al.* 2014, Deklerck *et al.* 2017, Evans *et al.* 2017). In addition to these studies, another prominent solution proposal is the use of near-infrared spectroscopy (NIRS) with chemometric methods, which evaluates the data obtained by analysing the chemical structure of the wood with statistical approaches, multivariate analysis (Pasquini 2003, Brereton 2007, Burns and Ciurczak 2008, Tsuchikawa and Kobori 2015). The common point of all these methods is that each of them is still in the experimental stage. In order for these methods to be counted among the traditional methods, the usability of these systems should be tested for different tree species.

CONTACT F. Digdem Tuncer fdigdemt@iuc.edu.tr

NIRS is an analytical method used to determine the chemical components of the material, characterized by quantitative and qualitative analysis. This method is based on measuring the vibrational energy reacted by the molecular bonds of organic substances. With the use of vibrational energy for analysis, NIRS appears as a non-destructive, low-costly, clean, and easily reproducible method (Tsuchikawa *et al.* 2003, Burns and Ciurczak 2008, Workman and Weyer 2012, Cao 2013).

Even NIRS is already applied in various industries, the use of NIRS in the forest products industry is in the development phase and has not yet been included in the standard methods. On the other hand, researches show that NIRS makes it possible to obtain highly accurate predictive information for many different purposes related to the forest products industry (Via *et al.* 2014, Wang *et al.* 2021), such as the content of lignin, cellulose, hemicellulose, extractives and the crystallinity of cellulose (Jiang *et al.* 2007), microfibril angle (Schimleck *et al.* 2005, Via *et al.* 2007), morphological characteristics of tracheid (Schimleck *et al.* 2004, Jones *et al.* 2005), mechanical properties (Hoffmeyer and Pedersen 1995, Kelley *et al.* 2004), density (Li *et al.* 2019), moisture content (Hans *et al.* 2015), shrinkage behaviour (Taylor *et al.* 2008), separation of sapwood-heartwood (Sandberg and Sterley 2009), determination of reaction wood (Chen *et al.* 2007), type and degree of degradation (Gierlinger *et al.* 2003, So *et al.* 2004), effects of wood modification (Bächle *et al.* 2012), content of wood pulp (Dahlman 2012) and discrimination of tropical wood species (Flaete *et al.* 2006, Tounis 2009, Braga *et al.* 2011, Espinoza *et al.* 2012, Shou *et al.* 2014, Lang *et al.* 2015) based on the chemical structure of wood.

In near-infrared (NIR) region, due to the changes in the path of light (scatter effects) and the complex spectra formed by the overlapping and combination of signals, problems could occur. Herewith, it becomes difficult to establish the direct connections of a particular feature with a particular component. For this reason, it is critical to use preprocessing methods to obtain the desired information from the spectrum (Burns and Ciurczak 2008, Rinnan *et al.* 2009, Workman and Weyer 2012, Wiesner *et al.* 2014). The main reason for using preprocessing methods is to reduce the noise in the data set and improve the spectral properties of interest. Choosing the most robust preprocessing method is very important for establishing a reliable model and for a successful modelling phase (Gholizadeh *et al.* 2015, Olivieri 2018). But it has two main problems; lack of a standard procedure for selection of the suitable preprocessing method (s) and the noise originating from the device can be seen as a feature (correlation of noise).

In many commercial cases, wood discrimination at the genus level is sufficient (Dormontt *et al.* 2015), but the importance of discrimination at the species level becomes apparent when the situations like where information about cultural and historical structures is needed (Horikawa *et al.* 2015, Hwang *et al.* 2016) or where some endangered species' trade is restricted while there is no commercial restriction for other species of similar appearance within the same genus (Gasson and MacLachlan 2010, Shou *et al.* 2014, Snel *et al.* 2018). Although microscopic methods are generally used in such cases where tree species identification is needed, there are situations where it is difficult to distinguish some tree species from each other microscopically, sometimes even impossible (Hather 2000, Schoch *et al.* 2004). In this study, identification of anatomically similar wood species (Scotch pine (*Pinus sylvestris* L.) and black pine (*Pinus nigra* J.F. Arnold)) that cannot be distinguished by traditional methods was investigated by NIR spectroscopy. In this way, an area where traditional methods are insufficient has been tried to be illuminated. Furthermore, frequently used preprocessing methods (Tuncer 2020) employed for wood discrimination via NIRS were compared. For this purpose, raw data were preprocessed using multiple scatter correction (MSC), standard normal variate (SNV), Savitzky–Golay for derivatives (1st and 2nd) and smoothing (Sm) and combination of these methods. For comparing the efficiency of different preprocessed data, partial least square discriminant analysis (PLS-DA), a multivariate data analysis, was used.

Materials and methods

For this work, 25 trees for each species of Scotch pine (*Pinus sylvestris* L.) and black pine (*Pinus nigra* J.F. Arnold) were taken from Ayancik (Sinop) and Sariyer (Istanbul) regions in Turkey (Table A1, Appendix). Sample collection was held by cutting wood discs and taking increment bores at breast height (1.30 m). For each disc, a woodblock was cut from bark to pith and all samples were air-dried at 20 ± 2°C with 65% + 5 relative humidity. For homogenization of the surface roughness, cross section of each sample was sanded with grit 80 and 180. A total of 376 spectra from 50 individual trees were collected from bark to pith (Table 1).

NIR spectra were collected by Antaris FT-NIR Analyzer (Thermo Nicolet Scientific, USA) equipped with an integrating sphere (operating in reflectance mode) at the spectral region from 4000 to 12000 cm^{-1} with 4 cm^{-1} resolution, and 64 scans averaged per scan. Depending on the changes in chemical structure through perpendicular to the longitudinal axis of the tree (radial direction) (Zobel and van Buijtenen 1989, Schweingruber 2007), more than one measurement area (5–10 per sample) in the radial direction were determined on the samples.

Data analysis and spectral preprocessing were carried out using open-access R software (R Core Team 2018) with various packages (Signal Developers 2013, Wehrens 2011, Kucheryavskiy 2019, Mevik *et al.* 2019). Partial least squares discriminant analysis (PLS-DA) was carried out according to

Table 1. Diameter at breast height (DBH), ages, number (No.) of trees and spectra.

Wood Species	DBH (cm) (min–max)	Age at DBH (min–max)	No. of trees (wood discs/ increment bores)	No. of spectra (train/test)
P. sylvestris	31–62	57–142	25 (10/15)	195 (143/52)
P. nigra	32–74	49–162	25 (10/15)	181 (128/53)
		Total	50	376

Straightforward Implementation of a statistically inspired modification of the partial least square (PLS) method SIMPLS algorithm (de Jong 1993, Kucheryavskiy 2019).

