

IR 및 NIR 스펙트럼과 주성분 분석을 통한 지종의 분류

김강재 · 엄태진[†]

접수일(2016년 1월 8일), 수정일(2016년 2월 1일), 채택일(2016년 2월 4일)

Classification of papers using IR and NIR spectra and principal component analysis

Kang-Jae Kim and Tae-Jin Eom[†]

Received January 8, 2016; Received in revised form February 1, 2016; Accepted February 4, 2016

ABSTRACT

In this study, we classified three copying papers and Korean, Chinese, and Japanese traditional papers using IR and/or NIR spectra and principal component analysis. Various chemicals are used when producing fine papers. In this case, the IR method to analyze functional groups is suitable for the classification of paper. On the other hand, NIR analysis is more suitable for the classification of traditional papers, as it uses nearly raw materials (pulp). Therefore, principal component analysis using IR and NIR depending on the paper production process will be the classification tool of paper.

Keywords: Chemometrics, principal component analysis (PCA), ATR, NIR, multivariate analysis, classification

1. 서론

화학계량학(chemometrics)이란 스펙트럼이나 크로마토그램과 같은 다양한 분석화학 데이터에서 원하는 정보를 도출하여 정성 및 정량분석을 가능하게 하는 수학적 통계 알고리즘을 총칭하는 것이다. 화학계량학에서는 주로 다변량 분석을 사용한다. 다변량 분석(multivariate analysis)은 단변량 분석과는 대조되는 방법으

로 여러 개의 독립변수에 대한 여러 개의 종속변수를 동시에 분석해보는 통계적 기법이다.¹⁾ 다변량 분석법은 크게 3가지 분석 모델이 존재한다. 기술모델(descriptive model)에는 주성분분석(principal component analysis, PCA), 기존 통계량(basic statistics), 집단분석(clustering)이 있으며, 두 번째 모델인 회귀 및 예측 모델(regression & predictive model)에는 다중회귀분석(multiple linear regression, MLR), 주성분회귀분석

• 경북대학교 농업생명과학대학 임산공학과(Dept. of Wood Science and Technology, College of Agriculture and Life Science, Kyungpook National University, Daegu, 702-701, Korea)

[†] 교신저자(Corresponding Author): E-mail: tjeom@knu.ac.kr

(principal component regression, PCR), 부분최소자승회귀분석(partial least square regression, PLSR)이 있다. 마지막으로 분류모델(classification model)에는 SIMCA법(soft independent modelling of class analogies, SIMCA), 서포트 벡터 머신(support vector machine, SVM), 선형판별분석(linear discriminant analysis, LDA) 및 부분최소자승판별분석(partial least squares-discriminant analysis, PLS-DA) 등이 있다.^{2,3)}

화학계량학적 방법은 제품의 종류 및 원산지 판별, 그리고 진위 분석 등에 많이 사용되고 있는데, Watanabe 등⁴⁾은 PLS를 이용하여 고목재의 수종을 분류하였으며 Schimleck 등⁵⁾은 NIR 스펙트럼과 PCA를 이용하여 유칼립투스의 수종 및 원산지 판별에 대하여 연구 결과를 보고하였다. 그리고 Horikawa 등⁶⁾은 다변량 분석을 통해 육송과 해송을 구분하였다. 이와 같이 목재 관련에서는 많은 연구가 진행되고 있는 화학계량학적 방법이 제지분야에서는 그 사례를 찾아보기가 힘들다.

앞서 언급한 여러 다변량 분석법 중, PCA는 다루기 힘든 고차원의 신호를 낮은 차원으로 줄이는 차원 감소 기법 중 하나로 다차원의 공간에 데이터가 분포되어 있을 때 가장 효율적으로 차원을 축소할 수 있는 축을 찾아 그 축으로 차원을 축소시키는 것이다(Fig. 1). 다시 말해, PCA에서는 얻고자 하는 정확한 축을 찾아내는 것이 아니라 여러 축 중 최적의 축을 찾아내는 것이 핵심이다.⁷⁾

본 연구에서는 국내에서 시판되고 있는 복사용지 3종과 한중일 전통지 4종의 IR/NIR 스펙트럼 데이터를 수집하여 주성분분석(PCA)으로 지종의 분류 가능성을 알아보고자 하였다.

2. 재료 및 방법

2.1 공시재료

본 연구에 사용된 복사용지는 현재 국내에서 가장 많이 유통되며 사용되어지고 있는 각 제지회사의 D 용지, H 용지 및 M 용지를 사용하였으며 전통지는 한국, 중국 및 일본에서 전통방식으로 제작한 전통지를 선정하였다. 한지는 문경과 의령에서 제조된 것이고 중국의 선지는 안휘성(安徽省)에서 제조한 것을 상해에서 구입하였으며 일본의 화지는 시마네현(Shimane, 島根縣)에서 제조되는 세키큐반시(Sekikyu banshi)를 사용하였다. Table 1에는 복사용지와 전통지의 성분 분석 자료를 나타내었다.

