J. Korean Wood Sci. Technol. 2017, 45(4): 399~408
pISSN: 1017-0715 eISSN: 2233-7180
https://doi.org/10.5658/WOOD.2017.45.4.399

Original Article

Partial Least Squares Analysis on Near-Infrared Absorbance Spectra by Air-dried Specific Gravity of Major Domestic Softwood Species[1]

Sang-Yun Yang[2] · Yonggun Park[2] · Hyunwoo Chung[2] · Hyunbin Kim[2] · Se-Yeong Park[2] ·
In-Gyu Choi[2,3,6] · Ohkyung Kwon[4] · Kyu-Chae Cho[5] · Hwanmyeong Yeo[2,3,†]

ABSTRACT

Research on the rapid and accurate prediction of physical properties of wood using near-infrared (NIR) spectroscopy has attracted recent attention. In this study, partial least squares analysis was performed between NIR spectra and air-dried specific gravity of five domestic conifer species including larch (*Larix kaempferi*), Korean pine (*Pinus koraiensis*), red pine (*Pinus densiflora*), cedar (*Cryptomeria japonica*), and cypress (*Chamaecyparis obtusa*). Fifty different lumbers per species were purchased from the five National Forestry Cooperative Federations of Korea. The air-dried specific gravity of 100 knot- and defect-free specimens of each species was determined by NIR spectroscopy in the range of 680-2500 nm. Spectral data preprocessing including standard normal variate, detrend and forward first derivative (gap size = 8, smoothing = 8) were applied to all the NIR spectra of the specimens. Partial least squares analysis including cross-validation (five groups) was performed with the air-dried specific gravity and NIR spectra. When the performance of the regression model was expressed as R^2 (coefficient of determination) and root mean square error of calibration (RMSEC), R^2 and RMSEC were 0.63 and 0.027 for larch, 0.68 and 0.033 for Korean pine, 0.62 and 0.033 for red pine, 0.76 and 0.022 for cedar, and 0.79 and 0.027 for cypress, respectively. For the calibration model, which contained all species in this study, the R^2 was 0.75 and the RMSEC was 0.37.

Keywords : near-infrared spectroscopy, partial least squares regression, air-dried specific gravity, major domestic softwood species

---

[1] Date Received February 3, 2017, Date Accepted May 8, 2017

[2] Department of Forest Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

[3] Research Institute of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

[4] National Instrumentation Center for Environmental Management (NICEM), Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

[5] KC Tech In Co. Ltd., 170 Sohyun-ro, Bundang-gu, Seongnam 13590, Republic of Korea

[6] Institutes of Green Bio Science and Technology, Seoul National University, 1447 Pyeongchang-daero, Daehwa-myeon, Pyeongchang 25354, Republic of Korea

[†] Corresponding author: Hwanmyeong Yeo (e-mail: hyeo@snu.ac.kr)

1. INTRODUCTION

Wood is a biological material, with a typical porous cell structure. Each cell is composed of a cell wall and cell lumen. Even within the same species, the thickness of the cell wall and cell lumen differ depending on the climate-growth associations of the trees. Furthermore, the cell wall thickness varies depending on the distance from the pith, pith eccentricity and earlywood/latewood ratio in the individual growth rings. The specific gravity of dry wood cell wall substance is 1.5, irrespective of the species. In contrast, wood cell lumens are filled with water or air. Therefore, the specific gravity of wood has a wide distribution. There is a direct correlation between the physical/mechanical performance of the wood and its specific gravity. Hence, the specific gravity of wood is probably the single most important intrinsic wood property in predicting its performance (Armstrong *et al*., 1984; Zhang, 1995; Forest Products Laboratory, 2010). Several traditional gravimetric methods are available to measure the specific gravity (Williamson and Wiemann, 2010). Alternatively, various non-destructive methods, such as near-infrared spectroscopy (NIR), have been developed based on the correlation between the specific gravity of wood and its chemical composition (Greaves *et al*., 1996; Bergsten *et al*., 2001; Isik and Li, 2003; Park and Telewski, 1993).