Several spectral preprocessing methods and their combinations were used to eliminate the scatter effects and removal of noises. Multiplicative scatter correction (MSC) (Geladi and Kowalski 1986, Naes *et al.* 2004, Mevik *et al.* 2019) and standard normal variate (SNV) (Barnes *et al.* 1989, Kucheryavskiy 2019), for smoothing (Sm) (15 point filter, first-order polynomial) (Kucheryavskiy 2019) and derivative (Dr) Savitzky–Golay (SG) (Savitzky and Golay 1964, Naes *et al.* 2004, Signal Developers 2013, Kucheryavskiy 2019, Olivieri 2018) algorithm and combination of these preprocessing methods (1.Dr + SNV, 1. Dr + MSC, Sm + 1Dr (3-point filter, first-order polynomial) and Sm + 2Dr (5-point filter, second-order polynomial)) were used in this work; 70% of the data sets were used for the calibration of the models, the remaining 30% was reserved for the control of the models with independent samples (validation). The samples were assigned to train and test groups at the tree level. For both species, the spectra obtained from 18 trees were divided into the training group, while the spectra obtained from 7 trees were divided into the test group. The performances of the models were evaluated based on accuracy (ACC), sensitivity (SEN) and specificity (SPE) via confusion matrix (Table 2) (Ciaburro and Venkateswaran 2017).

Table 2 Confusion matrix.

Confusion matrix		Predicted values: *True (1)*	Predicted values: *False (0)*
Actual Values	*True (1)*	True positive (TP)	False negative (FN)
	False (0)	False positive (FP)	True negative (TN)

$$ACC(\%) = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$SEN(\%) = TP/(TP + FN) \quad (2)$$

$$SPE(\%) = TN/(TN + FP) \quad (3)$$

For variable selection and interpretation, the relationship between the latent variables used in PLS-DA and the original variables was determined by selectivity ratio (SR) (Rajalahti *et al.* 2009).

$$SR_i = \nu_{expl,i}/\nu_{res,i} \quad i = 1, 2, 3\ldots \quad (4)$$

Results and discussion

As can be seen in Figure 1, NIR spectra of two pine species have quite similar spectral curves and overlaps, like their anatomical characteristics. Based on this visual similarity, it is hard to make any discrimination by individual samples. This caused a necessity to use preprocessing methods for revealing the hidden relationship.

In order to find out the best preprocessing method for modelling, different methods were tried alone or combined with others (Figure 2). Since visual differentiation could not achieve by just preprocessing methods, the use of chemometric analysis has become a requirement. Based on this requirement and for comparison of the effects of preprocessing methods to classifying wood species, PLS-DA models were used (Table 3).

According to the correct classification rate (accuracy) without any preprocessing methods (raw data), classification of pinewood species was acquired (84.3%). It was determined

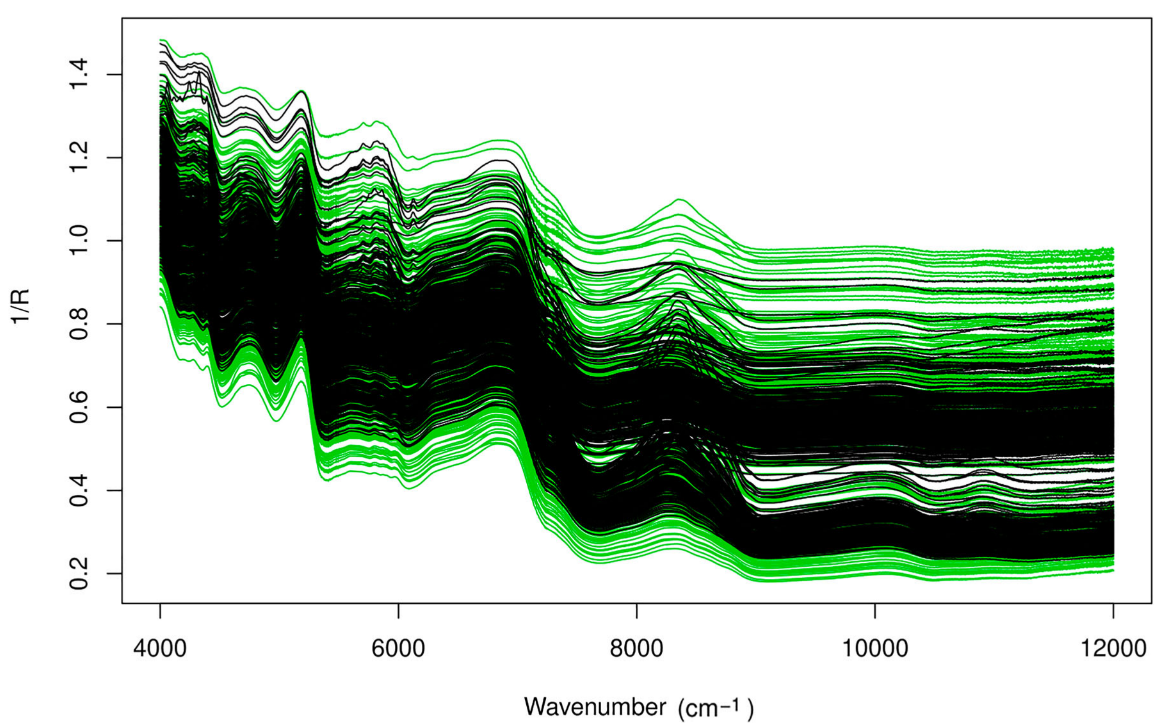


Figure 1. NIR spectra of pine species: Scotch pine (green), black pine (black).

