

## 근적외선 분광학과 화학계량학에 기반한 청정 바이오연료 실시간 품질 측정 기술 개발

조형수, 류준형<sup>†</sup>, 유 준\*

부경대학교 화학공학과  
608-739 부산광역시 남구 용당동 산 100  
<sup>†</sup>동국대학교 원자력 및 에너지 공학부  
780-714 경북 경주시 석장동 707

(2011년 7월 21일 접수; 2011년 8월 16일 수정본 접수; 2011년 8월 24일 채택)

### Development of an On-line Measurement Method for Clean Biofuel Based on Near Infrared Spectroscopy and Chemometrics

Hyeongsu Cho, Jun-hyung Ryu<sup>†</sup>, and J. Jay Liu\*

Department of Chemical Engineering, Pukyong National University, San 100, Yongdong-Dong, Nam-Gu, Busan 608-739, Korea

<sup>†</sup>Department of Nuclear and Energy System Engineering, Dongguk University, 707 Seokjang-Dong, Gyeongju, Gyeongbuk 780-714, Korea

(Received for review July 21, 2011; Revision received August 16, 2011; Accepted August 24, 2011)

#### 요 약

화석연료 사용에 따른 기후 변화 문제 및 자원 고갈에 대한 염려 때문에 바이오 연료는 그 중에서 가장 바람직한 재생에너지 대안이다. 하지만 대량 생산을 할 수 있는 경제성확보가 없으면 바람직하고 현실적인 대안이라 할 수 없다. 실시간 품질측정 시스템 개발은 바이오연료의 대량생산에 위하여 반드시 필요하다. 바이오연료 중 바이오에탄올은 ASTM (American Society for Testing & Materials) D4806-10과 같은 품질 규격을 따르며 이에 명시된 평가방법을 대량 생산에 적용하려면 여러 어려움이 뒤따른다. 따라서 본 연구에서는 바이오에탄올 제품 품질 분석의 저가격화와 시간단축 및 생산 공정의 모니터링 실시간화를 해결하기 위해 근적외선 분광분석법과 화학계량학의 부분최소자승법(Partial Least Squares)을 융합한 실시간 분광 모니터링 시스템을 새롭게 제안하였다. 수종의 전처리와 부분최소자승법을 이용하여 모델링한 결과 스펙트럼의 산란 보정, 노이즈 감소, 검량선의 유지보수 등에서 Savitzky-Golay 전처리가 가장 우수함을 보였다. 본 연구를 통해 실시간으로 품질 측정뿐만 아니라 분광분석 장치로 여러 성분을 동시에 측정 가능함을 보임으로서 비용의 절감도 가능할 것으로 예상된다. 그리고 상관계수 R<sup>2</sup>이 0.99 이상으로 실험실분석의 대체도 가능함을 보였다.

**주제어** : 바이오에탄올, 대량생산, 실시간 품질측정, 부분최소자승법, Savitzky-golay법

**Abstract** : It is an important issue to develop quality assessing system for biofuel for the purpose of accelerating the mass production of biofuel. It is particularly challenging to conduct testing method in the mass production of bioethanol while meeting quality specifications such as ASTM (American Society for Testing & Materials) D4806-10. In order to address this challenge, this paper proposes on-line spectroscopic quality assesment system based on Near Infrared spectrum and Partial Least Squares method in Chemometrics. As a result of testing a number of preprocessing methods and Partial Least Squares, it was found out that Savitzky-Golay method showed the best performance in terms of spectrum correction, noise reduction, and model maintenance. The proposed system allows us to assess multiple quality components continuously using spectroscopic facilities with the cheap cost. Since the value of R<sup>2</sup> is more than 0.99, it is possible to replace the laboratory analysis.