NIR is electromagnetic wave ranging from the 780 and 2500 nm. Furthermore, the low energy level of NIR allows rapid analysis of wound-free sample surfaces (Williams and Norris, 2004; Burns and Ciurczak, 2007). NIR spectroscopy has recently been actively studied in the field of wood science (Tsuchikawa and Kobori, 2015) and is known to accurately and readily predict the physical, mechanical and chemical properties of wood (Tsuchikawa, 2007; Cooper *et al*., 2011; Schwanninger *et al*., 2011; Schimleck *et al*., 2011; Hans *et al*., 2013; Yang *et al*., 2015). For instance, several researchers have used NIR spectroscopy to predict the specific gravity or density of wood (Schimleck and Evans, 2003; Via *et al*., 2005; Hein *et al*., 2009; Kothiyal and Raturi, 2011).

In this study, five wood species [larch (*Larix kaempferi*), Korean pine (*Pinus koraiensis*), red pine (*Pinus densiflora*), cedar (*Cryptomeria japonica*), and cypress (*Chamaecyparis obtusa*)] were collected from the five nationwide National Forestry Cooperative Federations of Korea. Regression models were developed to estimate the specific gravity of each species by applying partial least squares (PLS) analysis to the NIR spectra acquired from the defect-free sites of the lumber.

2. MATERIALS and METHODS

2.1. Material

Fifty lumbers of each species [larch (*Larix kaempferi*), Korean pine (*Pinus koraiensis*), red pine (*Pinus densiflora*), cedar (*Cryptomeria japonica*), and cypress (*Chamaecyparis obtusa*)], having dimensions of 50 × 100 × 1200 mm (thickness × width × length), were purchased from several of the National Forestry Cooperative Federation located throughout Korea. Two specimens of 20 × 50 × 100 mm (R × T × L direction) were cut from each lumber. The NIR spectra were acquired at the widest face of the specimen. The specimens had no defects, such as knots and splits. A total 100 specimens were prepared to evaluate their air-dried specific gravity.

2.2. Measurement of Air-dried Specific Gravity

The specific gravity of wood is based on its oven-dried weight but the volume may be in the green, air-dry or oven-dry condition. In the instance of air-dried specific gravity, the volume is in an air-dry condition. In order to induce the air-dry condition of the specimens, all samples were humidified in a constant temperature and humidity room maintained at 25℃ and 60% relative humidity for 2 weeks. The length of the air-dried specimens was measured by a caliper to determine the volume of the air-dried samples. The oven-dried mass of the specimens was measured after oven-drying at 105℃ until constant weight.

2.3. NIR Spectrometry

All NIR reflectance spectra were acquired using a SpectraStar 2500XL (Unity Scientific, US) from 680-2500 nm at intervals of 1 nm, then converted to absorbance [log (1/R)] spectra for calibration. The spectral region ranging from 1300-1400 nm was excluded due to the measurement limit of the instrumentation. Each spectrum was averaged by the 12 scans and two spectra were obtained per specimen at different positions. A total of 200 NIR spectra were used for PLS regression analysis for each species.

A PLS analysis finds relation between the matrix **X** which contains the spectra of the samples and the vector **Y** which stores the described properties. The result is a following equation.

