Rapid Prediction of the Chemical Information of Wood Powder from Softwood Species Using Near-Infrared Spectroscopy

Se-Yeong Park, a Jong-Chan Kim, a Seunghoon Yeon, a Sang-Yun Yang, a Hwanmyeong Yeo, a, b and In-Gyu Choi a, b, c, *

Five different softwoods were used to investigate fast methods for predicting quantitative chemical information via near-infrared (NIR) spectroscopy. In biomass-related industries, fast collection of chemical information from a feedstock is needed. Prior to predicting quantitative information, a principal component analysis (PCA) using NIR spectra was conducted to evaluate the possibility of discriminating the softwoods. As a result of PCA, the five species were divided into three groups. This result indicated that the extractive compounds were key factors because the powder samples were separated by species having a similar extractive content. The partial least square (PLS) method was applied to develop a calibration model for predicting chemical composition. This model showed good performance in predicting the extractive and lignin content of all species. The calibration results of the extractive and lignin content for all species were indicated as $R^2 = 0.99$. The cross-validation of the components for all species also showed an excellent value of $R^2 = 0.98$ and 0.97, respectively. Based on our results, it was possible to suggest a useful tool for providing rapid information about wood used in the bioenergy and pulp production fields.

Keywords: Softwood species; Near-infrared spectroscopy; Partial least square regression; Chemical composition; Extractive; Lignin; Non-destructive measurement

Contact information: a: Department of Forest Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea; b: Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea; c: Institutes of Green-Bio Science and Technology, Seoul National University, Pyeongchang 25354, Republic of Korea; * Corresponding author: cingyu@snu.ac.kr

INTRODUCTION

With an increase in use of wood materials in various industries, it is necessary to rapidly know chemical and physical information about the wood. In particular, the chemical properties are influenced by the condition of the wood, and if the wood has too many defects, the chemical composition may differ from a normal specimen of the wood. In addition, the chemical structure is different depending on the section of the wood (Onnerud 2003; Wadenbäck et al. 2004). Furthermore, a natural phenomenon related to wood degradation or deterioration via external factors can result in the restriction of the use of wood (Schultz et al. 2007; Zhang et al. 2009; Ozdemir et al. 2015). Therefore, it is important to quickly find defects and abnormal conditions in wood materials on the spot. Although the defects and differences on the surface of the wood can be observed easily, it is harder to know the chemical information of other wood-based materials such as sawdust,
chips, and powder, which are not uniform in shape. For this reason, a faster and easier method to determine chemical information of the wood materials is needed.

The traditional wet analysis of wood chemical properties is time-consuming and inefficient; it is typically conducted in a laboratory and is labor intensive. To overcome this inconvenience, spectroscopic techniques such as near-infrared (NIR) and visible infrared have been developed to predict the chemical composition of wood.

Particularly, NIR spectroscopy has been introduced as a promising method for the prediction of wood properties. The technique is a simple task with no chemical destruction and easy preparation of samples (Yeh et al. 2004; Schimleck 2008; Xu et al. 2013). Many studies have already suggested the use of time-saving analytical models based on NIR spectroscopy. This allows cellulose, lignin, and extractive contents in wood or biomass to be measured automatically (Ishizuka et al. 2014; Li et al. 2015). It is possible to predict more precise information, such as the syringyl/guaiacyl ratio and the mono-sugar distribution. In addition, other studies have been reported for predicting mechanical properties such as moisture content, density, and air-dried specific gravity (Leblon et al. 2013; Yang et al. 2017). Such information can improve the efficiency in not only wood industries but also bioenergy-based plant designs (Alves et al. 2012; López et al. 2017). However, the biomass industry has fewer attempts to apply the NIR system for the rapid prediction of the chemical properties of a wood.

In this study, NIR spectroscopy was employed for collecting the chemical information of five softwoods grown in the Republic of Korea. Characteristically, specimens for the experiment were prepared as powder state to eliminate the influence of the wood direction. Wood samples were used to develop regression models for predicting chemical composition using the partial least square (PLS) technique. The calibration and cross-validation results of the extractive and lignin contents were evaluated. In addition, a principal component analysis (PCA) was tested for the classification of softwood species.