**Keywords** : Bioethanol, Mass production, On-line quality assessment, Partial least square method, Savitzky-golay method

#### 1. 서 론

최근 인간이 화석연료를 연소시킴으로써 대량으로 배출되는 이산화탄소가 지구온난화를 일으키는 온실가스의 주범으

로 지목되면서 화석연료의 사용에 대한 환경적, 정치적 및 사회적 논란이 거세지고 있다[1]. 즉 화석연료 사용에 따른 환경오염과 그로 인한 기후 변화 문제 및 자원 고갈에 대한 염려 때문에 재생에너지에 대한 수요는 지속적으로 증가하고 있다. 기존에 사용되는 화석연료시설과의 연계성을 고려할 때 바이오 연료는 그 중에서 가장 바람직한 재생에너지 대안이

\* To whom correspondence should be addressed.  
E-mail: jayliu@pknu.ac.kr

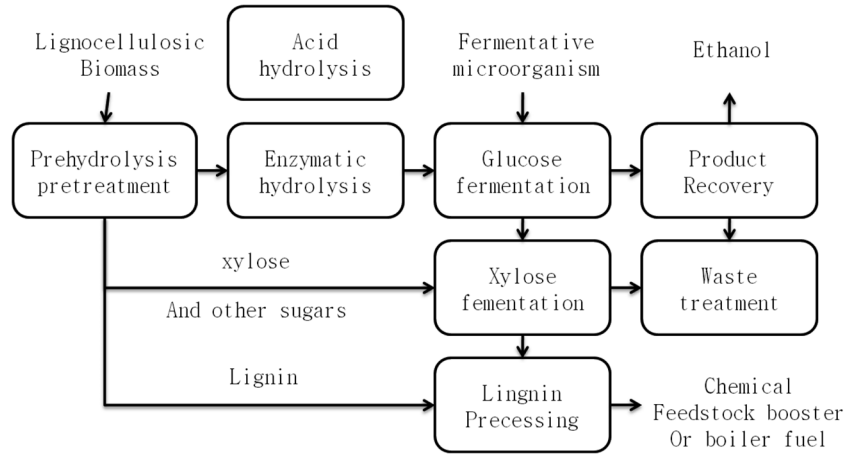


Figure 1. A typical Bioethanol production process from lignocellulosic biomass.

다. 하지만 대량 생산을 할 수 있는 경제성 확보가 없으면 바람직하고 현실적인 대안이라 할 수 없다.

이미 1970~1980년대부터 바이오연료 도입을 선도해 온 브라질, 미국, EU 등 바이오연료 강국의 경우, 다양한 바이오매스에 대한 공정개발은 거의 완료된 상태이며 대량생산 체계를 위한 기술개발 또한 매우 성숙된 단계이다. 하지만 한국의 경우 OECD국가 중 신재생 에너지 비중은 유럽, 미국 일본 등 선진국과 큰 격차를 보이고 있다[2]. 이 격차를 줄이기 위해서는 공정개발은 물론이며 대량 생산 기술개발이 동시에 이루어져야 한다. 하지만 2000년부터 최근 10여년간 바이오연료 관련 국가연구과제 현황을 살펴보면 70% 이상의 과제가 개별 공정개발에 국한되고 있으며, 바이오연료의 대량생산 기술개발은 전무한 실정이다.

바이오 연료 중 하나인 바이오에탄올은 당을 포함하거나 당으로 전환될 수 있는 녹말 또는 섬유소로부터 생산된다. 바이오 에탄올은 주로 옥수수 전분을 주원료로 하여 미생물과 효소를 이용한 생물학적 공정으로 생산되고 있으며, 가솔린 대체 연료 혹은 첨가제, 연료전지의 연료로 사용되고 있다. 에탄올은 기본적으로 당 화합물로부터 만들어지므로 전분 이외의 탄수화물로도 생산이 가능한데, 자연계에 가장 풍부한 물질 중의 하나인 단당류나 전분, 그리고 셀룰로오스가 대표적인 예이다. 일반적인 목질계 바이오매스로부터 에탄올의 전환 공정은 Figure 1에 나타내었다.

바이오에탄올 생산 공정(Figure 1)은 주로 소·중규모의 회분식 전환 공정으로 이루어져있다[3]. 그리고 연속식 공정과는 달리 강한 비선형성을 보이기 때문에, 세심한 제어와 모니터링, 및 적정 종말점 검출법(EPD, End-Point Determination)과 같은 대량생산 관련 기술의 중요성이 더 인식되어야 한다. 또한 작업의 복잡성과 실행이 어려운 작업 순서의 존재 같은 다루기 어려운 여러 가지 특성들 때문에 대량생산 기술 개발에 많은 노력이 필요하다. 그러나 무엇보다도 중요한 문제는, 앞서 서술한 여러 난제들을 극복하고 대량생산 기술을 개발하여도 회분식 공정의 중간 또는 최종제품의 품질을 실시간으로 모니터링 할 수 없다면, 상품성 있는 제품의 대량생산

기술의 실제 적용은 불가능하다는 점이다. 따라서 바이오연료 생산 공정의 실시간 품질 측정 기술 개발이 아주 시급하다 할 수 있다.