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

Where **b** is the regression coefficient matrix and **e** is the model error matrix. In the PLS model, the matrix **X** is decomposed in to scores, **t**, and weights, **w**, for example, $\mathbf{XW} = \mathbf{T}$, where $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_k)$ is chosen so that $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_k)$ has maximum covariance with **Y** (k is the number of factors). Thus, spectral information and properties are used at the same time in the calibration phase. One of the very important considerations in the PLS model development is the selection of the number of factors k to be included in the model. The similarity of the NIR spectrum was evaluated by standardized Mahalanobis distance (global distance, GD) and t-value to select the population used in the regression model. Any sample exceeding the similarity (GD = 3, t-value = 2) was excluded from the population because spectrum with low similarity can adversely affect the prediction performance of the regression model (Shenk and Westerhaus,

Table 1. Distribution of air-dried specific gravity of specimens

Species	Air-dried specific gravity						EMC* (%)	Average	SD
	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8			
Larch	0	4	28	64	1	3	12.3	0.518	0.064
Korean pine	0	26	55	19	0	0	11.7	0.439	0.061
Red pine	0	20	62	18	0	0	12.8	0.451	0.053
Cedar	17	62	21	0	0	0	14.3	0.348	0.046
Cypress	0	25	57	17	1	0	12.3	0.438	0.059
Total	17	137	223	118	2	3		0.439	0.078

* Equilibrium moisture content.

1991a; Shenk and Westerhaus, 1991b).

Standard normal variate (SNV) and detrend (quadratic fit) were used to reduce the electrical noise and error. Also, the forward gap first derivative (gap size = 8, smoothing = 8) was applied to deconvolute the original spectra due to broad overlapping absorption bands in the NIR region. A PLS regression model was developed for predicting the air-dried specific gravity by NIR spectroscopy with UCal NIR calibration software (version 3.0, Unity Scientific, US) including cross-validation (five groups). The cross-validation technique is used to convince the reliability of prediction models. In the cross-validation procedure, samples are randomly divided into equal-sized groups. One of them is excluded for validation test and remaining groups are used to establish air-dried specific gravity prediction models having different numbers of factors. The models are validated by the excluded group. This process iterates until all the samples are used as model construction and validation.

Models were constructed by raw spectra (unpreprocessed) and preprocessed spectra (SNV, detrend and forward first derivative) for evaluating the effect of mathematical preprocessing. The reliability of the regression model was evaluated 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). The optimal model was determined to have the lowest RMSECV. For a regression model with high reliability, R^2 is close to 1, and RMSEC and RMSECV close to zero.

3. RESULTS and DISCUSSION

3.1. Distribution of Air-dried Specific Gravity

The air-dried specific gravity distribution, average equilibrium moisture content, average air-dried specific gravity and standard deviation (SD) of the specimens used in this study are shown in Table 1. The species with the highest average air-dried specific gravity (0.518) was larch. In decreasing order, the specific gravity of the remaining species was red pine (0.451), Korean pine (0.439), cypress (0.438) and cedar

Table 2. Results of PLS modeling for predicting air-dried specific gravity by raw NIR spectra

Species	n	Factors (k)	Calibration		Validation	
			R^2	RMSEC	R^2	RMSECV
Larch	172	4	0.43	0.033	0.32	0.035
Korean pine	180	8	0.61	0.035	0.55	0.038
Red pine	178	6	0.55	0.033	0.41	0.035
Cedar	185	7	0.69	0.025	0.62	0.027
Cypress	186	5	0.73	0.028	0.64	0.029
Total	877	9	0.73	0.038	0.71	0.039

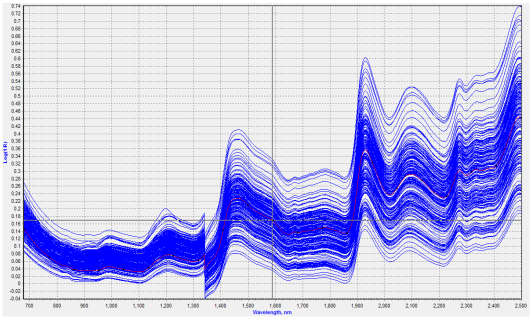

Fig. 1. Raw NIR absorbance spectra [log (1/R)] of larch.