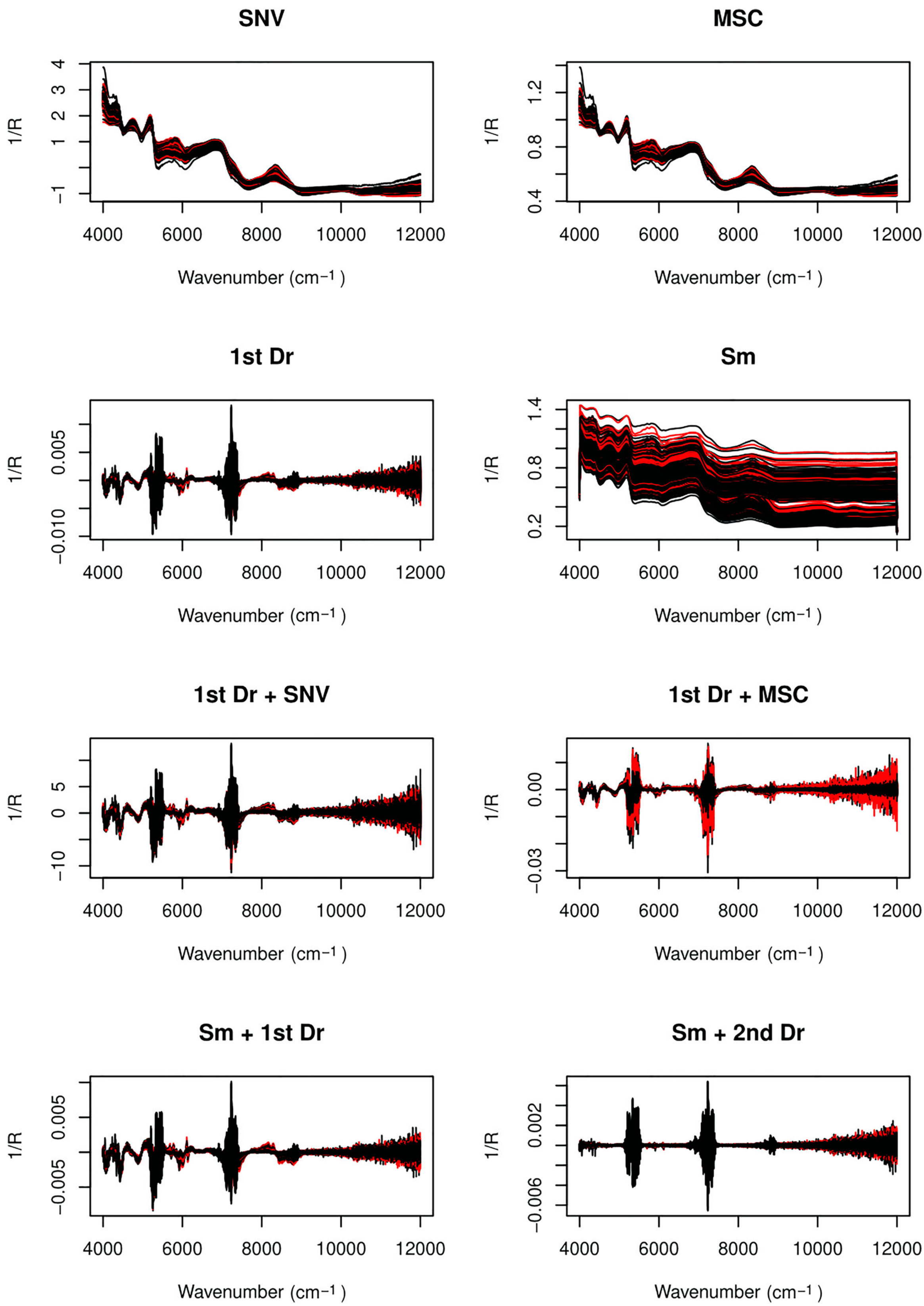


Figure 2. NIR spectra of pine species; from left to right SNV, MSC, 1Dr, Sm, 1.Dr + SNV, 1.Dr + MSC, Sm + 1.Dr, Sm + 2.Dr.

that not all the preprocessing methods improve the model performance. Smoothing with 1st derivatives (Sm+1Dr) enhanced 14.3% improvement and has the best performance (95%) for classification of pine species by means of PLS-DA. On the other hand, standard normal variate (SNV) has a negative effect and diminished the model performance by 13.10% (Table 3). These result consistent with the recommendation of not to use multiple preprocessing methods, as it may result in

Table 3. Performance of PLS-DA models with respect to preprocessing methods.

	Calibration			Validation			
Preprocessing methods	*ACC*	*SPE*	*SEN*	*ACC*	*SPE*	*SEN*	Improvement of accuracy (%) (test)
Raw	0.974	1	0.945	0.843	0.843	0.833	-
SNV	0.978	1	0.953	0.733	0.740	0.727	−13.10
MSC	0.989	1	0.977	0.838	0.818	0.860	0
1. Dr	1	1	1	0.933	0.881	1	10.71
1. Dr + SNV	1	1	1	0.933	0.895	0.979	10.71
1. Dr + MSC	1	1	1	0.943	0.897	1	11.90
Sm	0.875	0.923	0.820	0.771	0.792	0.754	−8.33
Sm + 1. Dr	1	1	1	0.952	0.912	1	14.29
Sm + 2. Dr	1	1	1	0.914	0.852	1	8.33

Note: ACC: Accuracy, SPE: Specificity, SEN: Sensitivity, SNV: Standard normal variate, MSC: multiplicative scatter correction, Sm: smoothing, Dr: derivative.

loss of meaningful data (Rinnan *et al.* 2009). Therefore, during the selection of preprocessing methods, it is better to carry out classic empirical research until a robust multivariate model is found (Olivieri 2018).

Considering all the studies conducted in the last 15 years about wood classification via NIRS, it has been determined that the accuracy rates for PLS-DA vary between 60% and 100% (Flaete *et al.* 2006, Tounis 2009, Horikawa *et al.* 2015, Yang *et al.* 2015, Bergo *et al.* 2016, Hwang *et al.* 2016, Lazarescu *et al.* 2017, Snel *et al.* 2018, Yong *et al.* 2019, Leandro *et al.* 2019, Pace *et al.* 2019, Zhou *et al.* 2020). Although this change in the correct prediction rates depends on the tree species and the applied preprocessing methods when comparing similar researches, it can be concluded that classification of Scotch pine and black pine wood species by the combination of NIRS and PLS-DA (95%) is quite successful.

Since the selection of the original variable is essential for understanding the modelled relationship (Xiaobo *et al.* 2010, Kvalheim 2020, Mehmood *et al.* 2020), mean spectra and latent variables of PLS-DA were examined in the current study. Visual inspection of the spectrum is the most commonly used method for classifying an unknown material in optical spectroscopy. This examination is carried out by looking at the areas where the absorbance peaks are located (shape) and the strength of absorbance peaks. If the peaks are seen at the same wavelengths in each of the compared materials, the absorbance differences can serve as a distinction in the classification of these materials (Burns and Ciurczak 2008). The mean spectra of pine species have some similarity in shape, but several minor differences in their absorption could be seen visually (Figure 3). It is thought that the differences between the species in the same genus arise from the quantitative differences in their chemical structures, not the qualitative ones (Fengel and Wegener 1983). Thus, their spectra appeared very similar to each other, and the difference is due to the strength of the absorption peaks. Those areas where the absorption differences were appearing at mean spectra were noted around 4167–4400 cm^{-1}, 4500–5267 cm^{-1}, 5500–6000 cm^{-1}, 6267–7000 cm^{-1}, 4733 cm^{-1}, 5167 cm^{-1} and 6833 cm^{-1}.