바이오연료에서 측정되어야 하는 것으로 바이오연료에 미량 포함되어 있는 불순물의 정량측정이 있으며, 이는 바이오연료의 대량생산 체계를 위한 제조공정의 최적화와 최종 제품의 품질을 확보하는데 있어 매우 중요하다. 바이오에탄올의 경우처럼 ASTM D4806-10(미국)과 같은 생산과 품질관리를 위한 표준이 제시되는 것도 이와 같은 이유때문이다. 하지만 전통적인 실험실 분석법(예를 들면, wet chemistry)으로 바이오매스 원료의 성분 또는 최종 제품의 불순물을 분석하려면 최소 4시간에서 최대 1일 이상이 소요되며, 여러 종류의 분석 장비와 많은 인력이 필요하므로 실제 산업 환경에 바로 적용하기에는 바람직하지 않다. 그러므로 본 논문에서는 근적외선 분광분석법과 화학계량학(Chemometrics)에 기반을 둔 분석기술을 개발하여 바이오연료 품질 분석의 저가격화와 시간단축 및 바이오연료 생산 공정의 모니터링 실시간화를 해결하고자 한다.

## 2. 실시간 분광 모니터링 시스템

바이오연료의 대량 생산을 위해서는 생산 공정에 공장단위의 제어 및 최적화와 같은 대량생산 관련 기술의 적용되어야 한다. 특히 공정에서 생산되는 바이오연료 제품의 품질을 분석하는데 있어서, 분석 비용의 절감 및 분석 시간단축이 선결되어야 할 문제이다. 여기서 분석 대상은 최종 생산물인 바이오연료에 미량 포함되어 있는 불순물을 정량적으로 측정할 필요가 있으며, 불순물을 적절하게 관리해야지만 제조공정의 최적화와 최종 제품의 품질 확보를 통해 바이오연료의 대량생산 체계를 구축할 수 있다. 미국의 경우 바이오에탄올의 생산과 품질관리를 위한 규격을 Table 1과 같은 ASTM D4806-10에 제시하였다[4].

바이오 에탄올의 경우 전체 제조 공정이 최적화되지 않았을 때 벤젠(Benzene), 올레핀(Olefin), 물, 메탄올 등의 불순물이

소량 생성된다. 이러한 불순물은 내연기관의 성능과 수명 그리고 배기가스 내 유해물질 등의 문제와 관련이 깊으므로 ASTM 표준에 주요 불순물로 분류되어 있고 이에 대한 다양한 분석방법 또한 제시하고 있다. 그러나 이 분석방법들에는 많은 한계점이 존재한다.

먼저 에탄올과 메탄올 및 벤젠의 측정방법으로 제시된 Gas Chromatography (GC) 측정법은 칼럼과 분석 시료 물질에 따라 측정시간이 다르며, 경우에 따라 전처리가 필요하다. 또한 휘발성 및 열안정성에 문제가 있고 시료의 기화에 따른 분석 물질의 파괴로 재사용이 불가능하다. 다음으로 올레핀 계열의 측정방법인 초임계유체크로마토그래피(SFC)는 체류시간, 선택도, 이론단수, 분리도 의존 온도, 압력, 밀도, 선속도, 이동상 조성 등의 많은 변수들에 대한 설정에 시간이 오래 걸리고 매우 복잡하다. 마지막으로 수분측정 방법으로 제시된 Karl Fisher법은 일반적으로 모든 시료의 수분 측정에 사용되지만 기기 조작에 있어 숙련자가 아니면 오차가 크고, 분석 시 시료가 파괴된다는 단점이 있다. 또한, 재현성이 좋지 않은 경우도 흔히 발생한다[5].