**EXPERIMENTAL**

**Materials**

Larch (*Larix kaempferi*), cedar (*Cryptomeria japonica*), cypress (*Chamaecyparis obtusa*), red pine (*Pinus densiflora*), and Korean pine (*Pinus koraiensis*) were purchased from the National Forestry Cooperative Federation in the Republic of Korea. Lumber pieces from each species were obtained as 50 × 100 × 600 mm (thickness × width × length) boards, which were cut into 50 samples. Each sample was stranded five times at 2 mm intervals. A total of 250 sawdust samples were prepared from each species. The sawdust was ground and filtered through a 40 mesh sieve. The ground powder was stored at room temperature in a sealed plastic bag.

**Methods**

*Near-infrared (NIR) spectrum measurement*

NIR spectra were obtained using a SpectraStar 2600 XL instrument (Unity Scientific, Milford, MA, USA). The 40 mesh size sample was placed in a cup holder with 100 × 50 mm (diameter × length), and 12 scans were recorded and averaged from 680 to 2600 nm wavelength at intervals of 1 nm. OMNIC 9.2 software (Thermo Scientific, Waltham, MA, USA) was used for spectra analysis. The spectra analysis was performed
with additional data pretreatment, and a Savitsky-Golay 2\textsuperscript{nd} derivative (polynomial order: 3, smoothing point: 21) was used to compare the differences more precisely between the wood species.

Selection of representative samples
A total of 250 samples for each species were prepared to obtain the NIR spectra. However, the chemical analysis of all samples was not possible. Thus, representative samples for each species were selected based on the Mahalanobis distance theory. The Mahalanobis distance in a multidimensional space was introduced to NIR spectroscopy research by Mark and Tunnel (1985). In a principal component space, the standardized Mahalanobis distance is applied to consist of clusters of NIR spectra as well as the selection of calibration samples. The distance is used in two ways: neighborhood H (NH), used when constructing a set of calibration samples, and global H (GH), used when determining the boundaries of the spectral population and the detection of outliers (Shenk and Westerhaus 1991a,b). NH was chosen for this study to select representative samples for further calibration of the set. The NH of larch, cypress, and cedar was set to 1.5, whereas the NH value of red pine and Korean pine was 1.0. A different NH value was applied to the selection step because the number of red and Korean pines was too small to develop the calibration model at 1.5. Therefore, the NH was reset to 1.0 for both pine species. Totally, 150 samples were selected as representatives (Larch: 29, Cypress: 25, Cedar: 26, Red pine: 29, Korean pine: 41).

Chemical composition analysis
The extractive and lignin contents were analyzed according to the National Renewable Energy Laboratory procedures (Sluiter et al. 2005, 2008). An alcohol-benzene (1:2, v/v) solution was used to extract the extractive compounds in woods via a Soxhlet extractor at 80 °C for 6 h. The solute was evaporated to remove the solvent. The extractive content was calculated by the oven-dry weight. The lignin content was analyzed via the Klasson lignin method using 0.3 g of wood powder, which was placed in a flask with 3 mL of 72% H\textsubscript{2}SO\textsubscript{4} for 1 h at 30 °C. Next, 84 mL of distilled water was added to the flask, and the hydrolysis reaction was continued at 120 °C for 1 h. When the reaction was finished, the solid and liquid fractions were separated using an aspirator. The weight of the solid residue as an acid insoluble lignin was calculated, and the acid hydrolyzed compounds in the liquid fraction (acid-soluble lignin) were measured via a UV/Vis spectrometer.

Classification via principal component analysis (PCA)
The score plot of the first principal component (PC1) and the second principal component (PC2) was based on the PCA, and it was used to observe the clusters in the selected samples. Unscrambler software (CAMO, 9.7, OSLO, Norway) was used for the multicomponent analysis with the PCA. To improve the spectral errors, a data pretreatment was performed using the Savitsky-Golay 2\textsuperscript{nd} derivative (polynomial order: 3, smoothing point: 21).