2.2 실험방법

2.2.1 ATR 및 NIR 분석

주성분분석을 위한 전 단계로 시료의 스펙트럼 데이터를 수집하기 위해 감쇠 전반사 적외선 분석기(Attenuated total reflection infrared spectroscopy, ATR)와 근적외선 분석기(Near Infrared spectroscopy, NIR)를 사용하였다. ATR과 NIR의 제원 및 분석 조건은 Table 2와 같다.

복사용지와 전통지는 ATR 측정을 위해 각각 다른 종이에서 10개의 시료를, NIR 측정을 위해선 8개씩의 시료를 준비하였다. 이들의 스펙트럼 분석을 실시하여 최종적으로 70개의 ATR 스펙트럼 데이터와 56개의 NIR 스펙트럼 데이터를 수집하였다. 수집된 스펙트럼 데이터는 Savitzky-Golay 알고리즘⁸⁾을 이용하여 5차 다항식에 의해 2차 미분하는 것으로 전처리를 실시하였다.

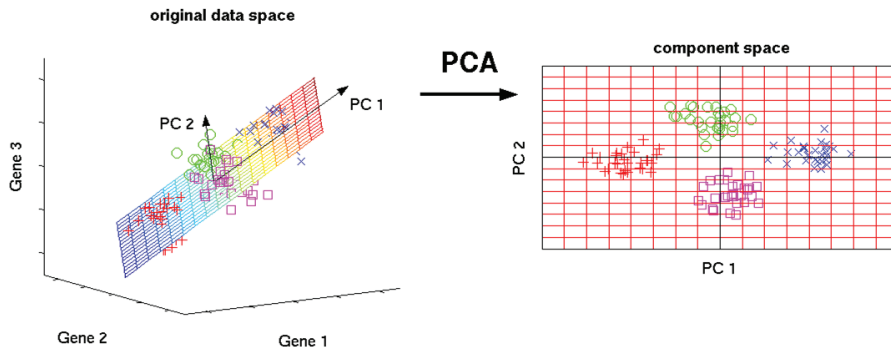


Fig. 1. Transformation model of principal component analysis (PCA).⁷⁾

Table 1. Component of copying papers and traditional papers

Sample		Basis weight, g/m ²	Ash, %	Holocellulose (α -cellulose), %	Lignin, %
Copying paper	D paper	80.6	10.9	84.1 (78.7)	2.3
	H paper	74.7	20.3	75.3 (71.4)	2.0
	M paper	74.5	21.8	73.9 (70.9)	1.6
Traditional paper	CHN_Shanghai	30.1	3.1	84.9 (75.1)	1.7
	JPN_Shimane	36.1	4.8	77.8 (67.5)	3.6
	KOR_Munkyung	41.4	5.4	81.1 (70.2)	1.0
	KOR_Uiryung	57.5	5.9	79.2 (69.6)	1.3

Table 2. Information and conditions of ATR and NIR spectrometer

	ATR	NIR
Model	Alpha-P, Bruker Optics	NPA-PDA1700, Bruker Optics Spectrum 100N, PerkinElmer
Resolution	2 cm ⁻¹	1 nm 16 cm ⁻¹
Scan number	16 times	16 times 32 times
Wavelength range	4,000 - 400 cm ⁻¹	900 - 1,700 nm 10,000 - 4,000 cm ⁻¹

2.2.2 주성분 분석(PCA)

PCA는 변수와 개체의 수가 많은 데이터를 분석하는 방법으로서 데이터의 단순화, 이상치 탐지, 변수 선택 및 분류 등을 주목적으로 한다.⁹⁾ 2차 미분으로 전처리된 126개의 스펙트럼 데이터들의 주성분 분석을 위해 소프트웨어 프로그램인 Unscrambler[®] Ver. 9.8(CAMO Software Inc., Woodbridge, NJ)을 이용하였다. 2차 미분 IR 스펙트럼 데이터는 2,000-400 cm⁻¹의 파장에서, 2차 미분 NIR 스펙트럼 데이터는 각각 8,000-4,000 cm⁻¹과 935-1,665 nm의 파장 범위에서 분석이 진행되었다.

3. 결과 및 고찰

3.1 복사용지 IR 및 NIR의 스펙트럼 분석

Fig. 2는 3종류 복사용지의 IR 및 NIR 스펙트럼과 이들의 2차 미분 스펙트럼을 나타낸 것이다. Fig. 2에서 좌측은 전처리를 하지 않은 스펙트럼 즉, 원본 스펙트럼 데이터를 나타낸 것이고 우측의 그래프는 2차 미분으로 전처리한 스펙트럼 데이터이다. IR 및 NIR의 미처리 스펙트럼(Fig. 2의 좌)에서 보면 3종류의 복사용지 모두

거의 비슷한 피크 패턴을 보이고 있다. 따라서 3종류의 복사용지는 IR이나 NIR 스펙트럼으로는 분류가 어렵다. 본 연구에 사용된 종이처럼 고체로 이루어진 시료들은 IR이나 NIR 측정 시 스펙트럼의 재현성이 떨어지기 때문에 분석 결과 향상을 위해 Fig. 2의 우측 그래프와 같이 전처리 과정이 필요하다.¹⁰⁾ 2차 미분으로 전처리한 3종류 복사용지의 그래프에서는 IR과 NIR 스펙트럼 모두에서 특정 영역의 차이가 조금씩 나타났다. 하지만 이것으로 3종류의 복사용지를 분류하는 데에는 한계가 있다. 따라서 2차 미분 스펙트럼으로 PCA 분석을 실시하였다.