이와 같은 전통적인 분석화학에 의한 불순물 정량 측정 방법에는 각기 다른 분석 장비와 시약을 필요로 하고 이를 운영·유지보수하기 위해서는 많은 인적 자원과 물질적 비용을 필요로 한다. 더구나 위와 같은 방법들은 측정시간이 오래 걸리므로 이 분석결과를 가지고 엔지니어가 지속적으로 생산 공정을 최적상태로 운전하는 것은 불가능하다. 다시 말하면, 대량생산 측정 시스템이 갖추어야 할 요소로는 첫째, 분석시간이 짧아야 하고, 둘째 생산 및 검사 비용을 최소화 하며, 셋째로 시료 처리가 간편해야 한다. 또한 검량모델의 정확도가 뛰어나 안정적인 공정 운영과 체계적인 품질관리에 도움

이 되어야 한다. 따라서 대량생산 기술을 통해 바이오연료의 생산 공정을 최적의 상태로 유지하기 위해서는 위의 요소를 만족하는 실시간 분광분석 시스템을 개발하는 것이 매우 필요하다.

기존에 사용되고 있는 분석화학 기법 중에서, 반복·재현성이 우수하고 기기가 매우 안정적이어서 검량 모델(Calibration model)만 완벽하게 만들면 오랜 기간 사용이 가능할 수 있는 근적외선 분광법을 사용하고자 한다. 기기의 유지 보수가 크게 필요치 않아 기기가 안정적이며 오래 사용 가능하고 동시에 다 성분 측정이 가능하다. 그리고 GC, SFC 등의 기존 분석 장비는 고가인데 비해, 근적외선 분광계는 비교적 간단한 장비이어서 실시간, 비파괴 분석이 가능하다.

그러나 바이오 에탄올의 최종제품과 같은 복잡한 혼합물을 분석하는 경우, 각 화합물 고유의 근적외선 스펙트럼 간에 중첩이 있어 데이터가 복잡하고 분석이 어려워 MLR (Multiple Linear Regression)과 같은 방법으로는 분석이 힘들다. 반면에 화학계량학 중 부분최소자승법(Partial Least Squares, PLS)은 복잡한 근적외선 스펙트럼을 가지고도 여러 혼합물 형태의 시료를 분석할 수 있고, 여러 파장과 조성/물질간의 상호관계를 만들어 검량하기 위한 최적의 방법이다.

개괄적으로 말하면 바이오연료 제품 품질 분석의 저가격화와 시간단축 및 바이오연료 생산 공정의 모니터링 실시간화를 해결하기 위해 본 연구에서 근적외선 분광분석법과 화학계량학(Chemometrics)을 융합한 실시간 분광 모니터링 시스템을 새롭게 제안하며, 제안된 시스템은 Figure 2와 같이 도식화 할 수 있다(Figure 2 참조).

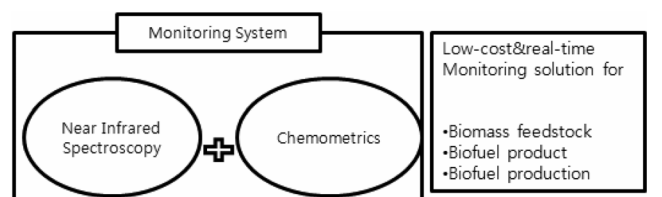
다음에는 이들 시스템 구성요소를 자세히 설명하겠다.

**Table 1.** Specification of Bioethanol given by ASTM D4806-10[4]

Property	Limit	ASTM method
Ethanol, volume %, min	92.1	D5501
Methanol, volume %, max	0.5	D5501
Solvent-washed gum content, mg/100mL, max	5.0	D381
Water, volume %, max	1.0	E203 or E1064
Inorganic Chloride, mass ppm (mg/L), max	10	D7319 or D7328
Copper, mg/kg, max	0.1	D1688
Acidity (as acetic acid CH <sub>3</sub> COOH), mass % (mg/L), max	0.007	D1613
pH	6.5 to 9.0	D6423
Sulfur, mass ppm, max	30.	D2622, D3120, D5453 or D7039
Total sulfate, mass ppm, max	4	D7318, D7319 or D7328
Benzene, vol% max	0.06	D5580
Olefins, vol% max	0.5	D6550
Aromatics, vol% max	1.7	D5580

**2.1. 근적외선 분광분석(NIR, Near Infrared spectroscopy)**

복사선의 흡수나 방출에 수반되는 광자의 에너지는 물질을 구성하는 원자나 분자의 에너지 준위간의 간격이나 분자의 운동진동수와 연관성을 갖는다. 이러한 전자기 복사선은 파장(Wavelength)에 따라 낮은 대역의 라디오파(Radio Wave)에서부터 마이크로파(Micro Wave), 적외선(Infrared, IR), 가시광(Visible, VIS), 자외선(Ultraviolet, UV), X선(X-Ray)등의 순서로 높아지며 각각의 파장에 따라 물질이 대응하는 에너지 형태가 다르다. 그 중에서 Figure 3에서와 같이 적외선은 근적외선(Near Infrared, NIR), 중적외선(Mid Infrared : MIR), 원적외선(Far Infrared, FIR)으로 다시 세분할 수 있다[6].