For understanding the modelled relationship, latent variables used in PLS-DA were determined by the selectivity

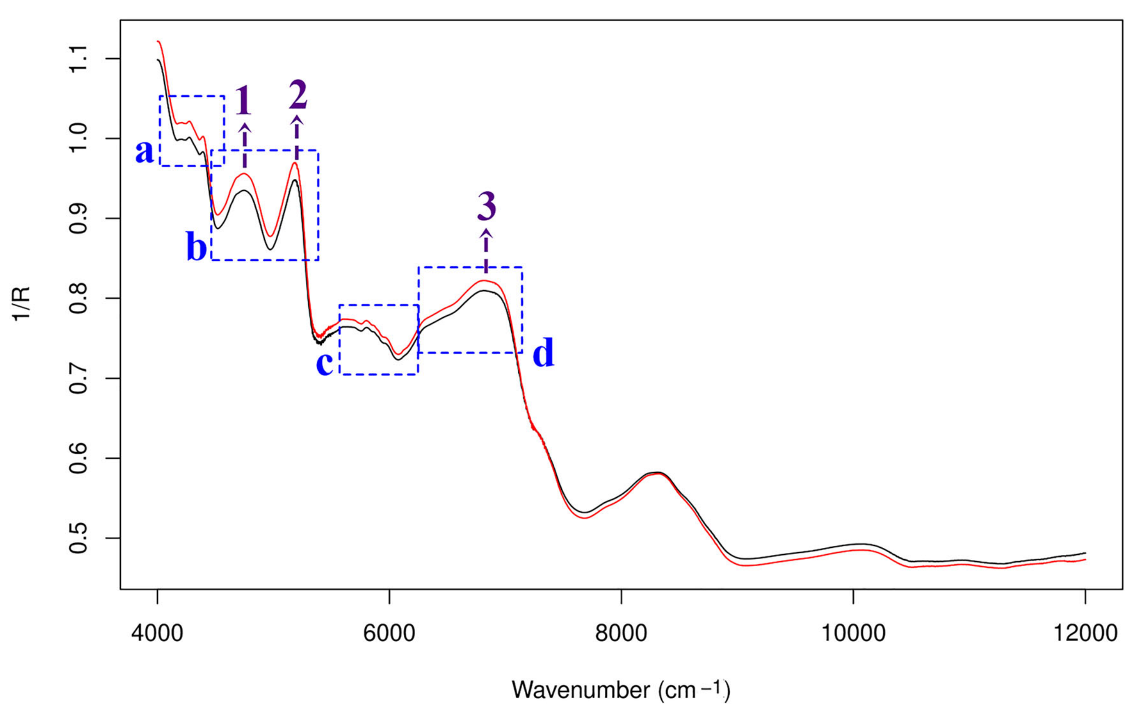


Figure 3. Mean spectra of Scotch pine (red) and black pine (black). a: 4167–4400 cm^{-1}, b: 4500 cm^{-1}, c: 5500–6000 cm^{-1}, d: 6267–7000 cm^{-1}, 1:4733 cm^{-1}, 2:5167 cm^{-1}, 3: 6833 cm^{-1}.

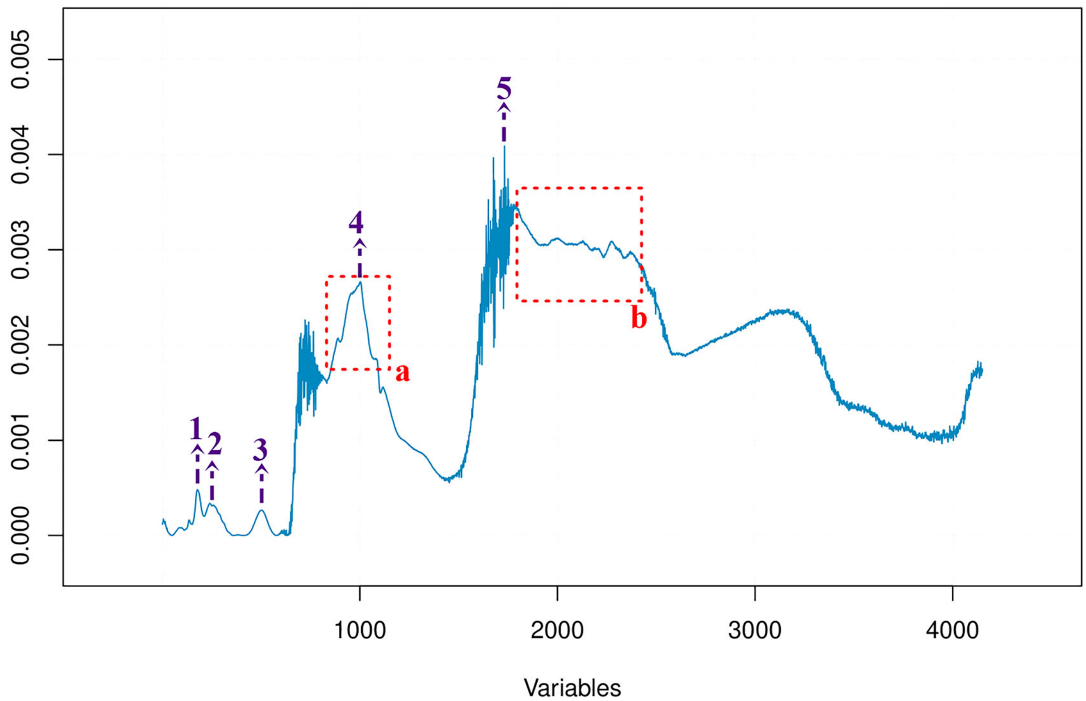


Figure 4. Selectivity ratio of raw data. a: 5680–6100 cm^{-1}, b:7444–8788 cm^{-1}, 1: 4336 cm^{-1}, 2: 4504 cm^{-1}, 3: 4966 cm^{-1}, 4: 5932 cm^{-1}, 5: 7360 cm^{-1}.

ratio. First 20 original variables according to selectivity ratios (SR) of the variables used in PLS-DA were 7444–8788 cm^{-1}, 7360 cm^{-1}, 5932 cm^{-1}, 5680–6100 cm^{-1}, 4336 cm^{-1}, 4504 cm^{-1} and 4966 cm^{-1} (Figure 4).