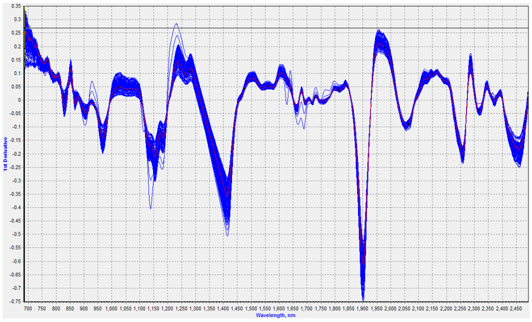

Fig. 2. Mathematically preprocessed [SNV, detrend, forward first derivative (gap size = 8, smoothing = 8)] NIR absorbance spectra [log (1/R)] of larch.

(0.348).

Valdes *et al.* (1987) reported that the population used for regression model development should have a wide range and uniform distribution to develop a reliable regression model. The distribution of the air-dried specific gravity population used in this study was slightly dense at the mean value. However, it had a wide range and was expected to be sufficient for developing the regression models.

3.2. Mathematical Preprocessing of NIR Spectra

Generally, NIR absorption bands are broad and overlapping. Also, they may differ due to density, temperature, surface roughness and particle size of the specimen. These factors result in scattering and baseline changes that induce prediction errors for calibration. Fig. 1 shows the raw NIR absorbance spectra (log (1/R)) of the larch specimens. There were baseline shifts in the absorbance spectra caused by scattering, with high absorption at 1450, 1940, 2100 and 2280 nm. Therefore, the raw spectra were mathematically preprocessed (SNV, detrend and forward gap first derivative (gap size = 8, smoothing = 8)), which removed the baseline offset and separated overlapping NIR bands (Fig. 2).

Table 3. Results of PLS modeling for predicting air-dried specific gravity by preprocessed spectra

Species	n	Factors (k)	Calibration		Validation	
			R^2	RMSEC	R^2	RMSECV
Larch	185	5	0.63	0.027	0.48	0.030
Korean pine	176	5	0.68	0.033	0.55	0.038
Red pine	187	6	0.62	0.033	0.46	0.037
Cedar	188	5	0.76	0.022	0.67	0.025
Cypress	197	7	0.79	0.027	0.70	0.031
Total	948	9	0.75	0.037	0.72	0.038

3.3. Development of Air-dried Specific Gravity Prediction Models

In this study, we developed regression models to predict the air-dried specific gravity of each species and of the combined species. Table 2 shows the reliability of the air-dried specific gravity prediction model developed by a PLS regression of raw spectra, and Table 3 shows that of preprocessed spectra. The number of population (n) determined by the similarity evaluation differed for each species. For example, in the development of the larch air-dried specific gravity regression model, the NIR spectrum obtained from the specimens with high air-dried specific gravity was excluded from the population.

As shown in Table 2 and 3, the calibration results were better than the validation results. It is natural because PLS regression analysis performed in this study contained cross-validation. In a PLS calculation, calibration set is used only for developing prediction models with varied factors. Too many factors of model will lead overfitting only the data and may not be suited to unknown sample. However, if not sufficient factors are selected, underfitting occurs (Martens and Næs, 1991). Thus, it is important to select optimum number of factors by cross-validation. In general, optimum model is selected at the first minimum RMSECV (Ruthenburg *et al.*, 2014). The models described in Table 2 and 3 had a similar factors developed by Shimleck and Evans (2003) and Stirling *et al.* (2007). Both PLS regression result of raw spectra and preprocessed spectra, combined species air-dried specific gravity prediction model had the highest (k = 9) optimum number of factor. Other models of individual species had lower factors. It was estimated that the spectral difference among species affected factor analysis, then led increment of optimum number of factor for combined species. This result also made the highest RMSEC and RMSECV for combined species air-dried specific gravity regression models.