3.2 2차 미분 스펙트럼을 이용한 복사용지의 주성분 분석

Fig. 3은 복사용지의 IR 및 NIR 스펙트럼을 2차 미분으로 전처리한 후 PCA를 실시한 결과를 나타낸 것이다. PC 1과 PC 2는 2차 미분 스펙트럼으로 주성분 분석을 실시하였을 때 다차원의 공간에서 가장 많은 시료의 점들이 가장 많이 분산된 방향을 PC 1이 되고 이에 직각방향에서 분산된 점들이 많은 방향이 PC 2가 된다.³⁾ PCA에서는 PC의 수를 임의로 지정할 수 있으며 본 연구에서는 PC 1에서 PC 4까지 지정하였다. 이 때 가장 높은 분류 정확성을 보인 것이 PC 1과 PC 2이므로 본 연구에서는

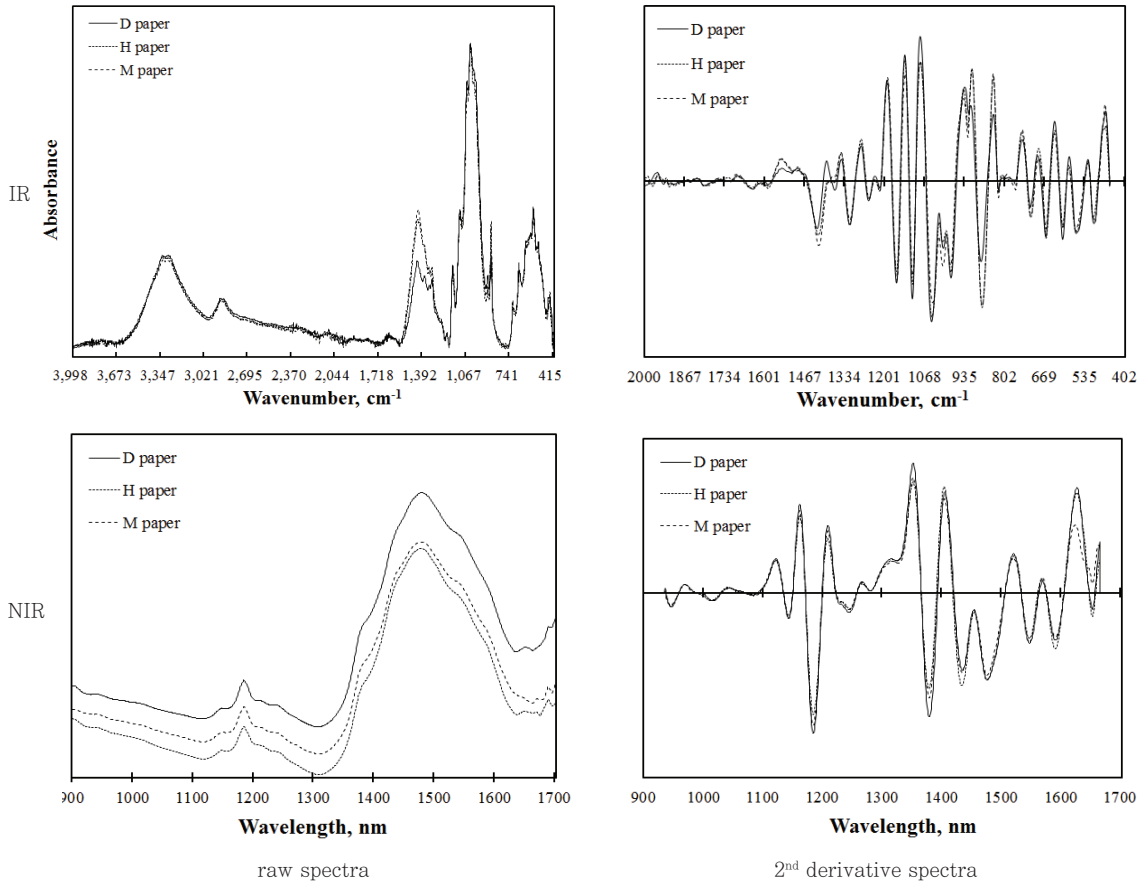


Fig. 2. IR and NIR raw spectra and 2nd derivative spectra of copying paper.

