

# 소나무와 금강송의 수종식별을 위한 화학계량학적 접근<sup>1</sup> - 근적외선 분광법과 다변량분석을 이용한 수종 분류 -

황 성 욱<sup>2</sup> · 이 원 희<sup>2,†</sup> · Yoshiki Horikawa<sup>2</sup> · Junji Sugiyama<sup>3,†</sup>

## Chemometrics Approach For Species Identification of *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. *erecta* Uyeki<sup>1</sup> - Species Classification Using Near-Infrared Spectroscopy in combination with Multivariate Analysis -

Sung-Wook Hwang<sup>2</sup> · Won-Hee Lee<sup>2,†</sup> · Yoshiki Horikawa<sup>2</sup> · Junji Sugiyama<sup>3,†</sup>

### 요 약

소나무와 금강송의 수종 분류를 위해 근적외선(NIR) 분광법과 주성분분석(PCA) 및 부분최소자승법 판별분석(PLS-DA)을 결합하여 수종 분류 모델을 설계하였다. 측정된 모든 NIR 스펙트럼을 이용하여 PCA를 실시한 결과 소나무와 금강송의 수종 분류는 불가능하였다. 그러나 2차 미분된 스펙트럼을 이용하여 시험편의 단면과 심·변재 구분에 따른 수종 분류에서는 변재부에서 수종 분류가 가능하였으며, 특히 방사단면의 변재에서는 명확하게 수종이 분류되었다. 그리고 개발된 PLS-DA 예측 모델을 통해 명확한 수종 분류가 가능하였다. 2차 미분으로 전처리된 스펙트럼을 이용하였을 때 가장 좋은 분류 결과 얻을 수 있었다. 2차 미분 스펙트럼을 이용한 예측 모델은 100%의 분류 정확도를 나타내었으며, 예측 모델의  $R_p^2$  값은 0.86, RMSEP는 0.38로 나타났다. 전처리하지 않은 스펙트럼과 2차 미분 스펙트럼을 이용한 예측 모델의 신뢰도는 유사하였다. 근적외선 분광법과 부분최소자승법 판별분석을 결합한 수종 분류 모델은 소나무와 금강송의 분류에 적합하였다.

### ABSTRACT

A model was designed to identify wood species between *Pinus densiflora* for. *erecta* Uyeki and *Pinus densiflora* Sieb. et Zucc. using the near-infrared (NIR) spectroscopy in combination with principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA). In the PCA using all of the spectra, *Pinus densiflora* for.

<sup>1</sup> Date Received June 3, 2015, Date Accepted July 16, 2015

<sup>2</sup> 경북대학교 농업생명과학대학 임산공학과. Department of Wood Science & Technology, College of Agriculture & Life Sciences, Kyungpook National University, Daegu 41566, Republic of Korea

<sup>3</sup> Research Institute for Sustainable Humanosphere, Kyoto University, Uji 611-0011, Japan

<sup>†</sup> 교신저자(Corresponding author): 이원희(e-mail: leewh@knu.ac.kr)

<sup>†</sup> 교신저자(Corresponding author): Sugiyama Junji(e-mail: sugiyama@rsh.kyoto-u.ac.jp)

*erecta* Uyeki and *Pinus densiflora* Sieb. et Zucc. could not be classified. In the PCA using the spectrum that has been measured in sapwood, however, *Pinus densiflora* for. *erecta* Uyeki and *Pinus densiflora* Sieb. et Zucc. could be identified. In particular, it was clearly classified by sapwood in radial section. And more, these two species could be perfectly identified using PLS-DA prediction model. The best performance in species identification was obtained when the second derivative spectra was used; the prediction accuracy was 100%. For prediction model, the  $R_p^2$  value was 0.86 and the RMSEP was 0.38 in second derivative spectra. It was verified that the model designed by NIR spectroscopy with PLS-DA is suitable for species identification between *Pinus densiflora* for. *erecta* Uyeki and *Pinus densiflora* Sieb. et Zucc.

**Keywords :** wood identification, near-infrared spectroscopy, principal component analysis, partial least square discriminant analysis, *Pinus densiflora* for. *erecta* Uyeki, *Pinus densiflora* Sieb. et Zucc.

## 1. 서 론

목재를 연구하는데 있어 수종 식별은 가장 기초적 이면서도 중요한 분야이다. 목재는 수종에 따라 조직 구조가 서로 다르며, 조직구조가 다르면 그 목재의 성질 또한 달라지기 때문에 목재산업 전반에 걸쳐 중요한 역할을 한다. 뿐만 아니라 수종 식별은 목재 문화재 연구분야에서도 기여하는 바가 매우 크다. 가장 일반적인 수종 식별법은 목재 샘플의 절편을 채취하여 해부학적 특징을 현미경으로 관찰하는 현미경 관찰법에 의해 수행된다. 현미경 관찰에 의한 수종 식별은 가장 정확한 방법임에도 불구하고 해부학적 특성이 동일한 유사 수종의 식별은 곤란하다는 한계가 있다. 식별이 곤란한 수종의 일반적인 예로 국내산 소나무와 금강송을 들 수 있다. 소나무는 대표적인 국산 침엽수종으로서 한반도 전역에 고루 분포하며, 금강송은 국산 소나무의 지역품종으로서 금강산으로부터 태백산맥을 따라 강원도와 경북 북부 지방에 자생하는 소나무 중 수간이 곧고 형질이 좋은 소나무를 말한다. 예로부터 금강송은 궁궐의 건축재 및 관재 등 특수 목적으로 사용되었던 소중한 목재 자원이다. 그러나 일반적인 소나무(육송)와 금강송은 해부학적 특성이 동일하여 전통적인 현미경 관찰법으로는 수종 식별이 불가능하다.

수종 식별이 곤란한 유사수종을 구분하기 위해 화학계량학(chemometrics)을 이용한 연구가 진행되고 있다. 화학계량학은 다양한 분석화학 데이터에서 원하는 정보를 도출하여 분석을 가능하게 하는 통계적

알고리즘을 총칭하는 것으로써 주로 다변량분석(multivariate analysis) 기법이 이용된다. 근적외선(NIR) 분광법을 다변량분석(multivariate analysis)과 결합하여 이용할 경우 목재의 수종 분류가 가능할 수 있다. 다변량분석은 단일변량분석과 대조되는 개념으로 여러 변수의 관계성을 동시에 고려하여 분석하는 통계적 기법이다(Yang, 2006). NIR 분광법과 다변량분석을 이용한 수종 분류 과정은 먼저 NIR 스펙트럼을 수집한 후 스펙트럼 전처리를 실시하고 주성분분석(PCA, principal component analysis), 부분최소자승법(PLS, partial least square), SIMCA (soft independent modelling of class analogy) 등 용도에 적합한 다변량분석 기법을 이용하여 수종 분류를 실시하게 된다. 다변량분석 기법 중 PCA가 X 공간상의 분산을 최대로 설명할 수 있는 성분을 구하는 것이라면, PLS는 X와 Y 공간상의 분산을 설명하는 성분을 도출할 때 종속변수 Y와의 상관관계를 고려하여 가중치를 계산하는 반복과정을 통해 잠재변수를 도출하게 된다는 것이 가장 큰 차이점이다. 데이터의 규모가 큰 경우에는 PLS가 PCA보다 예측력이 높고 노이즈 제거 효과도 우수하다(Lee *et al.*, 2014).