When comparing the absorbance differences in mean spectra with selectivity ratio, it was seen that they were quite compatible with each other. These findings were consistent with the near-infrared studies, which carried out on the classification of different wood species held by PLS-DA (Table 4) (Gierlinger *et al.* 2004, Braga *et al.* 2011, Pastore *et al.* 2011, Espinoza *et al.* 2012, Shou *et al.* 2014, Lang *et al.* 2015, Snel *et al.* 2018). Although the assignment of bands to which chemicals or bonds they relate to has been calculated since the inception of infrared spectroscopy, it is a difficult step as the number and intensities of bonds vary widely even for simple molecules. In complex materials such as wood, these calculations were considered much more complex and difficult to implement. These difficulties arise from the complex molecular structures of wood components, as well as the fact that even a simple functional group can change the molecular environment considerably. Thereby bands and vibrations in these areas were discussed instead of the chemical components to which the bands are related in the near-infrared spectrum. When dividing the near-infrared spectrum into regions, 1st and 2nd overtones of OH, NH stretching vibration, 1st CH combination, 2nd and 3rd CH, $_{Ar}CH$, NH stretching vibrations were appearing in 12000–7000 cm^{-1} region which is affected by physical changes such as particle size and colour, it is preferred in working subjects such as thermally treated woods, density, etc. Hereby, this region is generally used for the establishment of calibration models for quantitative purposes (Tsuchikawa *et al.* 2003, Schwanninger *et al.* 2004, 2011, Bächle *et al.* 2012, Popescu *et al.* 2018). Below the 7000 cm^{-1} regions, 1st overtone of OH stretching vibrations mainly from carbohydrates dominated the region of 7000–6000 cm^{-1}, 1st overtone of H_2O, RCO_2H, RCOR, $_{Ar}CH$, SH, $CONH_2$, CH, CH_2, CH_3 and aromatic CH stretching vibrations and O-H combination band appear in 6000–5000 cm^{-1}, band combinations of NH,

Table 4. The positions of the wavenumbers with the highest discriminating power in species-based classification.

	Wavenumber (cm^{-1})	Wood species
Gierlinger *et al.* (2004)	6700–6300	*Larix decidua*
		Larix kaempferi
		Larix eurolepis
Braga *et al.* (2011)	8995–7498	*Swietenia macrophylla*
	8246–7498	*Carapa guianensis,*
	7502–4247	*Cedrela odorata*
		Micropholis melinoniana
Pastore *et al.* (2011)	4249–7501	*Swietenia macrophylla*
	5448–6100	*Carapa guianensis*
	5448–5774	*Cedrela odorata* L.
	4249–4600	*Micropholis melinoniana*
Espinoza *et al.* (2012)	5405–4000	Hybrid pine species
Shou *et al.* (2014)	9090–6250	*Pterocarpus santalinus*
	9090–7142	*Dalbergia louvelii*
	10000–6060	*Pterocarpus soyauxii*
Yang *et al.* (2015)	6789, 5195,	*Pometia* sp.
	4780, 4411	*Instia* sp.
		Couratari sp.
Snel *et al.* (2018)	6097–5682	*Dalbergia cearensis* Ducke
	5348–5236	*Dalbergia tucurensis* Donn. Sm.
	4808–4630	*Dalbergia decipularis* Rizzini & A. Mattos
		Dalbergia sissoo DC.
		Dalbergia stevensonii Standl.
		Dalbergia latifolia Roxb.
		Dalbergia retusa Hemsl.
		Dalbergia nigra (Vell.) Benth

H_2O, ROH, RNH_2, $CONH_2$, CHO, CH, CH_2, CH_3, $CONH_2$ were observed around 5000–4000 cm^{-1}, and 5200–4900 cm^{-1} is mainly related to water bands.

When our findings were evaluated with former studies, it was concluded that the range of 4000–7000 (9000) cm^{-1} is effective in wood identification. This spectral range is used for qualitative and quantitative purposes and is mainly dominated by stretching vibrations and combinations of wood components (cellulose, hemicellulose, lignin and extractives) (Xiaobo *et al.* 2010, Schwanninger *et al.* 2011).

Conclusion

This study demonstrated the effects of different preprocessing methods on the classification of hard to distinguished wood species. Near-infrared spectra of pinewood samples were collected and compared with each other. Results of the research show that it is effective to use NIR spectroscopy along with PLS-DA to the classification of the Scotch pine and black pine wood are impossible to distinguish each other by means of microscopic investigations (traditional methods). The correct classification rate without any preprocessing method was found at 84%. After applying different preprocessing methods, accuracy rates were differed between 73% (SNV) and 95% (smoothing and 1st derivative). It could be easily concluded that preprocessing methods are effective on the correct prediction rates of the models, but this effect is not always positive. It has been observed that the range of 4000–7000 cm^{-1} is useful for species-based discrimination. Further investigation should focus on this spectral range for the classification of different wood species, and various preprocessing methods and their combinations should be on trial.

This work showed that near-infrared spectroscopy (NIRS) coupled with the right chemometric methods (preprocessing and modelling) would be a fast, accurate and non-destructive technique for wood identification, even in the distinction of indistinguishable wood species. Also, the selection of the right preprocessing method affects the success of this technique.

Acknowledgements

NIR spectra were collected by Antaris FT-NIR Analyzer (Thermo Nicolet Scientific, USA) at the Forest Products Laboratory, Madison, USA. The data of this study were obtained from the PhD thesis entitled "Utilization of near infrared spectroscopy in wood identification".

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by International Doctoral Research Fellowship Program from TUBITAK 2214-A [grant number 1059B141800316].

ORCID

F. Digdem Tuncer http://orcid.org/0000-0002-6588-4789
Dilek Dogu http://orcid.org/0000-0001-7223-3987
Esra Akdeniz http://orcid.org/0000-0002-3549-5416

References

Bächle, H., Zimmer, B. and Wegener, G. (2012) Classification of thermally modified wood by FT-NIR spectroscopy and SIMCA. *Wood Science and Technology*, 46(6), 1181–1192.

Barnes, R. J., Dhanoa, M. S. and Lister, S. J. (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43(5), 772–777.

Bergo, M. C. J., Pastore, T. C. M., Coradin, V. T. R., Wiedenhoeft, A. C. and Braga, J. W. B. (2016) Nirs identification of Swietenia Macrophylla is robust across specimens from 27 countries. *IAWA Journal*, 37(3), 420–430.

Braga, J. W. B., Pastore, T. C. M., Coradin, V. T. R., Camargos, J. A. A. and Da Silva, A. R. (2011). The use of near infrared spectroscopy to identify solid wood specimens of Swietenia macrophylla(cites appendix II). *IAWA Journal*, 32(2), 285–296.

Brereton, R. G. (2007) *Applied Chemometrics for Scientists* (West Sussex: John Wiley& Sons).

Burns, D. A. and Ciurczak, E. W. (2008) *Handbook of Near Infrared Analysis* (Boca Raton: Taylor & Francis).

Cao, N. (2013) Calibration optimization and efficiency in near infrared spectroscopy. Iowa State University Capstones, Theses and Dissertations. Iowa State University.

Chen, Q., Hu, Z., Chang, H. and Li, B. (2007) Micro analytical methods for determination of compression wood content in loblolly pine. *Journal of Wood Chemistry and Technology*, 27(3–4), 169–178.