The mathematical preprocessing could improve the quality of the NIR data for PLS regression analysis. The effect of mathematical preprocessing had been reported our previous study (Yang *et al.*, 2015). Comparing regression results of preprocessed spectra with raw spectra,

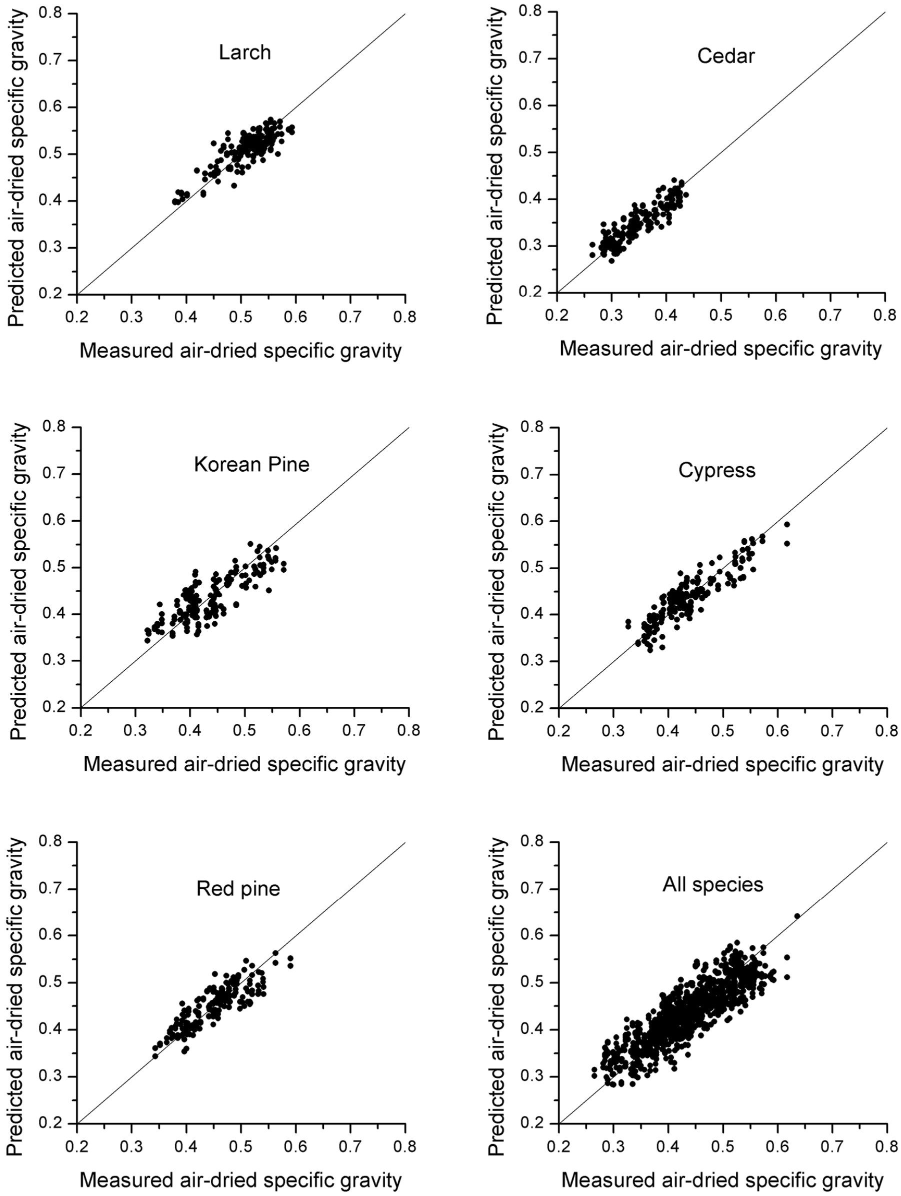

Fig. 3. Results of PLS calibration for air-dried specific gravity of larch, Korean pine, cedar, cypress, red pine and all species.

the coefficient of determination (R^2), the root mean square errors of calibration (RMSEC) and cross-validation (RMSECV) were improved after several mathematical preprocessings (SNV, detrend and forward first derivatives). Several preprocessing applied in this study clearly enhance the precision of air-dried specific gravity prediction model.