국내에서는 화학계량학적 방법을 이용하여 목재의 물리적 성질을 예측하기 위한 모델 개발 연구가 수행된 바 있다. Eom *et al.* (2010)은 수학적 전처리를 실시한 NIR 스펙트럼과 회귀분석을 결합하여 목재의 표면함수율 예측이 가능하다는 것을 확인하였으며, Chang *et al.* (2015)은 낙엽송 목분의 함수율 예측에 NIR 분광법을 적용하여 신뢰성 높은 예측 모델

을 개발하였다. 그리고 Yang *et al.* (2015)은 NIR 분광법과 PLS 회귀(Partial least squares regression)를 결합하여 섬유평화점 이하에서 국산 침엽수 및 백합나무의 함수율 예측모델을 개발하였다.

화학계량학적 방법을 이용한 수종분류에 대한 연구는 주로 국외에서 진행되어 왔다. Lewis *et al.* (1994)은 NIR-FTR 분광법과 뉴럴 컴퓨팅(neural computing)을 결합하여 목재의 수종 분류를 연구하였으며, Antti *et al.* (1996)은 NIR 분광법과 PLS 모델을 통해 3가지 수종(Swedish pine, Swedish spruce, Polish pine)이 혼재된 목재칩의 분류 가능성을 연구하였다. 그리고 Brunner *et al.* (1996)은 NIR 분광법과 군집분석(cluster analysis)을 이용하여 12가지 수종을 구분하는 것이 가능하다고 보고하였다. Schimleck *et al.* (2000)은 유칼리나무의 NIR 스펙트럼과 PCA를 이용하여 수종 및 원산지 판별에 대한 연구를 소개하였다. Tsuchikawa *et al.* (2003a; 2003b)은 NIR 스펙트럼과 마할라노비스의 범거리(Mahalanobis's generalized distance)를 결합하여 목재의 비파괴적 수종 판별법을 보고한 바 있다. Gierlinger *et al.* (2004)은 FT-NIR 스펙트럼 측정과 군집분석의 계통도(dendrogram)를 결합한 PCA를 통해 잎갈나무속 수종의 분류를 시도하였으며, SIMCA(soft independent modelling of class analogy)를 이용하여 수종의 분류 및 차이를 연구하였다. Watanabe *et al.* (2011)은 PLS를 이용하여 편백, 삼나무, 비자나무, 측백나무 등 고목재의 수종 분류를 성공하였다. 그리고 Pastore *et al.* (2011)은 NIR 분광법과 PLS 판별분석을 통해 마호가니 유사수종으로부터 진품 마호가니를 구분하는 연구를 실시하였으며, Horikawa *et al.* (2015)은 다변량분석을 이용하여 육송과 해송을 구분하였다. 이와 같이 목재의 수종 분류를 위해 분광분석과 다변량분석이 결합된 다양한 화학계량학적 접근이 이루어지고 있다.

본 연구에서는 전통적인 수종 식별법인 현미경 관찰법의 한계점을 보완하고 수종 식별을 위한 화학계량학적 접근의 가능성을 확인하기 위해 근적외선(NIR) 분광법과 주성분분석(PCA, principal component analysis) 및 부분최소자승법 판별분석(PLS-DA,

partial least square discriminant analysis)을 결합하여 유사수종인 소나무와 금강송의 수종 분류를 시도하였다. 그리고 소나무와 금강송의 수종 판별에 영향을 미치는 인자에 대해 고찰하였다.

## 2. 재료 및 방법

### 2.1. 공시재료

NIR 분광법과 다변량분석을 이용한 수종 분류에 관한 연구를 위하여 소나무(*Pinus densiflora* Sieb. et Zucc.)와 금강송(*Pinus densiflora* for. *erecta* Uyeki)의 목편이 이용되었다. 두 수종 모두 경상북도 울진군에서 채취되었다. 금강송은 울진군 소광리에서 채취되었으며, 소나무는 울진 남부지역에 위치한 영덕군과의 경계지역에서 채취되었다. 수종당 각 5개 시험편이 사용되었으며, 모든 시험편은 각기 다른 개체에서 채취되었다.

### 2.2. 근적외선(NIR) 분광분석

각 시험편의 심재와 변재, 횡단면과 방사단면의 NIR 스펙트럼을 측정하였다. 각 시험편의 측정부위(횡단면-변재, 횡단면-심재, 방사단면-변재, 방사단면-심재)당 12회 측정하여 총 480개의 NIR 스펙트럼 데이터를 수집하였다. NIR 스펙트럼은 FT-NIR(PerkinElmer Spectrum 100N)을 이용하여 분해능 16  $\text{cm}^{-1}$ 로 각 측정점당 32회 반복 스캔하였다. 측정된 스펙트럼은 Savitzky-Golay 알고리즘(Savitzky and Golay, 1964)에 의해 2차 미분으로 전처리하였다. 2차 미분된 스펙트럼과 전처리를 하지 않은 스펙트럼을 이용하여 PCA와 PLS-DA를 실시하였으며, 스펙트럼 전처리에 따른 수종 분류 결과를 비교하였다.

### 2.3. 다변량 분석

PCA와 PLS-DA는 소프트웨어 Unscrambler v.9.8(CAMO Software, Inc., Woodbridge, NJ)을 이용하여 NIPALS (nonlinear iterative partial least square) 알

고리층을 기반으로 수행되었으며, NIR 스펙트럼의 파장영역 8,000 - 4,000  $\text{cm}^{-1}$  범위에서 실시되었다.

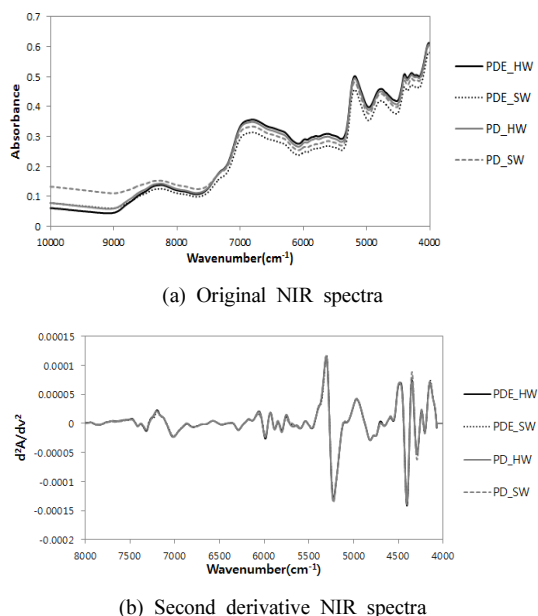
PCA는 변수와 개체의 수가 많은 거대한 데이터를 분석하기 위한 방법으로 데이터의 축약 및 단순화, 모델링, 이상치 탐지, 변수 선택, 분류, 예측, 혼합 해제 등을 목적으로 한다(Wold *et al.*, 1987). PCA를 이용하여 거대하고 복잡한 상호 의존적인 변수들을 서로 독립적인 새로운 변수로 만들어 데이터를 간소화할 수 있으며, 이 과정에서 새롭게 생성된 독립적인 변수들을 주성분(PC, principal component)이라 한다.