Ciaburro, G. and Venkateswaran, B. (2017) *Neural network with R. Packt* (Birgmingham: Packt Publishing).

Clark, J. Y. (2003) Artificial neural networks for species identification by taxonomists. *BioSystems*, 72(1-2), 131–147.

de Andrade, B. G., Basso, V. M. and de Figueiredo Latorraca, J. V. (2020) Machine vision for field-level wood identification. *IAWA Journal*, 41(4), 681–698.

de Jong, S. (1993) SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3), 251–263.

Dahlman, N. (2012) Near infrared spectroscopy – an introductory study on measurement techniques and tools for analysing moisture content in pulp and paper media. Thesis (M.Sc.), Karlstads University, Karlstad, Sweden.

Deklerck, V., Finch, K., Gasson, P., Van den Bulcke, J., Van Acker, J., Beeckman, H. and Espinoza, E. (2017) Comparison of species classification models of mass spectrometry data: Kernel discriminant analysis vs random forest; A case study of Afrormosia (Pericopsis elata (Harms) Meeuwen). *Rapid Communications in Mass Spectrometry*, 31(19), 1582–1588.

Dormontt, E. E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J. C., Koch, G., Lee, S. L., Kanashiro, M., Rimbawanto, A., Thomas, D., Wiedenhoeft, A. C., Yin, Y., Zahnen, J. and Lowe, A. J. (2015) Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biological Conservation*, 191, 790–798.

Espinoza, J. A., Hodge, G. R. and Dvorak, W. S. (2012) The potential use of near infrared spectroscopy to discriminate between different pine species and their hybrids. *Journal of Near Infrared Spectroscopy*, 20(4), 437–447.

Espinoza, E. O., Lancaster, C. A., Kreitals, N. M., Hata, M., Cody, R. B. and Blanchette, R. A. (2014) Distinguishing wild from cultivated agarwood (Aquilaria spp.) using direct analysis in real time and time of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 28(3), 281–289.

Esteban, L. G., de Palacios, P., Conde, M., Fernández, F. G., García-Iruela, A. and González-Alonso, M. (2017) Application of artificial neural networks as a predictive method to differentiate the wood of Pinus sylvestris L. and Pinus nigra Arn subsp. salzmannii (Dunal) Franco. *Wood Science and Technology*, 51(5), 1249–1258.

Evans, P., Heady, R. and Cunningham, R. (2008) Identification of yellow stringybark (Eucalyptus muelleriana) and silvertop ash (E. sieberi) wood is improved by canonical variate analysis of ray anatomy. *Australian Forestry*, 71(2), 94–99.

Evans, P. D., Mundo, I. A., Wiemann, M. C., Chavarria, G. D., McClure, P. J., Voin, D. and Espinoza, E. O. (2017) Identification of selected CITES-protected Araucariaceae using DART TOFMS. *IAWA Journal*, 38(2), 266–2S3.

Fengel, D. and Wegener, G. (1983) *Wood: Chemistry, ultrastructure, reactions* (Berlin: Walter de Gruyter).

Flaete, P. O., Haartveit, E. Y. and Vadla, K. (2006) Near infrared spectroscopy with multivariate statistical modelling as a tool for differentiation of wood from tree species with similar appearance. *New Zealand Journal of Forestry Science*, 36(2), 382–392.

Gasson, P. and MacLachlan, I. R. (2010) PCA of cites listed Pterocarpus santalinus (Leguminosae) wood. *IAWA Journal*, 31(2), 121–138.

Geladi, P. and Kowalski, B. R. (1986) Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185, 1–17.

Gholizadeh, A., Boruvka, L., Saberioon, M. M., Kozák, J., Vašát, R. and Nemecek, K. (2015) Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. *Soil and Water Research*, 10(4), 218–227.

Gierlinger, N., Jacques, D., Schwanninger, M., Wimmer, R., Hinterstoisser, B. and Paques, L. E. (2003) Rapid prediction of natural durability of larch heartwood using Fourier transform near-infrared spectroscopy. *Canadian Journal of Forest Research*, 33, 1727–1736.

Gierlinger, N., Schwanninger, M. and Wimmer, R. (2004) Characteristics and classification of Fourier-transform near infrared spectra of the heartwood of different larch species (Larix sp.). *Journal of Near Infrared Spectroscopy*, 12, 113–119.

Giménez, A. M., Moglia, J. G., Figueroa, M. E. and Calatayu, F. (2014) Comparative wood anatomy of Maytenus in Northwestern Argentina (South America). *Madera y Bosques*, 20(17), 95–110.

Hans, G., Leblon, B., Cooper, P., La Rocque, A. and Nader, J. (2015) Determination of moisture content and basic specific gravity of Populus tremuloides (Michx.) and Populus balsamifera (L.) logs using a portable near-infrared spectrometer. *Wood Material Science & Engineering*, 10(1), 3–16.

Hather, J. G. (2000) *The Identification of Northern European Woods. First* (London: Archetype Publications Ltd).

Hellberg, E. and Carcaillet, C. (2003) Wood anatomy of West European Betula: Quantitative descriptions and applications for routine identification in paleoecological studies. *Ecoscience*, 10(3), 370–379.

Hermanson, J. C. and Wiedenhoeft, A. C. (2011) A brief review of machine vision in the context of automated wood identification systems. *IAWA Journal*, 32(2), 233–250.

Hoffmeyer, P. and Pedersen, J. G. (1995) Evaluation of density and strength of Norway spruce wood by near infrared reflectance spectroscopy. *Holz als Roh- und Werkstoff*, 53, 165–170.

Horikawa, Y., Mizuno-Tazuru, S. and Sugiyama, J. (2015) Near-infrared spectroscopy as a potential method for identification of anatomically similar Japanese diploxylons. *Journal of Wood Science*, 61(3), 251–261.

Hwang, S. W., Horikawa, Y., Lee, W. H. and Sugiyama, J. (2016) Identification of Pinus species related to historic architecture in Korea using NIR chemometric approaches. *Journal of Wood Science*, 62(2), 156–167.

Jiang, Z. H., Yang, Z., So, C. L. and Hse, C. Y. (2007) Rapid prediction of wood crystallinity in Pinus elliotii plantation wood by near-infrared spectroscopy. *Journal of Wood Science*, 53(5), 449–453.

Jiao, L., Yin, Y., Cheng, Y. and Jiang, X. (2014) DNA barcoding for identification of the endangered species Aquilaria sinensis: Comparison of data from heated or aged wood samples. *Holzforschung*, 68(4), 487–494.

Jones, P. D., Schimleck, L. R., Peter, G. F., Daniels, R. F. and Clark, A. (2005) Non-destructive estimation of Pinus taeda L tracheid morphological characteristics for samples from a wide range of sites in Georgia. *Wood Science and Technology*, 39(7), 529–545.