As a result of the calibration for the specific gravity prediction model of larch, R^2 was 0.63 and RMSEC was 0.027, which was very low compared to the average air-dried specific gravity of larch (0.518). The cross-validation results of a group not used in the calibration showed that R^2 was 0.48 and RMSECV was 0.030. The R^2 and RMSEC were 0.68 and 0.033 for Korean pine, 0.62 and 0.033 for red pine, 0.76 and 0.022 for cedar, and 0.79 and 0.027 for cypress, respectively. For the calibration model that contained all species in this study, the R^2 was 0.75 and the RMSEC 0.37 (Fig. 3). Thus, NIR spectroscopic calibrations for predicting air-dried specific gravity of several conifer species were successful. Cross-validation results showed a relatively lower reliability than the calibration results, however, the RMSECV values remained low for air-dried specific gravity prediction. The performances of models varied by species. The R^2 of cross-validation for larch, Korean pine and red pine were relatively lower than that of other species. It was supposed that earlywood/latewood ratio of the NIR acquisition face affected the performance of models. The NIR spectrum was acquired from the tangential or radial face of specimen which mixed with earlywood and latewood. The specific gravity of latewood is extremely higher than that of earlywood. The three species which showed low cross-validation results had more distinct and thicker latewood than cedar and cypress. Thus, differences of the between the air-dried specific gravity the NIR acquisition area and that of whole the specimen necessarily arose. This could be one of the reason decreasing the prediction performance of models.

4. CONCLUSION

In this study, PLS regression models that can predict the air-dried specific gravity of larch, Korean pine, red pine, cedar, cypress, which were obtained from the several of National Forestry Cooperative Federation in Korea, were developed by NIR spectroscopy. Using NIR spectra acquired from the specimen, models which can predict the air-dried specific gravity of the species used in this study were developed by PLS analysis. Comparing PLS regression analysis using raw spectra and mathematical preprocessed spectra, more precise and general model could be developed when mathematical preprocessing was performed. The prediction performance of models varied by the species. It was supposed that the anatomical difference, earlywood-latewood ratio, affected to the results.

ACKNOWLEDGEMENT

This work was financially supported by the Forest Science & Technology Projects (Project No. S111616L060110) provided by the Korea Forest Service.

REFERENCES

Armstrong, J.P., Skaar, C., deZeeuw, C. 1984. The effect of specific gravity on some mechanical properties of some world woods. Wood Science

Technology 18(2): 137～146.

Bergsten, U., Lindeberg, J., Rindby, A., Evans, R. 2001. Batch measurements of wood density on intact or prepared drill cores using x-ray microdensitometry. Wood Science and Technology 35(5): 435～452.

Burns, D.A., Ciurczak, E.W. 2007. Handbook of near-infrared analysis 3rd ed. CRC press. Boca Raton. USA.

Cooper, P.A., Jeremic, D., Radivojevic, S., Ung, Y.T., Leblon, B. 2011. Potential of near-infrared spectroscopy to characterize wood products. Canadian Journal of Forest Research 41(11): 2150～2157.

Forest Products Laboratory. 2010. Wood Handbook: Wood as an Engineering Material. General Technical Report 190, USDA Forest Products Laboratory, Madison, WI, USA.

Greaves, B.L., Borralho, N.M., Raymond, C.A., Farrington, A. 1996. Use of a Pilodyn for the indirect selection of basic density in Eucalyptus nitens. Canadian Journal of Forest Research 26(9): 1643～1650.

Hans, G., Leblon, B., Stirling, R., Nader, J., LaRocque, A., Cooper, P. 2013. Monitoring of moisture content and basic specific gravity in black spruce logs using a hand-held MEMS-based near-infrared spectrometer. The Forestry Chronicle 89(5): 607-620.