PLS-DA 모델의 개발 과정은 먼저 소나무와 금강송의 NIR 스펙트럼을 각 수종별 그룹으로 분류하여 소나무 그룹의 스펙트럼에는 임의의 가상 변수값을 '-1'로 지정하고 금강송 그룹의 스펙트럼에는 가상 변수값 '1'을 종속변수로 지정하였다. 각 수종별 NIR 스펙트럼 240개 중 160개의 스펙트럼을 'calibration set'으로 지정하였으며, 나머지 80개의 스펙트럼은 'prediction set'으로 지정하였다. 즉 calibration set과 prediction set의 분류는 2 : 1 비율로 무작위(random) 분류되었다. 총 480개의 스펙트럼 중 160개의 prediction set을 PLS-DA 모델에 적용하여 소나무와 금강송의 판별을 실시하였다. 소나무와 금강송의 판별은 가상 변수값 1과 -1의 중간값 0을 기준으로 예측값이 0보다 작으면 소나무, 0보다 크면 금강송으로 분류하는 방법으로 실시되었다. PLS-DA 모델의 calibration 성능을 평가하기 위해 결정계수( $R_c^2$ , coefficient of determination for calibration)와 평균제곱근오차(RMSEC, root mean square error of calibration)를 이용하였다. 그리고 개발된 모델은 prediction의 결정계수( $R_p^2$ )와 평균제곱근오차(RMSEP)를 이용하여 평가하였다. 예측 정확도는 prediction set의 총 샘플 개수에서 올바르게 수종 판별된 샘플 수의 비율로 나타내었다.

### 3. 결과 및 고찰

#### 3.1. 소나무와 금강송의 NIR 스펙트럼 분석

Fig. 1에 소나무와 금강송의 NIR 스펙트럼을 나타



Notes) PDE, *Pinus densiflora* for. *erecta* Uyeki; PD, *Pinus densiflora* Sieb. et Zucc.; CS, cross section; RS, radial section; HW, heartwood; SW, sapwood.

**Fig. 1.** Original and second derivative NIR spectra of *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. *erecta* Uyeki.

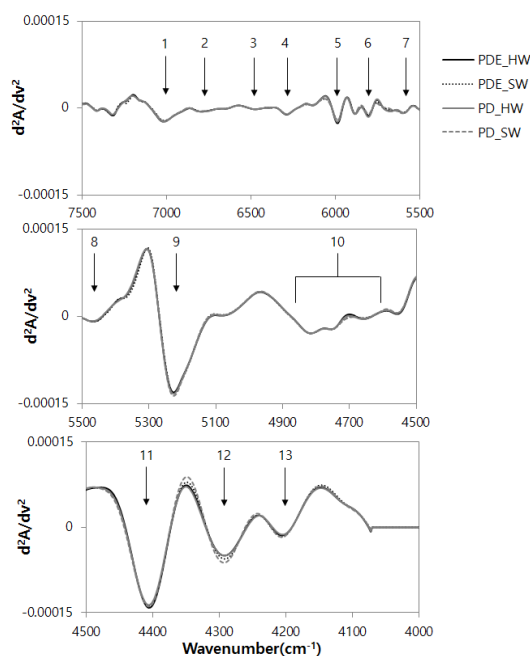
내었다. Fig. 1 (a)는 전처리를 실시하지 않은 스펙트럼, Fig. 1 (b)는 2차 미분으로 전처리한 스펙트럼이다. 목재와 같은 고체의 NIR 측정은 스펙트럼의 재현성이 좋지 않기 때문에 분석결과를 향상시키기 위해 전처리 과정이 필요하다. Fig. 1 (b)의 2차 미분으로 전처리된 소나무와 금강송의 NIR 스펙트럼은 거의 동일한 패턴을 나타내어 스펙트럼 분석만으로는 수종 분류가 불가능하였다.

Fig. 2에 소나무와 금강송의 2차 미분 스펙트럼과 목재의 주요 성분들의 스펙트럼 영역을 나타내었으며, 각 스펙트럼 영역에 상응하는 목재 성분들을 Table 1에 나타내었다. Fig. 2에서 볼 수 있듯이 셀룰로오스를 나타내는 대부분의 스펙트럼 밴드 [7,000 (peak 1 in Fig. 2; Tsuchikawa *et al.*, 2003b); 6,718 (peak 2; Tsuchikawa *et al.*, 2003b); 6,450 (peak 3; Tsuchikawa *et al.*, 2003b); 6,287 (peak 4;

**Table 1.** Absorption bands and corresponding vibration modes for wood in NIR wavelength region (Tsuchikawa *et al.*, 2005)

| Peak | Wavenumber (cm <sup>-1</sup> ) | Bond vibration          | Remarks                                  |
|------|--------------------------------|-------------------------|------------------------------------------|
| 1    | 7,000                          | OH str. first overtone  | Amorphous regions in cellulose           |
| 2    | 6,718                          | OH str. first overtone  | Semi-crystalline regions in cellulose    |
| 3    | 6,450                          | OH str. first overtone  | Crystalline regions in cellulose         |
| 4    | 6,287                          | OH str. first overtone  | Crystalline regions in cellulose         |
| 5    | 5,980                          | CH str. first overtone  | Aromatic skeletal due to lignin          |
| 6    | 5,800                          | CH str. first overtone  | Furanose/pyranose due to hemicellulose   |
| 7    | 5,587                          | CH str. first overtone  | Semi- or crystalline region in cellulose |
| 8    | 5,464                          | OH str. + 2 × CO str.   | Semi- or crystalline region in cellulose |
| 9    | 5,219                          | OH str. + OH def.       | Water                                    |
| 10   | 4,890-4,620                    | OH str. + CH def.       | Cellulose                                |
| 11   | 4,404                          | SH str. first overtone  | Cellulose and hemicellulose*             |
| 12   | 4,280                          | CH str. + CH def.       | Semi- or crystalline region in cellulose |
| 13   | 4,198                          | CH def. second overtone | Holocellulose                            |

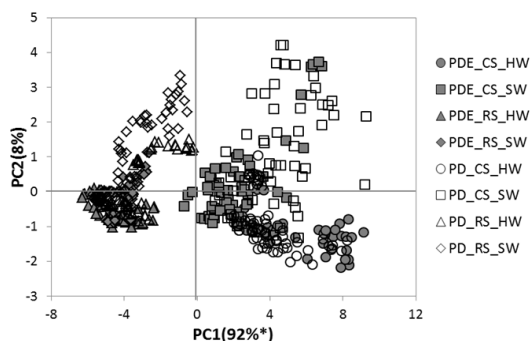
Notes) str., stretching; def., deformation; \*, cited from Schwanninger *et al.*, (2011).