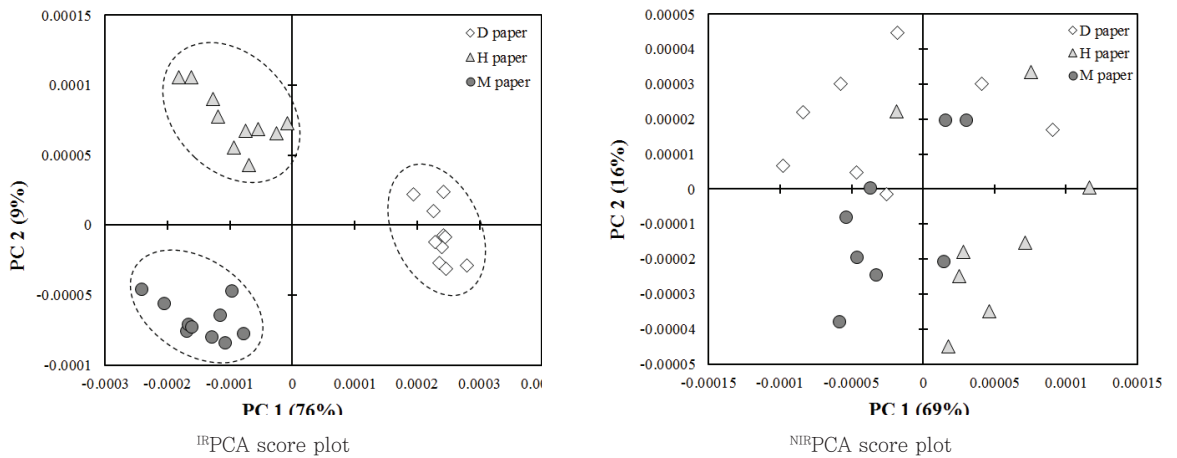


Fig. 3. PCA score plots of copying paper on the PC 1 and PC 2.

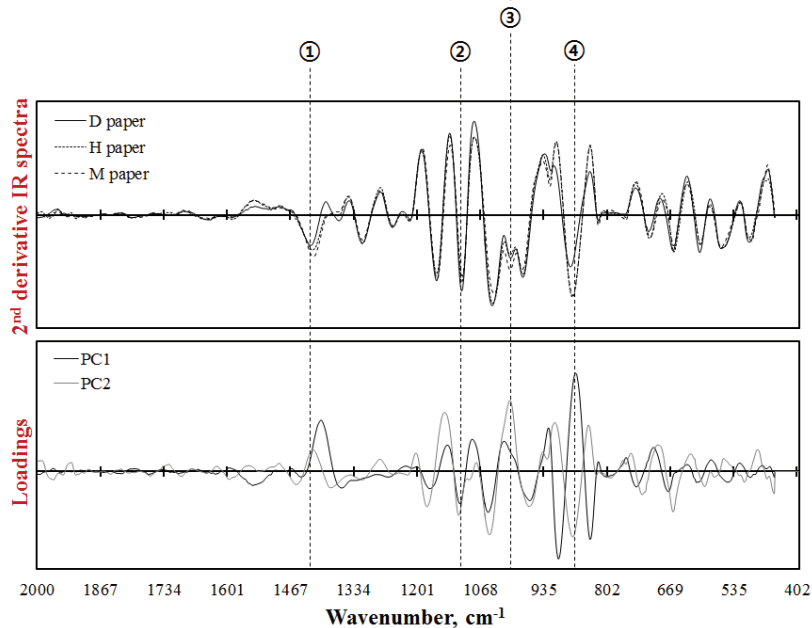


Fig. 4. Comparison between the 2nd derivative IR spectra from copying papers and the spectra from PC 1 and PC 2 in PCA.

이 두 PC를 주로 이용하였다. 또한 PCA 그래프에서 각 주성분(PC 1, PC 2)과 함께 표기된 백분율 값을 '설명할 수 있는 분산(explained variance)'으로 PCA 모델의 총 분산 중에서 해당 주성분을 가장 잘 설명할 수 있는 분산의 양을 말한다.¹⁰⁾

Fig. 3에서 보면, IR-PCA에서는 3종류의 복사용지가 PC 1의 0을 중심으로 D 복사용지와 H, M 복사용지가 분류되는 것을 볼 수 있으며 PC 2에서는 음의 값에서 M 복사용지, 0의 값에선 D 복사용지 그리고 양의 값에선 H 복사용지가 완벽히 분류되었다. 하지만 NIR-PCA 결과에서는 그래프의 좌측 상단에서 우측 하단으로 복사용지가 어느 정도 분류가 되는 것처럼 보이나 각 복사용지의 스코어 플롯이 혼재되어 있는 부분이 있어 완벽히 분류되었다고 할 수가 없다. 이러한 결과로 보아, 복사용지와 같이 종이 제조 시 많은 종류의 약품과 무기 충전물이 들어가는 지종의 경우에는 NIR 스펙트럼을 이용한 PCA 분류보다는 성분 내 작용기를 분석하는 IR 스펙트럼을 이용하여 PCA를 실시하는 것이 더욱 효과적이라고 볼 수 있을 듯하다. 그러므로 복사용지 IR-PCA로부터 지종 분류가 나타난 스펙트럼의 피크를 추적하기 위해 Fig. 4와 같이 2차 미분 스펙트럼과 PC 1 및 PC 2의 로딩값을 비교하여 보았다.