**Figure 2.** The proposed online measurement system design using Near Infrared spectroscopic analysis and Chemometrics.

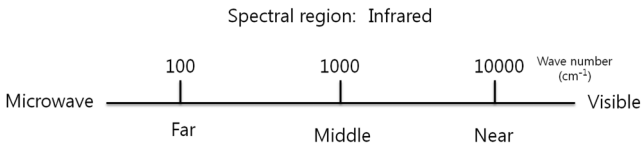


Figure 3. Electromagnetic Spectrum (Infrared region).

근적외선은 가시광선과 중적외선(mid-infrared) 사이에 존재하는 빛으로, 800 nm에서 2,500 nm (12,000~4,000 cm<sup>-1</sup>) 사이에 존재한다. 결국 가시광선보다는 에너지가 낮고 중적외선보다는 에너지가 높다. 근적외선에서의 흡수는 주로 중적외선에 유래되는 -CH, -OH, -NH 작용기 분자진동의 결합대(combination band)와 배음대(overtone band)로 나타난다. 결합대와 배음대로 나타나는 근적외선에서의 흡수는 흡광도가 많이 약해지기 때문에 중적외선에 흡수가 강한 -CH, -OH, -NH 작용기의 정보가 주로 나타나게 된다.

근적외선에서는 탄화수소의 경우 1차, 2차 배음대와 2개의 결합대가 나타나며 파장이 짧은 쪽으로 갈수록 감도는 약해지면서 흡수대는 넓어진다. 결국, 근적외선 지역에서 파장이 낮은 지역(1,100~1,650 nm)을 사용할 경우, 보통 경로길이가 10~20 nm 정도가 가능하며, 반대로 파장이 긴 지역(1,650~2,500 nm)을 사용할 경우 경로 길이를 1~4 mm 정도로 사용하여야 한다. 2차 배음대 흡수위치의 이론적으로 약간 다르게 나타나게 되며, 이는 배음대로 갈수록 분자 운동의 비대칭성(anharmonicity)이 증가되기 때문이다[7].

## 2.2. 화학계량학(Chemometrics)

근적외선 분광학(Near-IR spectroscopy)에서는 12,000~4,000 cm<sup>-1</sup>의 파장을 갖는 적외선을 시료에 조사하여 흡수스펙트럼을 측정하는데, 이 대역은 주로 분자들의 기본 밴드가 나타나는 영역으로 각 분자의 특징한 흡수밴드가 비교적 깨끗하게 분리되어 있어 특정 화합물의 존재 유무를 확인하는데 쓰여왔다. 일반적으로 시료에는 한 가지 화합물만이 존재하는 것이 아니고 여러 가지 화합물이 섞여 있는 경우가 많다. 특히 관심 있는 분자의 스펙트럼이 혼합물을 구성하는 다른 성분의 스펙트럼과 겹치게 되는 경우에는 화합물의 유무를 확인하기가 매우 어렵게 된다. 이 때 측정된 혼합물 시료의 스펙트럼으로부터 구성성분의 순수 스펙트럼을 분리해 낼 수 있다면 특정 유기화합물의 존재 유무를 확인하는데 큰 도움이 될 것이다.