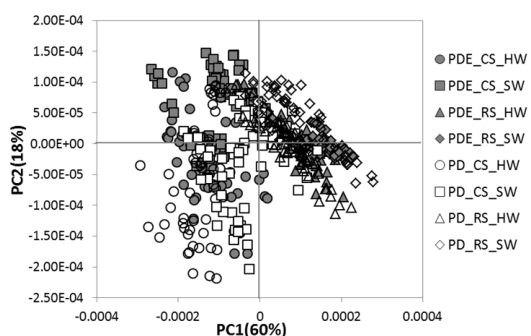
Notes) PDE, *Pinus densiflora* for. *erecta* Uyeki; PD, *Pinus densiflora* Sieb. et Zucc.; CS, cross section; RS, radial section; HW, heartwood; SW, sapwood.

**Fig. 2.** Second derivative NIR spectra of *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. *erecta* Uyeki (Tsuchikawa *et al.*, 2005).

Tsuchikawa *et al.*, 2003b); 5,587 (peak 7; Siesler *et al.*, 2008); 5,464 (peak 8; Siesler *et al.*, 2008); 4,890-4,620 (peak 10; Osborne *et al.*, 1993); 4,198 cm<sup>-1</sup> (peak 13; Hitoshi and Tsuchikawa, 2003)]는 동일한 패턴을 나타내었다. 그러나 셀룰로오스의 반결정 또는 결정영역을 나타내는 4,280 cm<sup>-1</sup> (peak 12; Osborne *et al.*, 1993) 밴드에서는 심재의 경우 소나무와 금강송이 동일하였으나, 변재에서는 소나무의 피크가 금강송에 비해 강하게 나타났다. 헤미셀룰로오스[5,800 cm<sup>-1</sup> (peak 6; Siesler *et al.*, 2008)]와 리그닌[5,980 cm<sup>-1</sup> (peak 5; Siesler *et al.*, 2008)]은 두 수종 사이에 별 다른 차이가 없었다. 그리고 수분을 나타내는 5,220 cm<sup>-1</sup> (peak 9; Siesler *et al.*, 2008) 밴드에서는 두 수종 모두 변재의 함수율이 다소 높게 나타났다. NIR 분광법과 다변량분석을 이용하여 소나무와 해송의 수종 분류 연구를 실시한 Horikawa *et al.* (2015)의 연구에서도 육송과 해송의 NIR 스펙트럼이 동일한 패턴을 나타내어 스펙트럼 분석만으로는 해당 수종의 구분이 곤란하다고 보고한 바 있다.



(a) Score plot using the original spectra



(b) Score plot using the second derivative spectra

Notes) PDE, *Pinus densiflora* for. *erecta* Uyeki; PD, *Pinus densiflora* Sieb. et Zucc.; CS, cross section; RS, radial section; HW, heartwood; SW, sapwood; \*, explained variance.

**Fig. 3.** PCA score plot of *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. *erecta* Uyeki on the PC1 and PC2.

### 3.2. 주성분분석(PCA)

#### 3.2.1. PCA를 이용한 소나무와 금강송의 수종 분류

소나무와 금강송의 수종 분류를 위해 각 수종별 모든 스펙트럼 데이터를 이용하여 PCA를 실시하였다(Fig. 3). Fig. 3의 PCA 스코어 플롯에서 볼 수 있듯이 소나무와 금강송은 구분되지 않았으며, 심·변재간 구분도 불가능하였다. NIR 분광법은 관능기에 민감하게 반응하기 때문에 수목의 화학적, 구조적 특성에 크게 영향을 받는다. 따라서 근본적으로 주요

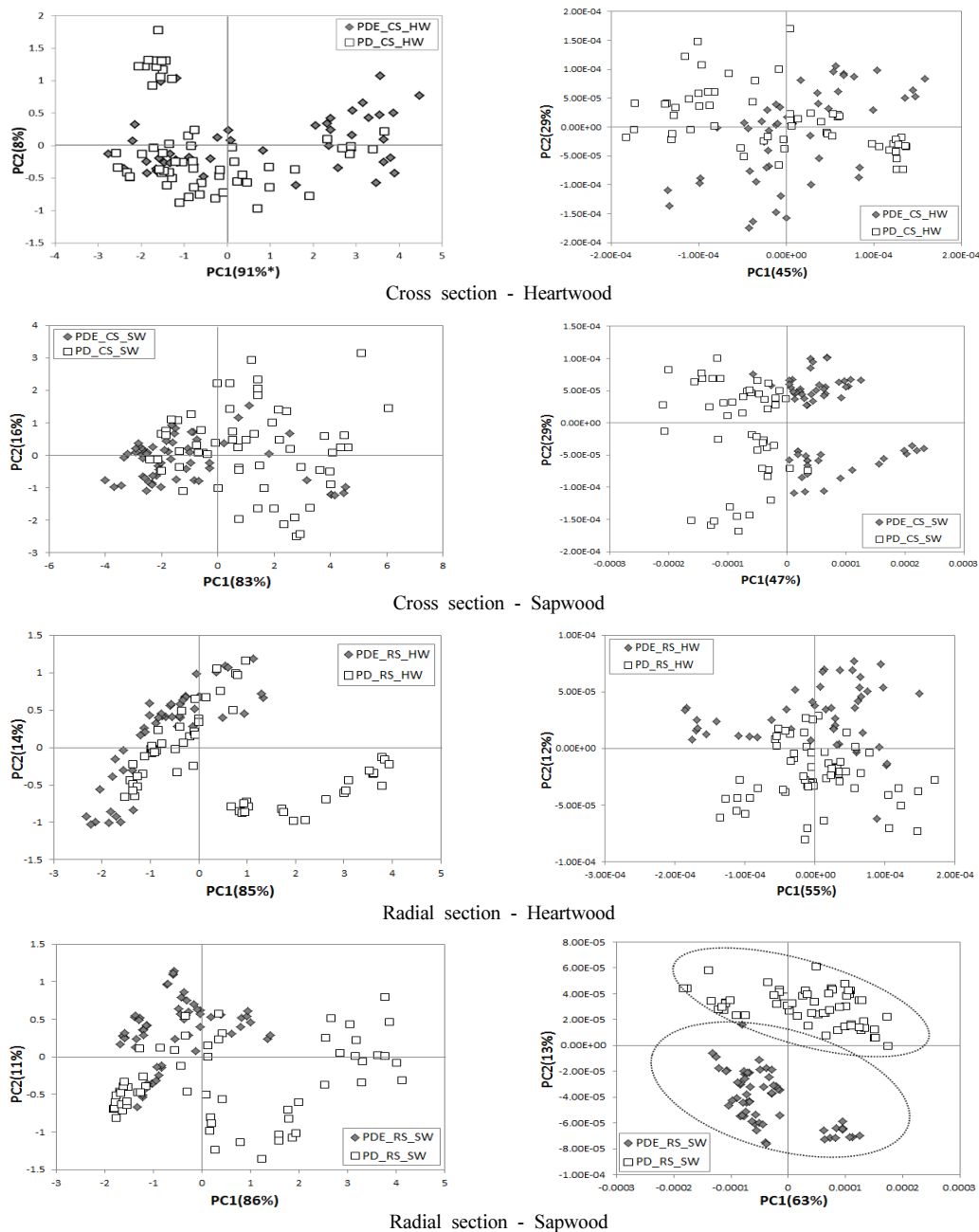
성분들이 화학적으로 다르지 않은 소나무와 금강송의 수종 분류는 쉽지 않다. 그러나 각 수종별로 분류되는 것은 아니지만 Fig. 3 (a)의 전처리하지 않은 스펙트럼의 스코어 플롯에서 PC1의 값 0을 기준으로 두 개의 그룹으로 구분되는 것을 볼 수 있다. 이것은 시험편의 스펙트럼 측정 단면에 따라 구분된 것으로 PC1의 값 0을 기준으로 좌측은 방사단면, 우측은 횡단면의 측정값이다. Fig. 3 (b)의 2차 미분으로 전처리된 스펙트럼의 경우에는 전처리하지 않은 스펙트럼을 이용한 스코어 플롯만큼 명확하지는 않지만 대략적으로 PC1의 값 0을 기준으로 좌측은 횡단면, 우측은 방사단면으로 구분되는 것을 확인할 수 있었다. Fig. 3의 스코어 플롯에서 각 주성분과 함께 표기되어 있는 백분율 값을 설명할 수 있는 분산(explained variance)을 나타낸다. 이 값은 PCA 모델의 총 분산 중에서 해당 주성분이 설명하는 분산의 양을 의미한다.