Regression model via partial least square (PLS)
A PLS analysis was employed to develop the prediction models for each species. Prior to model evaluation, preprocessing conditions were set. The standard normal variate (SNV), and a detrend was used to lower the electrical error. In addition, a forward gap 1\textsuperscript{st} derivative (gap size = 8, smoothing = 8) was applied. A PLS regression model for the
extractive and lignin content via NIR spectroscopy was calibrated with the preprocessed spectra and the results of the chemical analysis using UCal NIR calibration software (version 3.0, Unity Scientific, Milford, MA, USA). The calibration set and cross-validation were divided into five groups followed by evaluation for each and all species. The model calibration was validated by the coefficient of determination (R$^2$), the root mean square error of calibration (RMSEC), and the root mean square error of cross-validation (RMSECV).

**RESULTS AND DISCUSSION**

**NIR Spectra Analysis**

The five softwood species used in this study are native to a wide range of Korea. There were differences in the chemical composition of each wood, which can be influenced by the environmental conditions, such as moisture and nutrients. The different chemical compositions and contents can be a problem in species identification. The 250 ground samples for each species were collected, and a total of 1,250 spectra reports were obtained. Unlike wood sawdust or lumber, the wood powder used in the study had no effect on the absorbance difference according to the directions for measurement. Fig. 1(A) shows the raw spectra of all samples. It was not easy to compare the spectral differences among the five species. It was difficult to interpret the precise spectral information without any preprocessing because the primary and secondary overtones are reflected in the near-infrared region. Overlapping and broad spectral data was observed. Therefore, a mathematical pretreatment using the Savitsky-Golay 2nd derivative was applied to compare the spectra among the five species, and the results are presented in Fig. 1(B). The spectral region ranging from 1600 to 1800 nm is commonly represented as the extractive and lignin band assignment, whereas the 2000 to 2300 nm wavelength is related to the cellulose and hemicellulose bands. The distinct differences at the 1600 to 1800 nm wavelength were attributed to the 1st overtone of the C-H bond originated from the –CH$_2$, -CH$_3$, and the =CH$_2$, found in aromatic compounds. More specifically, the peak of 1672 and 1685 nm may be originated from the C-H stretch in aromatics (Bailлерesa et al. 2002; Fujimoto et al. 2007; Workman and Weyer 2007; Schwanninger et al. 2011). Thus, the five species used in this study exhibited unique patterns, especially in the aromatic compounds, which is a key factor in gathering information for wood discrimination.

![Fig. 1. NIR spectra of all selected species (A) raw spectra and (B) 2nd derivative spectra depending on species](image-url)
Chemical Composition of Representative Samples

Prior to performing the principal component analysis (PCA) for the evaluation of the discriminability of five softwood species, it was necessary to collect the chemical composition of the wood powders. However, it is difficult to analyze all of 1,250 samples. Thus, dozens of samples were selected as representative samples with similar chemical properties for each species. Table 1 summarizes the chemical composition of the selected samples based on the neighborhood distance. In larch, the content of the extractives was 1.6 to 5.5%, and the lignin content was 23.3 to 29.7%. The contents of the extractives and lignin were similar in cedar and cypress. In particular, the lignin contents of cedar and cypress were higher than other species. Previous research noted similar results using the same species (Japanese cedar and Japanese cypress) collected in the Japanese Archipelago (Ishizuka et al. 2014). The extractive content of red pine ranged from 3.9 to 18.3%, and that of Korean pine was between 3.3 and 22.7%. Both pine species belong to the same Pinus family and are known to accumulate large amount of resin (Ekeberg et al. 2006; Kim et al. 2010). They showed a broad range of extractive content.