Hein, P.R.G., Lima, J.T., Chaix, G. 2009. Robustness of models based on near infrared spectra to predict the basic density in Eucalyptus urophylla wood. Journal of near infrared spectroscopy 17(3): 141～150.

Isik, F., Li, B. 2003. Rapid assessment of wood density of live trees using the Resistograph for selection in tree improvement programs. Canadian Journal of Forest Research 33(12): 2426～2435.

Kothiyal, V., Raturi, A. 2011. Estimating mechanical properties and specific gravity for five-year-old Eucalyptus tereticornis having broad moisture content range by NIR spectroscopy. Holzforschung 65(5): 757～762.

Park, W.K., Telewski, F.W. 1993. Measuring maximum latewood density by image analysis at the cellular level. Wood and Fiber Science 25(4): 326～332.

Martens, H., Næs, T. 1991. Multivariate calibration. John Wiley & Sons, Chichester, U.K.

Ruthenburg, T.C., Perlin, P.C., Liu, V., McDade, C.E., Dillner, A.M. 2014. Determination of organic matter and organic matter to organic carbon ratios by infrared spectroscopy with application to selected sites in the improve network. Atmospheric Environment 86: 47～57.

Schimleck, L.R., Monteiro de Matos, J., da Silva Oliveira, J., Bolzon Muniz, G. 2011. Non-destructive estimation of pernambuco (Caesalpinia echinata) clear wood properties using near infrared spectroscopy. Journal of Near Infrared Spectroscopy 19(5): 411～419.

Schwanninger, M., Rodrigues, J.C., Gierlinger, N., Hinterstoisser, B. 2011. Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression. Part 1. Wavenumber-selection and evaluation of the selected range. Journal of Near Infrared Spectroscopy 19(5): 319～329.

Shenk, J.S., Westerhaus, M.O. 1991a. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. Crop science 31(2): 469～474.

Shenk, J.S., Westerhaus, M.O. 1991b. Populations structuring of near infrared spectra and modified partial least squares regression. Crop Science 31(6): 1548～1555.

Shimleck, L.R., Evans, R. 2003. Estimation of air-dry density of increment cores by near infrared spectroscopy. Appita Journal 56(4): 312～317.

Stirling, R., Trung, T., Breuil, C., Bicho, P. 2007. Predicting wood decay and density using NIR spectroscopy. Wood and Fiber Science 39(3): 414～423.

Tsuchikawa, S. 2007. A review of recent near infrared research for wood and paper. Applied Spectroscopy Reviews 42(1): 43～71.

Tsuchikawa, S., 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.

Valdes, E.V., Hunter, R.B., Pinter, L. 1987. Determination of quality parameters by near infrared reflectance spectroscopy in whole-plant corn silage. Canadian journal of plant science 67(3): 747～754.

Via, B.K., So, C.L., Shupe, T.F., Stine, M., Groom, L.H. 2005. Ability of near infrared spectroscopy to monitor air-dry density distribution and variation of wood. Wood and Fiber Science 37(3): 394～402.

Williams, P., Norris, K. 2004. Near-infrared technology in the agricultural and food industries 2nd ed. American Association of Cereal Chemists, Inc., Minnesota. USA.

Williamson, G.B., Wiemann, M.C. 2010. Measuring wood specific gravity… correctly. American Journal of Botany 97(3): 519～524.

Yang, S.Y., Han, Y., Park, J.H., Chung, H., Eom, C.D., Yeo, H. 2015. Moisture Content Prediction Model Development for Major Domestic Wood Species Using Near Infrared Spectroscopy, The Korean Society of Wood Science Technology 43(3): 311～319.

Zhang, S.Y. 1995. Effect of growth rate on wood specific gravity and selected mechanical properties in individual species from distinct wood categories. Wood Science and Technology 29(6): 451～465.