그리고 전처리하지 않은 스펙트럼과 2차 미분 스펙트럼 모두에서 횡단면의 스코어 플롯이 방사단면보다 높은 분산을 나타내었다. 이처럼 시험편의 측정 단면에 따라 구분이 되는 것은 목재의 횡단면과 방사단면의 구조적인 차이에 기인한다고 판단된다. NIR 분광법은 시료에 조사된 근적외선의 반사를 통해 시료의 화학적 정보를 나타내게 되는데, 횡단면의 경우 조사된 근적외선이 세포내강 깊숙이 침투된 후 반사되어 방사단면과는 다소 상이한 스펙트럼 특성을 나타낸 것으로 추정된다. 이러한 횡단면의 구조적 특성상 스펙트럼의 재현성은 낮아질 것이며, 그 결과 횡단면의 스코어 플롯이 방사단면보다 높은 분산을 나타내는 것이라 판단된다.

#### 3.2.2. 심재와 변재의 구분에 의한 수종 분류

Fig. 4에 소나무와 금강송의 각 단면별 심재와 변재에서 전처리하지 않은 스펙트럼과 2차 미분으로 전처리된 스펙트럼의 PCA 결과를 나타내었다. 각 단면에서 심재와 변재의 PCA를 살펴보면, 전처리하지 않은 스펙트럼의 경우 각 단면의 심재와 변재 모두에서 소나무와 금강송의 플롯이 혼재되어 수종 분류가 불가능하였다.

소나무와 금강송의 수종식별을 위한 화학계량학적 접근



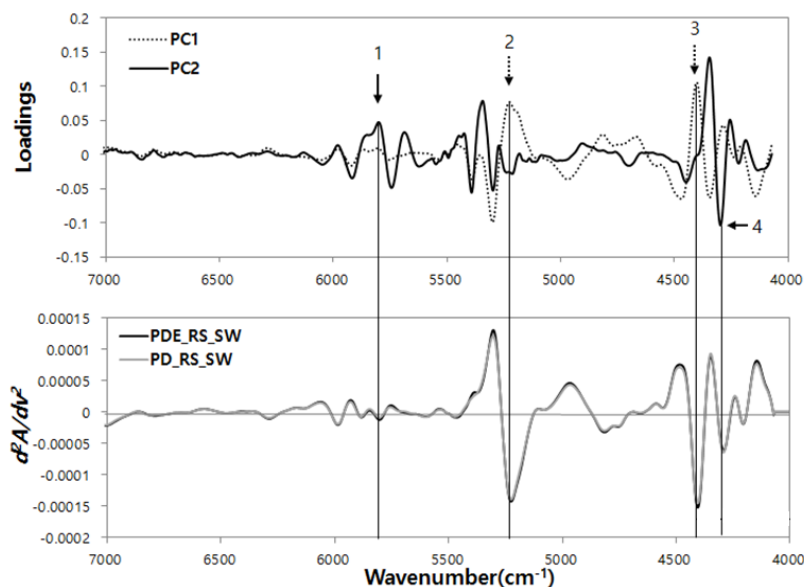
Notes) PDE, *Pinus densiflora* for. *erecta* Uyeki; PD, *Pinus densiflora* Sieb. et Zucc.; CS, cross section; RS, radial section; HW, heartwood; SW, sapwood; \*, explained variance.

**Fig. 4.** PCA score plot of *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. *erecta* Uyeki by measuring position of NIR spectra.

**Table 2.** Assignment of representative NIR absorption bands in Fig. 5

| Peak | Wavenumber (cm <sup>-1</sup> ) | Bond vibration          | Remarks                                  |
|------|--------------------------------|-------------------------|------------------------------------------|
| 1    | 5,800                          | CH str. first overtone  | Furanose/pyranose due to hemicellulose   |
| 2    | 5,220                          | OH str. + OH def.       | Water                                    |
| 3    | 4,404                          | SH str. first overtone  | Cellulose and hemicellulose              |
| 4    | 4,280                          | CH def. second overtone | Semi- or crystalline region in cellulose |

Notes) str., stretching; def., deformation



**Fig. 5.** Comparison between the spectrum obtained from PC1 and PC2 in PCA and the second derivative spectra obtained from *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. erecta Uyeki.

그러나 2차 미분으로 전처리된 스펙트럼을 이용한 경우 횡단면-심재에서는 플롯이 혼재되어 분류가 되지 않았지만 횡단면-변재, 방사단면-심재, 방사단면-변재에서는 수종별 스코어 플롯의 군집된 형태를 확인할 수 있었다. 횡단면-변재의 경우 PC1의 값 0을 기준으로 좌측에는 소나무, 우측에는 금강송의 플롯이 분포되어 수종별 분류가 가능하였다. 방사단면-심재의 경우 일부 플롯이 다소 혼재되어 있지만 대략적으로 PC2의 0을 기준으로 위쪽에는 금강송, 아래쪽에는 소나무가 분포하였다. 가장 뛰어난 수종 분류 결과를 나타낸 것은 방사단면-변재였다. Fig. 4 (b)의

방사단면-변재의 PCA 결과에서 볼 수 있듯이 PC2의 값 0을 기준으로 위쪽에는 소나무, 아래쪽에는 금강송이 위치하여 거의 완벽하게 수종별로 군집이 형성되는 것을 확인할 수 있었다.

가장 뛰어난 수종 분류 결과를 나타낸 방사단면-변재에서 어떤 요인에 의해 수종이 분류되었는지를 확인하기 위해 PC의 적재(loading)값과 2차 미분 스펙트럼을 비교하였다(Fig. 5). 적재값은 PCA를 통해 생성된 새로운 변수 PC에 대해서 기존의 변수(스펙트럼)가 얼마나 기여하는지를 나타낸다. 2차 미분으로 전처리된 횡단면-변재의 경우 PC1의 값에 따라



좌우로 각각 소나무와 금강송이 분류되었기 때문에 PC1의 적재값과 2차 미분 스펙트럼을 비교하였다. 그리고 가장 명확한 분류 결과를 나타낸 방사단면-변재의 경우 PC2의 값을 따라 위, 아래로 각각 수종이 분류되어 PC2의 적재값을 비교하였다.