Table 1. Summary of Extractive and Lignin Contents in Selected Samples for Each Species

<table>
<thead>
<tr>
<th>Species</th>
<th>NH*</th>
<th>Number of Selected Samples</th>
<th>Extractive (%)</th>
<th>Lignin (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Min.</td>
<td>Max.</td>
</tr>
<tr>
<td>Larch</td>
<td>1.5</td>
<td>29</td>
<td>1.6</td>
<td>6.6</td>
</tr>
<tr>
<td>Cedar</td>
<td>1.5</td>
<td>25</td>
<td>0.7</td>
<td>7.1</td>
</tr>
<tr>
<td>Cypress</td>
<td>1.5</td>
<td>26</td>
<td>1.3</td>
<td>6.2</td>
</tr>
<tr>
<td>Red pine</td>
<td>1.0</td>
<td>29</td>
<td>3.9</td>
<td>18.3</td>
</tr>
<tr>
<td>Korean pine</td>
<td>1.0</td>
<td>41</td>
<td>3.3</td>
<td>22.7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>150</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*NH: Neighborhood distance

Wood Classification via PCA

The PCA for the selected 150 samples was performed using the raw spectra data. If the score plot are scattered clearly by the different wood species, it would be better to differentiate them by each species. As shown in Fig. 2(A), the species boundary was not clear based on PC1 (86%) and PC2 (10%), except for the larch. Additional preprocessing was applied to eliminate errors that occur during the acquisition of NIR spectra, and mathematical processing was performed. In the NIR absorbance region, sharp peaks appeared at 1380 to 1480 nm and 1830 to 1950 nm, which were originated from the measurement limit of instrumentation and moisture (Fig. 1(B)). Therefore, the spectra with the 2nd derivative pretreatment was processed by eliminating the two regions additionally for a precise comparison. After preprocessing, a cluster formation in the PCA plot was clearly observed as shown in Fig. 2(B). This means that it is important to establish a specific spectral region that is attributed to the wood components. Finally, the cluster was divided into three groups (1 group: larch; 2 group: cedar and cypress; 3 group: red pine and Korean pine). Meanwhile, Fig. 2(C) shows the comparison between the PC1 loading value and the spectra with the 2nd derivative treatment for all selected species. The black line indicates the loading value, while the color lines are the spectra with the 2nd derivative.
Fig. 2. PCA analysis score plots based on NIR spectra and its loading; A: raw spectra, B: 2nd derivative spectra with removing the measurement limit of the instrumentation and the moisture, C: PC1-loading

Loading refers to the degree of difference between the spectra, and intensive peaks at specific regions mean the presence of the big difference. The spectral pattern in the loading value was similar to the peaks in the spectra with the 2nd derivative. The spectral changes at 1600 to 1800 nm can be explained by the C-H bonds (-CH$_3$ and -CH) attributed
to the aromatic compounds such as the extractives and lignin. In addition, the spectral changes in the 2100 to 2400 nm region appear due to the difference in absorptions by the cellulose and hemicellulose C-H bonds (-CH$_2$ and CH). These results indicated that there are structural differences among each species although they were softwood species with similar wood components.

**Prediction of Chemical Composition via PLS Model**

The development of calibration model was conducted via a PLS method for the prediction of the chemical composition of the five species. For the development of the calibration formula using NIR spectra, SNV, detrend, the forward gap, and the 1$^{st}$ derivative were applied as a mathematical preprocessing. The predicted chemical components via NIR spectra and wet chemistry measurements were generated by the PLS regression procedure and shown in Fig. 3. Several previous papers mentioned that separating the specific spectrum range improved the calibration performance because it was more effective for good calibration results (He and Hu 2013; Zhou et al. 2015; Yang et al. 2016). In our study, the full spectral range was applied to develop the calibration model, considering the measurement limit of the instrumentation and the moisture. The correlation was high between the predicted and measured chemical contents, demonstrating the feasibility of the PLS regression in predicting the chemical composition of the five softwood species.
Fig. 3. PLS calibration results of extractive and lignin contents of larch, cedar, cypress, red pine, Korean pine and all species
Results of the calibration and validation models for the quantitative compositional analysis of both each and all species are shown in Table 2. Specifically, each species showed the calibration results for the extractive content with a $R^2$ of more than 0.94 of and a RMSEC of less than 0.48. For lignin, the calibration results were obtained with a $R^2$ greater than 0.97 and a RMSEC less than 0.43. The cross-validation results of the extractives also showed good prediction performance for each species, although larch indicated a relatively lower value with $R^2 = 0.78$ when compared to other species. The $R^2$ was 0.93 for cedar, 0.86 for cypress, 0.98 for red pine, and 0.97 for Korean pine, respectively. The detailed RMSECV values of the extractive contents for each species are presented in Table 2. For the lignin after cross-validation, the $R^2$ was shown as 0.92 for larch, 0.88 for cedar, 0.91 for cypress, 0.91 for red pine and 0.89 for Korean pine. All selected samples regardless of the species showed an excellent performance in the crossvalidation as well as the calibration results with a $R^2 = 0.99$ for both the extractives and lignin. As a result, it is expected that the prediction of the extractive and lignin contents of each and all species used in our study will perform well.