Fig. 4는 2차 미분 IR 스펙트럼과 PCA 스코어 플롯에서 분류가 된 PC 1과 PC 2의 로딩값을 비교한 그래프이다. 로딩값은 PCA에서 생성된 변수 PC에 대하여 기존의 변수(2차 미분 스펙트럼)가 얼마나 기여하는지를 나타낸 것이다.¹⁰⁾ 여기서는 PC 1과 PC 2에 의해 복사용지의 지종이 분류가 되었기 때문에 이 둘의 로딩값을 비교하였다. 그 결과, 각 PC의 분류에 영향을 미치는 4개의 주요 피크가 관찰되었다. 이들의 피크는 Table 3에 나타내었다.

Table 3. Assignment of main IR absorption peaks by PCA of copying paper

Peak	Wavenumber	Functional group
①	1,428 cm ⁻¹	-COOH or CO=
②	1,108 cm ⁻¹	C-O bond stretch
③	1,004 cm ⁻¹	HC-CH ₂
④	874 cm ⁻¹	Substd,* benzene ring

*substd : substituted

복사용지를 분류하는데 있어서 PC 1은 2개의 주요 피크가 영향을 미쳤다. 그 두 피크는 1,108 cm⁻¹(②)과 874 cm⁻¹(④)의 영역이었다.¹¹⁾ ②의 피크는 C-O 결합을 하는 피크로 셀룰로오스와 AKD와 같은 사이즈제 사이의 결합에 해당하는 것이고 ④의 피크는 펄프 내에 존재하는

치환된 benzene ring에 해당하는 것이다. 그 중에서도 PC 1의 분산에 가장 큰 영향을 미치는 피크는 ④ 피크이며 이는 펄프화 과정에서 생성된 -SH와 -OH 이온이 α-O-4와 β-O-4 결합을 방해하여 재축합을 방지하는 것으로¹¹⁾ 이 작용기의 생성 차이에 따라 복사용지가 분류되는 것으로 판단된다.

PC 2의 분산에 영향을 미치는 피크는 4개로 1,428(①), 1,108, 1,004(③) 및 874 cm⁻¹가 그것이다. 이 피크들은 Esteves 등¹²⁾과 Largo-Goosens 등¹³⁾의 연구에서도 나타난 바와 같이 각각 -COOH or CO=, C-O 결합, HC-CH₂ 및 치환된 benzene ring을 나타낸다. 그 중에서도 PC 2의 분산에 가장 많은 영향을 미친 피크는 ③과 ④였다. ③의 피크는 종이 제조 시 사용된 PAM과 같은 지력증강제가 셀룰로오스 수산기와 결합되는 부분의 영역으로 예상되며 이에 따라 PC 2에서 지종을 분류하는 피크로 선별된 것이다. 이상 4가지의 주요 피크로 PC 1과 PC 2에 의해 복사용지의 지종이 분류된 것을 알 수 있었다.

3.3 전통지의 IR/NIR 스펙트럼 분석

일반적으로 한중일의 전통종이는 그 원료(펄프)의 종류가 상이하다. 한지는 주로 닥나무 인피섬유를 많이 사용하는 반면, 중국의 선지는 꾸지나무와 벗짚을 혼합하여 사용하며 일본의 화지는 주로 삼지닥나무가 그 원료인 것으로 알려져 있다. Fig. 5에서 보면, 원료가 서로 다른 전통지 임에도 IR이나 NIR의 스펙트럼 차이는 크게 나타나지 않았다. 따라서 Fig. 5의 우측 그래프와 같이 2차 미분으로 전처리를 실시한 다음 주성분 분석을 실시하였다.

3.4 2차 미분 스펙트럼을 이용한 전통지의 주성분 분석

Fig. 6은 Fig. 5에서 수집된 2차 미분 스펙트럼 데이터를 PCA에 적용하여 분산된 스코어 플롯을 나타낸 그래프이다. 복사용지와는 달리, 전통지에서는 PCA에 의한 분