이와 같이 여러 가지 화합물이 섞여있는 경우에는 복잡한 근적외선스펙트럼을 분석하기 위해서 예측력이 우수하고 잡음 등에 강건한 성능을 가진 검량 모델을 개발하는 화학계량학(Chemometrics)의 역할이 중요하며, 이를 통해 신속하고 정확한 화학적·물리적 성분의 예측을 가능하게 해준다. 화학계량학은 전자공학이나 통계학 등에서 많이 사용되는 방법으로 중첩이 심한 근적외선 스펙트럼에서도 원하는 조성 및 물성의 검량을 가능하게 한다. 즉 어떤 입력 변수 자체의 수가 너무 많거나 이론적인 해석이 쉽지 않아 기존의 일반 회귀분석으로는 상관관계를 찾기 어려울 때, 우리가 알고 있는 출력변

수를 이용해 입력변수와 통계적인 상관관계를 만드는 방법이 화학계량학이다. 이는 실험적인 검량기법으로 본 연구에서는 분광분석에서 입력변수는 표본 샘플의 스펙트럼에 해당되며 샘플의 조성이 출력변수이다. 주로 사용되는 화학계량학기법으로는 다중선형회귀법(Multiple Linear Regression), 부분최소자승법(Partial Least Squares) 등이 있다[7].

### (1) 다중선형회귀법(Multiple Linear Regression)

다중선형회귀법(MLR)은 입출력 변수가 각각 한 개씩인 경우에 적용하는 단순 선형 회귀법(Simple Linear Regression)을 입출력 변수가 각각 여러 개인 다변수 데이터에 적용할 수 있도록 확장한 방법이다.  $m$ 번의 실험으로 얻어진 입력 및 응답 변수들의 값을 응답 변수 행렬( $Y$ ) 및 입력 변수 행렬( $X$ )로 표시하면 다음의 행렬식으로 표시할 수 있게 된다.

$$Y = XB + E \tag{1}$$

여기서  $B(n \times p)$ 는 입력 변수 행렬  $X(m \times n)$  및 응답 변수 행렬  $Y(m \times p)$ 의 선형 관계식을 나타내는 계수 행렬이며,  $E$ 는 오차를 나타내는 행벡터로 이루어진  $(m \times p)$  행렬이다. 오차를 최소화 소로 만드는 계수 행렬  $B$ 는 최소 자승법에 의해 다음과 같이 구해진다.

$$B = (X^T X)^{-1} X^T Y \tag{2}$$

식 (2)로 구해진 행렬  $B$ 를 이용하면 식 (1)에 의해 입력 변수 값에 대한 응답 변수의 예측 값을 계산할 수 있다. 이와 같이 MLR은 사용하기가 간편하고 입력 변수들이 갖고 있는 기여도 등을 해석하기 쉬운 장점을 갖고 있다. 반면에 몇 가지 단점들을 가지고 있는데, 첫째는 입력 변수들이 서로 선형적으로 독립적이지 못할 때는 적용되기 곤란하다. 행렬  $X$ 의 열들이 서로 선형적으로 독립적이지 못할 때, (2)에 사용되는 행렬  $X^T X$ 가 full rank를 가지지 못해 역행렬이 존재하지 않으며, 따라서 계수 행렬  $B$ 를 구할 수 없다는 것이다. 또한 선형적으로 모두 독립적이라고 할지라도 입력변수들 간의 상관관계가 매우 크면 입출력 변수 값이 조금만 변해도 계수 행렬  $B$ 값의 변화가 너무 커져서 응답 변수의 예측에 대한 신뢰도가 떨어진다 단점이 있다. 두 번째는 입력 변수의 개수  $n$ 이 관측 수  $m$ 에 비하여 비슷하거나 그리 크지 않을 경우도 마찬가지로 행렬  $B$ 의 신뢰도가 떨어져 예측 성능이 나빠지는 단점이 있다[8].

### (2) 부분최소자승법(Partial Least Squares)

부분최소자승법 역시 응답 변수 행렬  $Y$ 와 입력 변수 행렬  $X$  사이의 선형관계를 모델링하는 투영법이다. 두 행렬은 다음과 같이 나뉜다.

$$X \cong \sum_{a=1}^A t_a p_a^T = TP^T \tag{3}$$

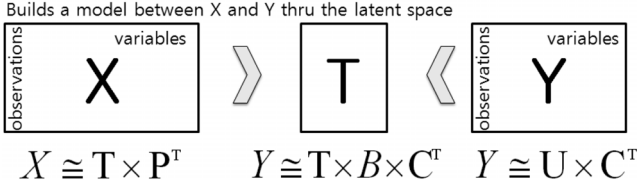


Figure 4. Indirect modeling by PLS.