Table 2. Calibration and Validation Indices for Prediction of Extractive and Lignin Contents Using NIR Spectra

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of Selected Samples</th>
<th>Constituent</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSEC</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Larch</td>
<td>29</td>
<td>Extractive</td>
<td>0.2639</td>
<td>0.9586</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lignin</td>
<td>0.1869</td>
<td>0.9894</td>
</tr>
<tr>
<td>Cedar</td>
<td>25</td>
<td>Extractive</td>
<td>0.0493</td>
<td>0.9992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lignin</td>
<td>0.0403</td>
<td>0.9996</td>
</tr>
<tr>
<td>Cypress</td>
<td>26</td>
<td>Extractive</td>
<td>0.3308</td>
<td>0.9364</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lignin</td>
<td>0.0605</td>
<td>0.9984</td>
</tr>
<tr>
<td>Red pine</td>
<td>29</td>
<td>Extractive</td>
<td>0.4139</td>
<td>0.9947</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lignin</td>
<td>0.2847</td>
<td>0.9692</td>
</tr>
<tr>
<td>Korean pine</td>
<td>41</td>
<td>Extractive</td>
<td>0.4769</td>
<td>0.9821</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lignin</td>
<td>0.4304</td>
<td>0.9675</td>
</tr>
<tr>
<td>All species</td>
<td>150</td>
<td>Extractive</td>
<td>0.4639</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lignin</td>
<td>0.4330</td>
<td>0.9864</td>
</tr>
</tbody>
</table>

CONCLUSIONS

1. The five softwood species used in this study had a significantly different content of extractives and lignin. In particular, the red pine and Korean pine showed the highest amount of extractive content and their maximum contents were shown as 18.3% and 22.7, respectively.

2. The possibility of classification of five softwoods via PCA was evaluated, indicating that the three groups were divided depending on their chemical properties (larch / cedar and cypress / red pine and Korean pine). However, it was difficult to classify each species clearly because some species had similar chemical structures. Therefore, more detailed research should be performed to identify unique information for each species.

3. As a result of the PLS analysis, it was possible to demonstrate the good performance on the prediction of the quantitative information for all species. To sum up, the $R^2$
values of the extractives and lignin in the cross-validation were shown as 0.98 and 0.97. These results showed the feasibility of a rapid prediction system via NIR, and thus they will be helpful for providing wood chemical information as a feedstock in various industries such as bioenergy as well as pulp and paper sciences.

ACKNOWLEDGEMENTS

This research was financially supported by the Forest Science and Technology Projects (Project No. 2016009C10-1719-AB01) and funded by the Korea Forest Service.

REFERENCES CITED


Baillèresa, H., Davrieuxa, F., and Ham-Pichavantb, F. (2002). “Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program,” Annals of Forest Science 59(5-6), 479-490. DOI: 10.1051/forest:2002032


Yang, Z., Li, K., Zhang, M., Xin, D., and Zhang, J. (2016). “Rapid determination of

chemical composition and classification of bamboo fractions using visible-near infrared spectroscopy coupled with multivariate data analysis,” *Biotechnology for Biofuels* 9, 35. DOI: 10.1186/s13068-016-0443-z


Article submitted: November 20, 2017; Peer review completed: January 20, 2018; Revised version received: January 24, 2018; Accepted: February 3, 2018; Published: February 13, 2018.
DOI: 10.15376/biores.13.2.2440-2451