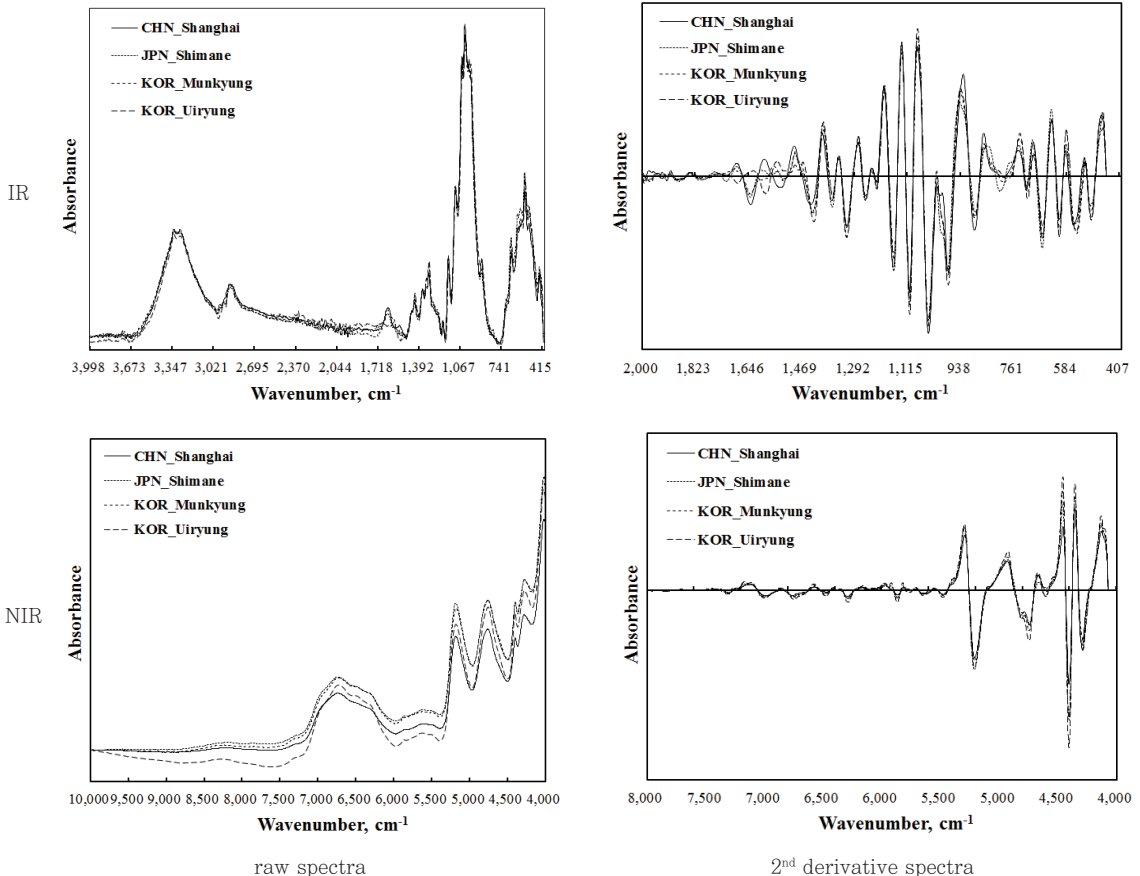


Fig. 5. IR and NIR raw spectra and 2nd derivative spectra of traditional paper.

류보다는 NIR-PCA에 의한 지종의 분류가 더욱 뚜렷이 관찰되었다. IR-PCA 결과를 보면 한국, 중국 및 일본의 종이에 따른 분류는 확연히 일어나지만 한국의 전통지의 스코어 플롯들이 혼재되어 존재하는 것을 볼 수 있다. 같은 원료(닥나무)를 사용한 한지의 경우, 만든 지역에 관계없이 플롯들이 혼합되어 있어 이 둘을 따로 주성분 분석을 실시한 결과 어느 정도의 분류는 이루어졌으나 각 성분의 분산량이 그리 높지 않아 확실한 분류가 이루어

지지 않았다. 반면에, NIR-PCA에서는 한중일 전통지의 분류뿐만 아니라 한지 내에서도 지역별로 분류가 명확히 이루어지는 것을 볼 수 있었다. 이러한 결과에 비추어 보면 전통지는 복사용지와는 반대로 원료 이외의 약품들이 거의 들어가지 않는 것으로 종이의 주성분인 셀룰로오스, 헤미셀룰로오스 및 리그닌 등을 분석할 수 있는 NIR이 상대적으로 더욱 적합할 것으로 사료된다.