$$Y \cong \sum_{a=1}^A u_a c_a^T = UC^T \quad (4)$$

$T$ 와  $U$ 는 score 행렬,  $P$ 와  $C$ 는 loading 행렬이다.  $A$ 개의 잠재변수(latent variables)를 갖는 모델은 cross validation을 이용하여 결정된다. 그리고 x-scores  $t_a$ 는  $X$  행렬( $a=1$ 일 때) 또는  $X$ 의 잔차 행렬( $a > 1$ 일 때)의 선형 조합이 된다.

$$t_a = X_{a-1} w_a, \quad X_a = X_{a-1} - t_a p_a^T \quad (5)$$

여기서  $w_a$ 는  $a$  번째 잠재변수의 가중치 벡터이다. 이는 다음과 같은 내적관계를  $T$ 와  $U$  사이의 공분산이 최대가 되도록 수행된다.

$$U \cong TB \quad (6)$$

여기서  $B$ 는 diagonal 행렬을 나타낸다. 결과적으로 PLS를 다음과 같은 예측모델로 나타낼 수 있다[9]. PLS 예측 모델에 대한 모델링 절차는 Figure 4를 참조할 수 있다.

$$Y \cong TBC^T = XW(P^T W)^{-1} BC^T \quad (7)$$

### 2.3. 스펙트럼 전처리

스펙트럼 전처리(Spectral Preprocessing)는 정확한 정량 검량을 위해서 기기조건 변화에 따른 스펙트럼의 변화를 제거 또는 보완하는데 요구되는 방법이다. 스펙트럼에 포함된 노이즈를 제거하여 보다 안정적이고 정확도가 높은 예측 모델을 만들기 위하여 이용된다. 이는 크게 MSC (Multiplicative Scattering Correction), SNV (Standard normal variate), OSC (Orthogonal Signal Correction), Savitzky-Golay법 등이 있다 [10].

#### (1) Multiplicative Scatter Correction (MSC)

산란현상은 파장에 따라 일정하지 않아 바탕선의 이동, 기울기 변화, 그리고 심할 경우 곡선 형태의 바탕선도 발생시킨다. 이를 보완하기 위해 데이터행렬의 평균 스펙트럼을 계산한 후 이를 이상 스펙트럼으로 보고 MSC를 이용하여 보정한다. 각 스펙트럼의 파장별 흡광도들과 평균 스펙트럼 흡광도와의 상관관계를 구하기 위해 회귀분석을 이용하여 기울기와 절편을 구한다. 따라서 이상적인 스펙트럼은 아래와 같다.

$$\bar{x}_j = \frac{\sum_{i=1}^m x_{i,j}}{m} \quad (9)$$

여기서,  $x_{i,j}$ 는 ( $m \times n$ ) 스펙트럼 행렬  $X$ 의 원소이며( $m$ 은 샘플 수,  $n$ 은 파장 수),  $\bar{x}_j$ 는 평균 스펙트럼( $1 \times n$ )이다. 이 평균 스펙트럼을 이용하여 각 샘플의 모든 파장에서의 흡광도 데이터  $x_i$ 에 대해 선형회귀를 취한다.

$$\text{즉, } x_i = a_i \bar{x}_j + b_i \mathbf{1} \quad (10)$$

선형회귀를 하여 구한  $a_i$  및  $b_i$ 값을 이용하여 다음 식과 같이 보정을 한다.

$$x_{i,msc} = \frac{(x_i - b_i \mathbf{1})}{a_i} \quad (11)$$

여기서,  $x_{i,msc}$ 는 샘플의 스펙트럼  $x_i$ 를 MSC로 보정한 스펙트럼 ( $1 \times n$ )이다.

MSC방법의 제약조건은 스펙트럼 변화와 농도간의 선형관계가 존재할 때만 사용할 수 있다는 점과 시료들이 화학적으로 유사할 경우에 만족할만한 결과를 얻을 수 있다는 것이다 [10].