Kelley, S. S., Rials, T. G., Groom, L. R. and So, C. L. (2004) Use of near infrared spectroscopy to predict the mechanical properties of six softwoods. *Holzforschung*, 58(3), 252–260.

Kucheryavskiy, S. (2019) mdatools: Multivariate Data Analysis for Chemometrics. R package version 0.9.4. https://CRAN.R-project.org/package=mdatools.

Kvalheim, O. M. (2020) Variable importance: Comparison of selectivity ratio and significance multivariate correlation for interpretation of latent-variable regression models. *Journal of Chemometrics*, 34(4), 1–10.

Lang, C., Costa, F. R. C., Camargo, J. L. C., Durgante, F. M. and Vicentini, A. (2015) Near infrared spectroscopy facilitates rapid identification of both young and mature Amazonian tree species. *PLoS One*, 10(8), 1–15.

Lazarescu, C., Hart, F., Pirouz, Z., Panagiotidis, K., Mansfield, S. D., Barrett, J. D. and Avramidis, S. (2017) Wood species identification by near-infrared spectroscopy. *International Wood Products Journal*, 8(1), 32–35.

Leandro, J. G. R., Gonzaga, F. B. and Latorraca, J. V. d. F. (2019) Discrimination of wood species using laser-induced breakdown spectroscopy and near-infrared reflectance spectroscopy. *Wood Science and Technology*, 53(5), 1079–1091.

Li, Y., Via, B. K., Young, T. and Li, Y. (2019) Visible-near infrared spectroscopy and chemometric methods for wood density prediction and origin/species identification. *Forests*, 10(12), 1078.

Marques, J. B. C., Callado, C. H., Rabelo, G. R., Neto, S., da, S. J. and Cunha, M. d. (2015) Comparative wood anatomy of species of Psychotria L. (Rubiaceae) in Atlantic Rainforest remnants of Rio de Janeiro State, Brazil. *Acta Botanica Brasilica*, 29(3), 433–444.

Mehmood, T., Sæbø, S. and Liland, K. H. (2020) Comparison of variable selection methods in partial least squares regression. *Journal of Chemometrics*, 34(6), 1–14.

Mevik, B., Wehrens, R. and Hovde Liland, K. (2019) pls: Partial Least Squares and Principal Component Regression. R package version 2.7-2. https://CRAN.R-project.org/package=pls.

Naes, T., Isaksson, T., Fearn, T. and Davies, T. (2004) *A User-Friendly Guide to Multivariate Calibration and Classification* (Chichester, West Sussex: Biddles, NIR Publications).

Olivieri, A. C. (2018) *Introduction to Multivariate Calibration* (Cham: Springer International Publishing).

Pace, J. H. C., Latorraca, J. V. F., Hein, P. R. G., Castro, J. P., Carvalho, A. M. and Silva, C. E. S. (2019) Wood species identification from Atlantic forest by near infrared spectroscopy. *Forest Systems*, 28, 3.

Pastore, T. C. M., Braga, J. W. B., Coradin, V. T. R., Magalhães, W. L. E., Okino, E. Y. A., Camargos, J. A. A., De Muñiz, G. I. B., Bressan, O. A. and Davrieux, F. (2011) Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixá. *Holzforschung*, 65(1), 73–80.

Pasquini, C. (2003) Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), 198–219.

Popescu, C.-M., Navi, P., Placencia Peña, M. I. and Popescu, M.-C. (2018) Structural changes of wood during hydro-thermal and thermal treatments evaluated through NIR spectroscopy and principal component analysis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 191, 405–412.

R Core Team. (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rajalahti, T., Arneberg, R., Kroksveen, A. C., Berle, M., Myhr, K. M. and Kvalheim, O. M. (2009) Discriminating variable test and selectivity ratio plot: Quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles. *Analytical Chemistry*, 81(7), 2581–2590.

Ravindran, P., Costa, A., Soares, R. and Wiedenhoeft, A. C. (2018) Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods*, 14(1), 1–10.

Rinnan, Å, Berg, F. v. d. and Engelsen, S. B. (2009) Review of the most common pre-processing techniques for near-infrared spectra. *TrAC – Trends in Analytical Chemistry*, 28(10), 1201–1222.

Sandberg, K. and Sterley, M. (2009) Separating Norway spruce heartwood and sapwood in dried condition with near-infrared spectroscopy and multivariate data analysis. *European Journal of Forest Research*, 128(5), 475–481.

Savitzky, A. and Golay, M. J. E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36 (8), 1627–1639.

Schimleck, L. R., Evans, R., Jones, P. D., Daniels, R. F., Peter, G. F. and Clark, A. (2005) Estimation of microfibril angle and stiffness by near infrared spectroscopy using sample sets having limited wood density variation. *IAWA Journal*, 26(2), 175–187.

Schimleck, L. R., Jones, P. D., Peter, G. F., Daniels, R. F. and Clark III, A. (2004) Nondestructive estimation of tracheid length from sections of radial wood strips by near infrared spectroscopy. *Holzforschung*, 58, 375–381.

Schoch, W., Heller, I., Schweingruber, F. H. and Kienast, F. (2004) Wood anatomy of central European species, available at: www.woodanatomy.ch.

Schwanninger, M., Hinterstoisser, B., Gierlinger, N., Wimmer, R. and Hanger, J. (2004) Application of Fourier transform near infrared spectroscopy (FT-NIR) to thermally modified wood. *Holz als Roh- und Werkstoff*, 62(6), 483–485.

Schwanninger, M., Rodrigues, J. C. and Fackler, K. (2011) A review of band assignments in near infrared spectra of wood and wood components. *Journal of Near Infrared Spectroscopy*, 19(5), 287–308.

Schweingruber, F. H. (2007) *Wood Structure and Environment* (Berlin, Heidelberg: Springer US).

Shou, G., Zhang, W., Gu, Y. and Zhao, D. (2014) Application of near infrared spectroscopy for discrimination of similar rare woods in the Chinese market. *Journal of Near Infrared Spectroscopy*, 22(6), 423–432.

Signal developers (2013) signal: Signal processing. URL:http://r-forge.r-project.org/projects/signal/.

Snel, F. A., Braga, J. W. B., da Silva, D., Wiedenhoeft, A. C., Costa, A., Soares, R., Coradin, V. T. R. and Pastore, T. C. M. (2018) Potential field-deployable NIRS identification of seven Dalbergia species listed by CITES. *Wood Science and Technology*, 52(5), 1411–1427.

So, C. L., Lebow, S. T., Groom, L. H. and Rials, T. G. (2004) The application of near infrared (NIR) spectroscopy to inorganic preservative-treated wood. *Wood And Fiber*, 36(3), 329–336.