PC1에서 높은 적재값을 나타낸 부분은  $5,220\text{ cm}^{-1}$  (peak 2 in Fig. 5)와  $4,404\text{ cm}^{-1}$  (peak 3) 영역이었다. 가장 높은 적재값을 나타낸 영역은  $4,404\text{ cm}^{-1}$ 로써 이 영역은 셀룰로오스와 헤미셀룰로오스로 할당되어 있다(Schwanninger *et al.*, 2011). 그리고 높은 적재값을 나타낸 또 다른 영역인  $5,220\text{ cm}^{-1}$ 는 수분을 나타내는 영역이다. 그러나 2차 미분 스펙트럼에서 소나무와 금강송의 수분 피크는 뚜렷한 차이를 나타내지 않았다.

PC2에서 목재의 주요 구성성분을 나타내는 스펙트럼 영역 중 높은 적재값을 나타낸 부분은  $5,800\text{ cm}^{-1}$  (peak 1)와  $4,280\text{ cm}^{-1}$  (peak 4) 영역이었다. NIR 스펙트럼에서  $5,800\text{ cm}^{-1}$ 은 헤미셀룰로오스 성분을 나타낸다(Siesler *et al.*, 2008). 그러나 이 영역에서 소나무와 금강송의 2차 미분된 NIR 스펙트럼은 뚜렷한 차이 없이 거의 동일한 패턴을 나타내었다. 그리고  $4,280\text{ cm}^{-1}$  영역은 PC2에서 가장 높은 적재값을 나타내었는데, 이 영역은 목재 성분 중 셀룰로오스의 반결정 또는 결정영역으로 할당되어 있다(Osborne *et al.*, 1993). Fig. 2 (peak 11)의 2차 미분 스펙트럼에서 볼 수 있듯이,  $4,280\text{ cm}^{-1}$  영역에서는 심재의 경우 소나무와 금강송의 차이가 없었으나 변재에서는 소나무의 피크가 금강송보다 강하게 나타났다. 따라서 소나무와 금강송의 변재부에 존재하는 셀룰로오스 반결정 또는 결정영역이 두 수종을 분류하는데 어느 정도 영향을 미친다는 것을 확인할 수 있었다. 이 영역을 제외한 다른 주요 성분들의 스펙트럼 영역에서는 뚜렷한 적재값을 나타내지 않았으며, NIR 스펙트럼의 차이도 없었다.

### 3.3. 부분최소자승법 판별분석(PLS-DA)

#### 3.3.1. PLS-DA를 이용한 수종 분류

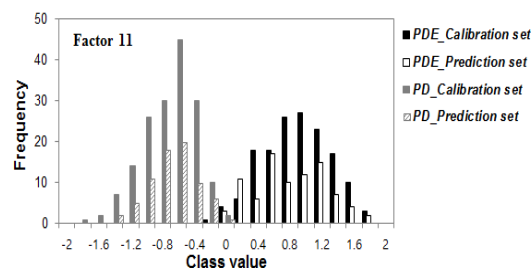
PLS-DA 모델을 이용한 소나무와 금강송의 수종

분류 결과를 Table 3과 Fig. 6에 나타내었다. PLS-DA 모델에서 소나무와 금강송의 수종 분류를 위한 최적의 factor를 결정하기 위해 잔차 분산(residual variance)을 검토하였다. 잔차 분산은 설명할 수 있는 분산(explained variance)과 대조되는 개념으로서 factor의 수가 증가할수록 설명할 수 있는 분산의 값은 증가하나 잔차 분산의 값은 감소한다. 전처리하지 않은 스펙트럼과 2차 미분된 스펙트럼 각각의 calibration set에서 잔차 분산 값의 감소폭이 크게 줄어드는 지점, 즉 잔차 분산의 감소가 안정화되는 요인(factor component)의 수를 최적의 factor로 선정하였다. Factor의 수가 높아질수록 수종 분류 모델의 신뢰도를 나타내는  $R^2$  값은 높아지고 RMSE 값은 낮아지지만, factor가 지나치게 높을 경우 예측 모델이 불필요한 정보까지 학습하게 되어 과적합(overfitting)을 초래할 수 있다. 따라서 수종 분류를 위한 적합한 factor를 선정하는 것이 중요하다. 이러한 과정을 통해 전처리하지 않은 스펙트럼의 경우 factor 11, 2차 미분된 스펙트럼은 factor 7을 최적의 요인으로 결정하였다. NIR 스펙트럼의 2차 미분을 통해 factor의 수는 상대적으로 감소하였다. Factor는 PLS-DA 모델 개발 과정에서 새롭게 생성된 독립 변수로서 PCA의 주성분과 의미가 유사하지만, factor의 계산 과정에서는 종속변수로서 각 수종의 가상 변수값(-1, +1)을 함께 고려한다는 점에서 주성분과는 다소 다른 개념이라 할 수 있다.

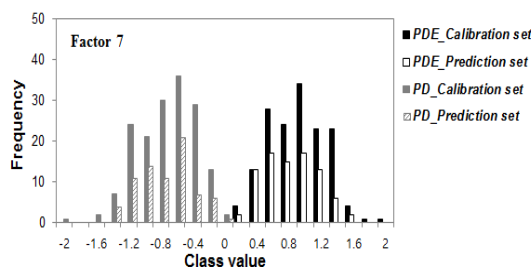
Table 3에서 볼 수 있듯이 2차 미분 스펙트럼을 이용한 수종 분류 결과 factor 7에서 소나무와 금강송의 분류 정확도는 100%로 나타났다. 그러나 전처리하지 않은 스펙트럼을 이용한 경우에는 일부 금강송 샘플의 분류값이 0 이하의 값을 나타내어 소나무로 판별되는 오류를 나타내었다. Calibration 모델에서 전처리하지 않은 스펙트럼을 이용하였을 경우(factor 11) 결정계수( $R_c^2$ ) 값은 0.83, 평균제곱근오차(RMSEC) 값은 0.41로 나타났으며, 2차 미분 스펙트럼을 이용한 경우(factor 7)에는  $R_c^2$  값이 0.84, RMSEC 값은 0.39로 나타나 스펙트럼 전처리 유무에 관계없이 선택된 모델은 유사한 신뢰도를 나타내었다. 그리고 prediction 모델에서는 전처리하지 않은

**Table 3.** Statistical summary of PLS-DA models

| Spectra Pretreatment      | PLS factors | Calibration set |       | Prediction set |       | Correct prediction (%) |
|---------------------------|-------------|-----------------|-------|----------------|-------|------------------------|
|                           |             | $R_c^2$         | RMSEC | $R_p^2$        | RMSEP |                        |
| Original spectra          | 11          | 0.83            | 0.41  | 0.80           | 0.45  | 97.5                   |
| Second derivative spectra | 7           | 0.84            | 0.39  | 0.86           | 0.38  | 100.0                  |



(a) Original spectra



(b) second derivative spectra

Notes) PDE, *Pinus densiflora* for. *erecta* Uyeki; PD, *Pinus densiflora* Sieb. et Zucc.