Fig. 6의 결과를 바탕으로 NIR-PCA로부터 얻어진 PC

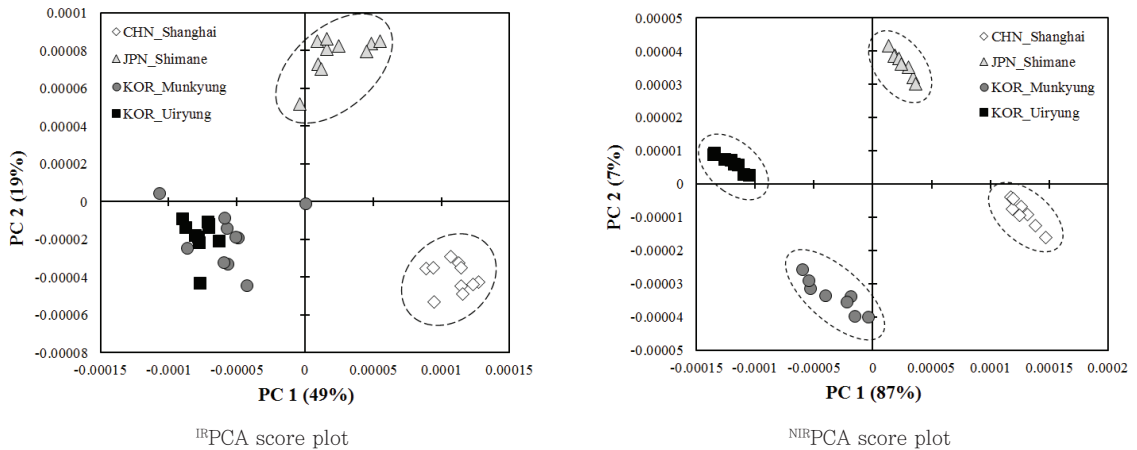


Fig. 6. PCA score plots of copying paper on the PC 1 and PC 2.

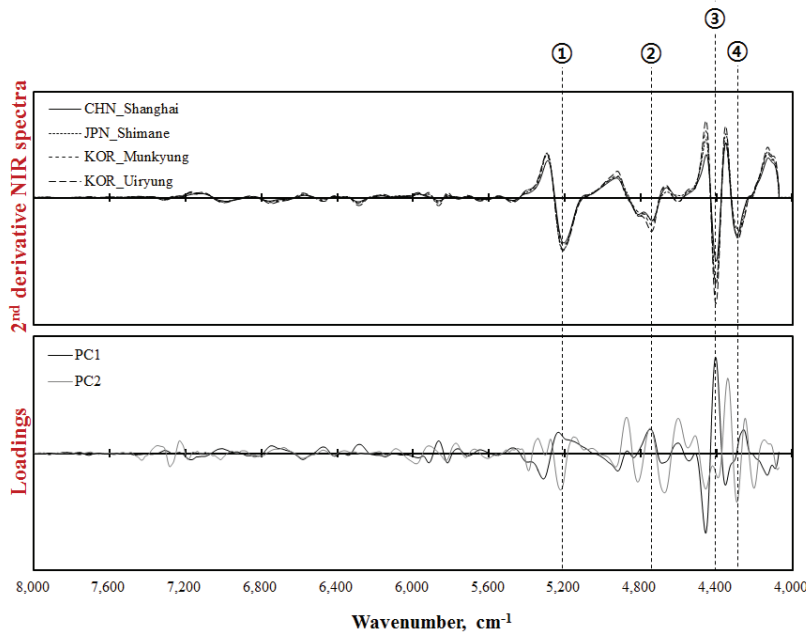


Fig. 7. Comparison between the 2nd derivative NIR spectra from copying papers and the spectra from PC 1 and PC 2 in PCA.

Table 4. Assignment of main NIR absorption peaks by PCA of traditional paper

Peak	Wavenumber	Bond vibration	Remarks
①	5,204 cm ⁻¹	OH str. ^{**} + OH def.	water
②	4,740 cm ⁻¹	O-H def. ^{***} + O-H str.	cellulose
③	4,404 cm ⁻¹	CH ₂ str. + CH ₂ Def.	cellulose, hemicellulose
④	4,284 cm ⁻¹	CH str. + CH def.	cellulose, hemicellulose, xylan

^{**}str.: stretching vibration, ^{***}def.: deformation vibration

의 로딩값과 2차 미분 스펙트럼을 비교하여 Table 4에 분류에 영향을 미치는 주요 피크 4가지를 확인하였다. PC 1의 분산에 영향을 미치는 피크는 4,740 cm⁻¹(②)과 4,404 cm⁻¹(③)이었으며 이들은 각각 셀룰로오스와 헤미셀룰로오스에 해당하는 것이다.¹⁴⁾ 각 전통지는 Table 1에서 언급한 것과 같이 α-셀룰로오스의 함량이 약 67~74%로 상이하게 나타나 이에 따른 분류의 결과와 예상할 수 있다. PC 2의 분산에는 4개의 피크가 모두 관여하였으며 이들은 5,204 cm⁻¹(①) 영역의 수분, 4,740 cm⁻¹(②) 영역의 셀룰로오스, 4,404 cm⁻¹(③) 영역에서의 셀룰로오스 혹은 헤미셀룰로오스 및 4,284 cm⁻¹(④) 영역의 헤미셀룰로오스 중 자일란에 의한 피크들이다.^{15,16,17)} 그러나 수분에 해당하는 ①의 피크의 수분은 NIR 측정시 목재에서는 함수율에 의해 적절한 피크가 형성이 되나 종이에서는 피크가 아닌 노이즈로 나타날 수도 있으므로 본 연구의 성분 분류에서는 제외하였다. NIR-PCA 데이터를 통해 얻은 PC 2의 주성분은 주로 자일란¹⁷⁾에 의한 분산량이 가장 높은 것으로 나타났다. 하지만 한지의 지역별 분류가 가능한 점 및 자일란이 지중 분류에 어떠한 영향을 미치는지에 대한 연구는 추가적으로 수행되어야 할 것으로 보인다.