Taylor, A. M., Baek, S. H., Jeong, M. K. and Nix, G. (2008) Wood shrinkage prediction using NIR spectroscopy. *Wood and Fiber Science*, 40(2), 301–307.

Tounis, E. (2009) Investigation of nir spectroscopy for identifying and sorting wood with respect to species , moisture content , and weathering, Thesis (M.Sc.), Ryerson University, Toronto, Canada.

Tsuchikawa, S., Inoue, K., Noma, J. and Hayashi, K. (2003) Application of near-infrared spectroscopy to wood discrimination. *Journal of Wood Science*, 49(1), 29–35.

Tsuchikawa, S. and Kobori, H. (2015) A review of recent application of near infrared spectroscopy to wood science and technology. *Journal of Wood Science*, 61(3), 213–220.

Tuncer, F. D. (2020) *Utilization of Near Infrared Spectroscopy in Wood Identification* (Istanbul University-Cerrahpasa).

Turhan, K. and Serdar, B. (2013) Support vector machines in wood identification: The case of three Salix species from Turkey. *Turkish Journal of Agriculture and Forestry*, 37(2), 249–256.

Via, B. K., So, C. L., Groom, L. H., Shupe, T. F., Stine, M. and Wikaira, J. (2007) Within tree variation of lignin, extractives, and microfibril angle coupled with the theoretical and near infrared modeling of microfibril angle. *IAWA Journal*, 28(2), 189–210.

Via, B. K., Zhou, C., Acquah, G., Jiang, W. and Eckhardt, L. (2014) Near infrared spectroscopy calibration for wood chemistry: Which chemometric technique is best for prediction and interpretation? *Sensors (Switzerland*), 14(8), 13532–13547.

Wang, Y., Xiang, J., Tang, Y., Chen, W. and Xu, Y. (2021) A review of the application of near-infrared spectroscopy (NIRS) in forestry. *Applied Spectroscopy Reviews*, 1–18.

Wehrens, R. (2011) *Chemometrics With R: Multivariate Data Analysis in the Natural Sciences and Life Sciences* (Heidelberg: Springer).

Wheeler, E. A. and Baas, P. (1998) Wood identification – A review. *IAWA Journal*, 19(3), 241–264.

Wiesner, K., Fuchs, K., Gigler, A. M. and Pastusiak, R. (2014) Trends in near infrared spectroscopy and multivariate data analysis from an industrial perspective. *Procedia Engineering*, 87, 867–870.

Workman, J. J. and Weyer, L. (2012) *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy* (2nd ed) (Boca Raton, FL: Taylor & Francis).

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M. and Hanpin, M. (2010) Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, 667(1–2), 14–32.

Yang, Z., Liu, Y., Pang, X. and Li, K. (2015) Preliminary investigation into the identification of wood species from different locations by near infrared spectroscopy. *BioResources*, 10(4), 8505–8517.

Yong, H., Qing-yuan, S., Min, R. and Yuan, H. (2019) Identification of wood species based on near infrared spectroscopy and pattern recognation method. *Spectroscopy and Spectral Analysis*, 39(03), 705–710.

Zhou, Z., Rahimi, S., Avramidis, S. and Fang, Y. (2020) Species-and moisture-based sorting of Green timber mix with near infrared spectroscopy. *BioResources*, 15(1), 317–330.

Zobel, B. J. and Van Buijtenen, J. P. (1989) *Wood Variation: Its Causes and Control* (Berlin: Springer-Verlag).

Appendix

Table A1. Identifier for Pinus sylvestris L.

Wood species	Sample collection	Provenances	Tree code	1.30-m diameter (cm)	Age	Coordinates	
Pinus sylvestris L.	Wood discs	Turkey, Sinop-Ayancik: Kepez Forest Management Department	S1	62	114	34.532131	41.766068
			S2	56	138	34.532128	41.755062
			S3	54	142	34.532144	41.766152
			S4	48	129	34.532155	41.766140
			S5*	46	121	34.532188	41.766112
			S6	50	137	34.532188	41.766066
			S7	62	139	34.532165	41.766013
			S8*	54	126	34.532170	41.765995
			S9*	54	132	34.532165	41.765918
			S10	36	138	34.532148	41.765977
	Increment bores	Turkey, Istanbul-Bahcekoy: Bentler Forest Management Department	S11	40	63	28.9882156	41.1755548
			S12	38	58	28.9883700	41.1753410
			S13*	32	61	28.9884696	41.1753757
			S14	44	70	28.9883973	41.1756475
			S15	44	62	28.9884778	41.17575
			S16*	34	65	28.9887654	41.1754801
			S17	36	66	28.9889469	41.1759666
			S18	36	57	28.9874645	41.1785613
			S19*	31	63	28.9873406	41.1784888
			S20	38	64	28.9872852	41.1784201
			S21	40	60	28.9871709	41.178479
			S22	40	59	28.9871635	41.1767316
			S23	36	66	28.9876135	41.1767215
			S24	42	66	28.9878576	41.1768542
			S25*	38	60	28.9879091	41.1769735

*Spectra collected from these trees were used for validation (test sets).

Table A2. Identifier for *Pinus nigra* J.F. Arnold.

Wood species	Sample collection	Provenances	Tree Code	1.30-m diameter (cm)	Age	Coordinates	
Pinus nigra J.F. Arnold	Wood discs	Turkey, Sinop-Ayancik: Kepez Forest Management Department	K1	60	109	34. 532100	41.773382
			K2*	66	117	34.532124	41.773365
			K3	64	127	34.532131	41.773301
			K4*	46	122	34.532189	41.773195
			K5	50	138	34.532204	41.773114
			K6	58	142	34.532146	41.773167
			K7*	66	147	34.532130	41.773148
			K8	74	134	34.532148	41.773137
			K9	64	155	34.532165	41.773199
			K10	64	162	34.532141	41.773194
	Increment bores	Turkey, Istanbul-Bahcekoy: Bentler Forest Management Department	K11*	42	68	28.9884797	41.1755729
			K12	38	72	28.9869191	41.1756155
			K13	58	67	28.9869014	41.1783159
			K14	58	58	28.9869014	41.1782833
			K15*	54	53	28.9872683	41.1783291
			K16	42	57	28.9871077	41.1783550
			K17	32	49	28.9871559	41.178400
			K18	44	55	28.9876897	41.1773978
			K19*	40	62	28.9876005	41.1773504
			K20	38	62	28.9872864	41.1786318
			K21	36	60	28.9873745	41.1786303
			K22*	40	64	28.9873149	41.1789665
			K23	38	57	28.9871626	41.1789387
			K24	40	61	28.9873103	41.1789337
			K25	42	64	28.9875773	41.1788513

*Spectra collected from these trees were used for validation (test sets).