**Fig. 6.** Histogram of class value computed by PLS-DA on the basis of the original (a) and second derivative spectra (b).

스펙트럼의  $R_p^2$  값은 0.80, RMSEP 값은 0.45로 나타났다. 2차 미분 스펙트럼의  $R_p^2$  값은 0.86, RMSEP 값은 0.38로 나타났다. Prediction 모델에서는 2차 미분으로 전처리된 스펙트럼을 이용한 모델이 전처리하지 않은 스펙트럼을 이용한 모델보다 신뢰도가 다소 높게 나타났다.

Fig. 6에는 전처리하지 않은 스펙트럼과 2차 미분 스펙트럼을 이용한 소나무와 금강송의 수종 분류 결과를 히스토그램으로 나타내었다. Fig. 6 (a)의 전처리하지 않은 스펙트럼의 히스토그램에서는 금강송의

일부가 소나무의 영역에 분포하는 것을 확인할 수 있으며, Fig. 6 (b)의 2차 미분 스펙트럼의 히스토그램에서는 분류값(class value) 0을 기준으로 음의 값 영역에는 소나무, 그리고 양의 값 영역에는 금강송이 위치하여 모든 샘플이 정확히 분류된다는 것을 확인할 수 있었다.

이상의 결과를 통해 수종 분류를 위한 PLS-DA 모델의 개발에 있어 2차 미분을 통한 스펙트럼의 전처리는 분류 정확도를 향상시킨다는 것을 확인할 수 있었다. 그리고 본 연구에서 개발된 PLS-DA 수종 분류 모델을 통해 소나무와 금강송을 정확히 판별할 수 있었다.

### 3.3.2. PLS-DA 모델의 수종 분류 영향 인자

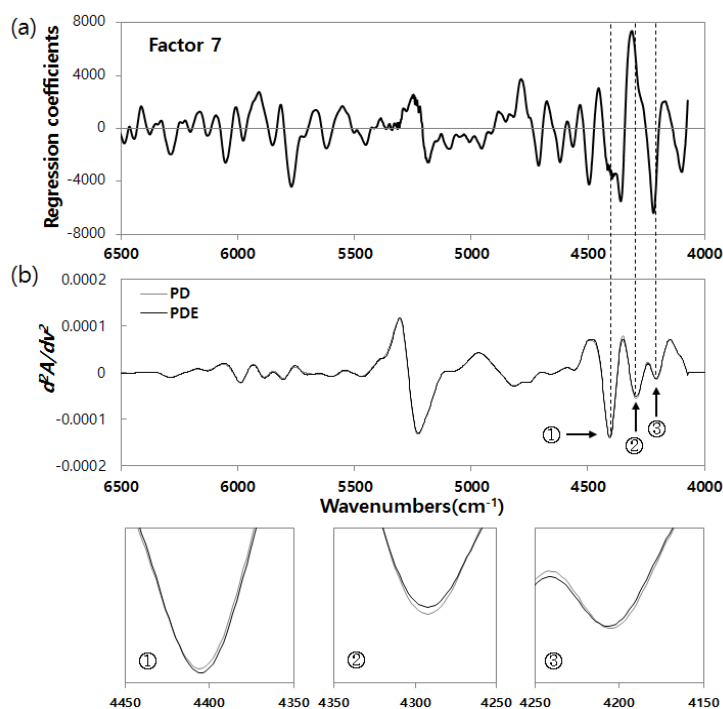
PLS-DA 모델을 통해 소나무와 금강송의 수종을 명확히 구분할 수 있었다. 특히 2차 미분 스펙트럼을 이용했을 경우 100%의 정확도로 수종 분류가 가능하였다. 이와 같이 소나무와 금강송의 수종 분류를 가능하게 한 영향 인자를 찾기 위해 2차 미분 스펙트럼을 이용한 분류 모델의 factor 7에 대한 회귀계수와 2차 미분 NIR 스펙트럼을 비교하였다. 회귀계수는 PLS-DA 모델에서 예측 변수에 주어지는 비중치를 의미한다. 회귀계수의 크기는 선택된 모델의 예측에 있어 얼마만큼의 영향을 미치는지에 대한 정도를 나타낸다.

Fig. 7 (a)의 회귀계수 스펙트럼에서 높은 값을 나타내는 영역들은  $4,404\text{ cm}^{-1}$ ,  $4,280\text{ cm}^{-1}$ ,  $4,198\text{ cm}^{-1}$  밴드이다. 소나무와 금강송의 분류에 큰 영향을 미치는 스펙트럼 영역은 주로  $4,500 - 4,100\text{ cm}^{-1}$  사이의 영역이었다. 이 영역들 이외에 Table 1에서 언급되었던 목재 주요 성분들의 스펙트럼 영역에서는 회귀계수 값이 높지 않았다. 근본적으로 소나무와 금강송의

**Table 4.** Assignment of representative NIR absorption bands in Fig. 7

| Peak | Wavenumber (cm <sup>-1</sup> ) | Bond vibration          | Remarks                                  |
|------|--------------------------------|-------------------------|------------------------------------------|
| 1    | 4,404                          | SH str. first overtone  | Cellulose and hemicellulose              |
| 2    | 4,280                          | CH def. second overtone | Semi- or crystalline region in cellulose |
| 3    | 4,198                          | CH def. second overtone | Holocellulose                            |

Notes) str., stretching; def., deformation



Notes) PDE, *Pinus densiflora* for. *erecta* Uyeki; PD, *Pinus densiflora* Sieb. et Zucc.

**Fig. 7.** Comparison between the spectrum of regression coefficients obtained from factor 7 in PLS-DA (a) and second derivative spectra obtained from *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for. *erecta* Uyeki (b).

주요 성분들은 동일하기 때문에 셀룰로오스, 헤미셀룰로오스, 리그닌과 같은 주요 성분들에 의한 수종 분류는 곤란하며, 그 결과 PLS-DA 모델에서도 주요 성분들의 회귀계수는 낮은 값을 나타내었다고 판단된다. 수분을 나타내는 5,220 cm<sup>-1</sup> 밴드의 회귀계수가 높지 않은 것으로 보아 소나무와 금강송의 분류

에 있어 함수율은 큰 영향을 미치지 않는 것으로 판단된다. 수종 분류에 있어 함수율 차이가 미치는 영향이 크지 않다는 것은 Tsuchikawa *et al.* (2003a)의 연구에서도 확인할 수 있다. 이들은 다양한 함수율 범위의 샘플을 이용한 수종 분류 연구를 실시하여 함수율 차이와 상관없이 수종 분류가 가능하다고 보

고하였다.