본 연구에서는 IR 및 NIR의 스펙트럼 데이터를 이용하여 지중을 분류하여 보았다. 그 결과 흥미롭게도 종이의 제조 방법에 따라 지중을 분류하는 분석법이 상이하였다. 하지만 실제로 이러한 지중 분류에 이상의 주성분들이 얼마나 영향을 미쳤는가에 대한 추가적인 연구가 필요할 것으로 보인다.

4. 결론

국내에서 시판되고 있는 복사용지 3종과 한중일 전통지 4종의 IR/NIR 스펙트럼 데이터를 수집하여 주성분 분석(PCA)으로 지중을 분류한 결과는 다음과 같다. 종

이 제조 시 다양한 약품이 사용되는 복사용지의 경우에는 작용기의 분석 위주로 이루어지는 IR 분석법이 주성분 분석을 통한 지중 분류에 적합하였다. 반면에 전통지와 같이 종이 제조 시 원료 이외에 다른 약품이 거의 들어가지 않는 경우에는 종이 내 셀룰로오스, 헤미셀룰로오스 및 리그닌 등을 분석하는 NIR 분석법이 더욱 정확하게 지중을 분류하였다. 따라서 종이 제조방법에 따라 IR과 NIR을 적절히 사용하면 PCA를 통해 지중 분류가 가능할 것으로 보인다.

Literature Cited

1. Yang, B. H., Understanding multivariate data analysis, CommunicationBooks, Seoul, Korea (2006).
2. Lee, S., Lohumi, S., Cho, B. K., Kim, M. S. and Lee, S. H., Development of nondestructive detection method for adulterated powder products using Raman spectroscopy and partial least squares regression, J. Korean Society for Nondestructive Testing, 34(4): 283-289 (2014).
3. Park, G. B., Multivariate data analysis, Hakjisa, Seoul, Korea (2000).
4. Watanabe, K., Abe, H., Kataoka, Y. and Nodhito, S., Species separation of aging and degraded solid wood using near infrared spectroscopy, Japanese Journal of Historical Botany, 19(1-2): 117-124 (2011).
5. Schimleck, L. R., Michell, A. J. and Vinden, P., Eucalyptus wood classification by NIR spectroscopy and principal components analysis, Appita Journal, 49(5): 319-324 (1996).

6. Horikawa, Y., Mizuno-Tazuru, S. and Sugiyama, J., Near-infrared spectroscopy as a potential method for identification of anatomically similar Japanese diploxylons, *J. Wood Science*, 61(3): 251-261 (2015).
7. Mattias S., Approaches to analyse and interpret biological profile data, Ph. D. thesis, University of Potsdam, Germany (2006).
8. Savitzky, A. and Golay, M. J. E., Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry*, 36(8): 1627-1639 (1964).
9. Wold, S., Esbensen, K. and Geladi, P., Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2(1-3): 37-52 (1987).
10. Hwang, S. W., Lee, W. H., Horikawa, Y. and Sugiyama, J., Chemometrics approach for species identification of *Pinus densiflora* Sieb. et Zucc. and *Pinus densiflora* for *erecta* Uye-ki, *J. Korean Wood Science and Technology*, 43(6): 701-713 (2015).
11. McDonald, R. G., *The pulping of wood* Volume I, McGraw-Hill Book Company, USA (1969).
12. Esteves, B., Velez Marques, A., Domingos, I. and Pereira, H., Chemical changes of heat treated pine and eucalypt wood monitored by FTIR, *Maderas. Cinc. Tecnol.*, 15(2): 245-258 (2013).
13. Largo-Goosens, A., Hernández-Altamirano, M., García-Calvo, L., Alonso-Simón, A., Álvarez, J. and Acebes, J. L., Fourier transform mid infrared spectroscopy applications for monitoring the structural plasticity of plant cell walls, *Front. Plant Sci.*, 30(303): 1-15 (2014).
14. Schwanninger, M., Rodrigues, J. C. and Fackler, K., A review of band assignments in near infrared spectra of wood and wood components, *Journal of Near Infrared Spectroscopy*, 19(5): 287-308 (2011).
15. Tsuchikawa, S., A review of recent near infrared research for wood and paper, *Applied Spectroscopy Reviews*, 42(1): 43-71 (2007).
16. Osborne, B. G., Fearn, T. and Hindle, P. H., *Practical NIR spectroscopy with applications in food and beverage analysis*, Longman Scientific and Technical, Harlow (1993).
17. Bassett, K. H., Marchessault, R. H. and Liang, C. Y., The infrared spectrum of crystalline polysaccharides. IX. The near infrared spectrum of cellulose", *J. Polymer Science Part A*, 1(5): 1687-1692 (1963).