가장 높은 회귀계수를 나타낸 스펙트럼 영역인  $4,280\text{ cm}^{-1}$  (peak 2 in Fig. 7b)와  $4,198\text{ cm}^{-1}$  (peak 3) 밴드는 각각 셀룰로오스의 반결정 또는 결정영역 (Osborne *et al.*, 1993)과 홀로셀룰로오스(Hitoshi and Tsuchikawa, 2003)를 나타낸다. Fig. 7 (b)의 2차 미분 스펙트럼에서  $4,280\text{ cm}^{-1}$  밴드는 소나무의 피크가 금강송 보다 다소 크게 나타났으며,  $4,198\text{ cm}^{-1}$  밴드는 두 수종간의 차이가 거의 없었다. 이 중에서 셀룰로오스의 반결정 또는 결정영역을 나타내는  $4,280\text{ cm}^{-1}$  밴드는 PCA를 이용한 수종 분류에서도 소나무와 금강송을 구분하는데 영향을 미친 영역이었다. 그리고 비교적 높은 회귀계수를 나타낸  $4,404\text{ cm}^{-1}$  (peak 1) 영역은 셀룰로오스와 헤미셀룰로오스를 나타낸다고 알려져 있다(Schwanninger *et al.*, 2011).

이상의 결과로부터 PLS-DA 모델을 이용하여 소나무와 금강송을 구분할 수 있다는 것을 확인하였다. 그러나 본 연구에 사용된 소나무와 금강송의 샘플은 별채된지 오래되지 않은 시험편이기 때문에 문화재에 사용된 부재와 같이 노화된 목재의 수종 분류에 그대로 적용하기에는 무리가 있다. 따라서 추후 연구에서는 고목재(archeological wood)의 수종 분류 및 최적의 측정 조건 결정에 관한 연구 수행이 진행되어야 할 것이다. 그리고 소나무와 금강송의 분류에 큰 영향을 미치는  $4,500 - 4,100\text{ cm}^{-1}$  영역에 대한 추가적인 연구가 필요하다.

## 4. 결 론

소나무와 금강송의 수종 분류를 위해 NIR 스펙트럼과 PCA 및 PLS-DA를 이용하여 수종 분류 모델을 설계하였다. 측정된 모든 NIR 스펙트럼을 이용하여 PCA를 실시한 결과 소나무와 금강송의 수종 분류는 불가능하였다. 그러나 2차 미분된 스펙트럼을 이용하여 시험편의 단면과 심·변재 구분에 따른 수종 분류에서는 변재부에서 수종 분류가 가능하였으며, 특히 방사단면의 변재에서는 명확하게 수종이 분류되었다. 그리고 개발된 PLS-DA 예측 모델을 이용하여 소나무와 금강송의 명확한 수종 분류가 가능하였

으며, 2차 미분을 통한 스펙트럼의 전처리는 수종 판별 정확도를 향상시켰다. 2차 미분 스펙트럼을 이용한 예측 모델은 100%의 분류 정확도를 나타내었으며, 예측 모델의  $R_p^2$  값은 0.86, RMSEP는 0.38로 나타났다. 전처리하지 않은 스펙트럼과 2차 미분 스펙트럼을 이용한 예측 모델의 신뢰도는 유사하였다. 본 연구에서 제안된 PLS-DA 모델은 소나무와 금강송 등 유사수종의 분류에 적합하였다. 그러나 수종 분류에 영향을 미치는 주요 성분들에 대한 추가적인 연구가 필요하다.

## REFERENCES

- Antti, H., Sjöström, M., Wallback, L. 1996. Multivariate calibration models using NIR spectroscopy on pulp and paper industrial applications. *Journal of Chemometrics* 10: 591-603.
- Brunner, M., Eugster, R., Trenka, E., Berganmin-Strotz, L. 1996. FT-NIR spectroscopy and wood identification. *Holzforchung* 50: 130-134.
- Chang, Y.S., Yang, S.Y., Chung, H., Kang, K.Y., Choi, J.W., Choi, I.G., Yeo, H. 2015. Development of moisture content prediction model for *Larix kaempferi* sawdust using near infrared spectroscopy. *Journal of the Korean Wood Science and Technology* 43(3): 304-310.
- Eom, C.D., Han, Y.J., Chang, Y.S., Park, J.H., Choi, J.W., Choi, I.G., Yeo, H. 2010. Evaluation of surface moisture content of *Liriodendron tulipifera* wood in the hygroscopic range using NIR spectroscopy. *Journal of the Korean Wood Science and Technology* 38(6): 526-531.
- Hitoshi, Y., Tsuchikawa, S. 2003. Near-Infrared spectroscopic comparison of antique and modern wood. *Applied Spectroscopy* 57(11): 320-340.
- Horikawa, Y., Mizuno-Tazuru, S., Sugiyama, J. 2015. Near-infrared spectroscopy as a potential

- method for identification of anatomically similar Japanese diploxylons. *Journal of Wood Science*, Published online.
- Lee, S., Lohumi, S., Cho, B.K., Kim, M.S., Lee, S.H. 2014. Development of nondestructive detection method for adulterated powder products using Raman spectroscopy and partial least squares regression. *Journal of Korean Society for Nondestructive Testing* 34(4): 283-289.
- Lewis, I.R., Daniel, N.W., Chaffin, N.C., Griffiths, P.R. 1994. Raman spectrometry and neural networks for the classification of wood types-1. *Spectrochimica Acta Part A: Molecular Spectroscopy* 50(11): 1943-1958.
- Osborne, B.G., Fearn, T., Hindle, P.H. 1993. (Eds.). *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman Scientific and Tech. Harlow.
- Pastore, T.C.M., Braga, J.W.B., Coradin, V.T.R., Magalhaes, W.L.E., Okino, E.Y.A., Camargos, J.A.A, de Muniz, G.I.B., Bressan, O.A., Davrieux, F. 2011. Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: discrimination of true mahogany, cedar, andiroba, and curupixa. *Holzforschung* 65: 73-80.
- Savitzky, A., Golay, M.J.E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627-1639.
- Schimleck, L.R., Michell, A.J., Vinden, P. 1996. Eucalypt wood classification by NIR spectroscopy and principal components analysis. *Appita Journal* 49: 319-324.
- Schwanninger, M., Rodrigues, J.C., Fackler, K. 2011. A review of band assignments in near infrared spectra of wood and wood components. *Journal of Near Infrared Spectroscopy* 19: 287-308.
- Siesler, H.W., Ozaki, Y., Kawata, S., Heise, H.M. (Eds.). 2008. *Near-infrared spectroscopy: principles, instruments, applications*. John Wiley & Sons.
- Tsuchikawa, S., Inoue, K., Noma, J., Hayashi, K. 2003a. Application of near infrared spectroscopy to wood discrimination. *Journal of Wood Science* 49: 29-35.
- Tsuchikawa, S., Siesler, H.W. 2003b. Near-Infrared spectroscopic monitoring of the diffusion process of deuterium-labeled molecules in wood. Part I: Softwood. *Applied Spectroscopy* 57(6): 187-198.
- Tsuchikawa, S., Yonenobu, H., Siesler, H.W. 2005. Near-infrared spectroscopic observation of the ageing process in archaeological wood using a deuterium exchange method. *Analyst* 130: 379-384.
- Watanabe, K., Abe, H., Kataoka, Y., Nodhito, S. 2011. Species separation of aging and degraded solid wood using near infrared spectroscopy. *Japanese Journal of Historical Botany* 19: 117-124.
- Wold, S., Esbensen, K., Geladi P. 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2: 37-52.
- Yang, B.H. 2006. *Understanding multivariate data analysis*. CommunicationBooks, Seoul, Korea.
- 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. *Journal of the Korean Wood Science and Technology* 43(3): 311-319.