#### (2) Standard normal variate (SNV)

SNV방법은 MSC와 마찬가지로 광산란 보정을 위한 방법으로, MSC에 비해 평균 스펙트럼과 같은 이상적인 스펙트럼이 필요하지 않으며, 대신에 각 파장에서의 흡광도를 샘플 내 전체 스펙트럼의 평균( $\bar{x}$ )과 표준편차( $w_i$ )로 정규화하여 광산란의 영향을 제거하는 방법이다. 광경로나 광원의 변동에 따른 스펙트럼의 변이 등이 SNV에 의해 보정될 수 있다고 보고되었다.

$$\bar{x}_i = \frac{\sum_{j=1}^n x_{i,j}}{n}, \quad w_i = \sqrt{\frac{\sum_{j=1}^n (x_{i,j} - \bar{x}_i)^2}{(n-1)}} \quad (12)$$

여기서,  $\bar{x}_i$ 와  $w_i$ 는 샘플  $i$ 의 모든 스펙트럼의 평균과 표준편차이다[10].

#### (3) Orthogonal Signal Correction (OSC)

전처리를 하는 이유는 농도 정보와 상관없는 구조적인 변동을 제거하기 위한 것이다. 수학적으로 이러한 전처리는 일종의 필터링이며, 이 때 성분  $y$ 는 전혀 고려되지 않는다. 따라서 스펙트럼  $x$ 에서 구조적인 노이즈를 제거할 때, (PLS 모델링의 경우) 스펙트럼의 변동이 큰 부분만 강조하게 되어 성분과 관련된 정보도 동시에 제거되는 경우가 있을 수 있다. 스펙트럼  $x$ 에서 성분  $y$ 와 상관없는 변수만 제거하면 성분과 관계가 높은 스펙트럼 성분만 남기 때문에, PLS에서 계산되

는 x-score  $t_a$ 를 성분  $y$ 와 직교시켜 성분에 직교하는 스펙트럼 성분을 계산하고 이를 원래의 스펙트럼에서 제거하면, 성분과 상관이 매우 높은 스펙트럼 성분만을 찾을 수 있다[11]. 즉 PLS의 NIPALS 알고리즘 (3)에서  $T$ 와  $Y$ 의 내적이 0이 되면 성분  $y$ 와 스펙트럼  $x$ 가 직교하게 된다. 따라서 직교 성분은  $T^* = (1 - Y(Y^T Y)^{-1} Y^T)T$ 와 같이 구할 수 있다. 이  $T^*$ 를 이용하여 성분  $Y$ 와 상관이 없는 스펙트럼 성분( $T^* \times P^T$ , OSC component)을 계산할 수 있다. 이를 원 스펙트럼에서 제거하면 상관이 높은 스펙트럼 성분( $X' = T^* \times P^T$ )을 얻을 수 있다.

(4) Savitzky-Golay (SG)

Savitzky-Golay 방법에 의한 신호 평활법은 다음과 같은 식에 의해서 스펙트럼 측정시스템에서 기인되는 노이즈를 감소시키는 방법이 된다.

$$Y_i^* = \frac{\sum_{j=1-n}^{i+n} Y_j c_j}{N} \tag{14}$$

위 식에서  $n$ 은 1 이상의 값을 가지며 이는 Savitzky-Golay의 방법의 적용에 있어서 이동 창의 크기( $2n+1$ )를 결정한다. 이 식은 어느 한 측정값의 잡음 신호를 제거하기 위하여 그 측정값을 중심으로 이동 창 내의 이전과 이후에 측정된 측정값들을 참조하는 방법을 취하게 한다. 식에서  $Y_i^*$ 은 잡음이 제거된  $i$  번째 측정값이 되고,  $Y_j$ 는  $Y_i^*$ 를 구하기 위해서 사용된 잡음을 제거하기 전의 측정값들을 의미하며,  $c_j$ 는 Savitzky-Golay의 방법의 적용에 있어서 잡음의 제거를 위해 기준이 되었던 함수관계와 최소자승의 원리에 따라 구해진 convolution integer이다[12]. 식에서  $N$ 은 일종의 scaling 상수이다.

3. 실험 및 결과

3.1. 실험 방법

본 실험에 사용된 시스템을 구체적으로 설명하고자 한다. 먼저 본 실험에서 NIR 기기는 Varian 660-IR을 사용하였다. 이 기기의 NIR 영역 광원, Beam Splitter, 검출기는 각각 Tungsten-Halogen Lamp, Near-IR Quartz, Lead Selenide를 사용하였다. 그리고 바이오 에탄올 표준샘플 측정 accessory로 Near-IR Fiber Optic Liquid Probe Tip을 사용하였다. 측정 옵션은 파장범위를 11,000~4,000  $cm^{-1}$ , resolution을 4.0  $cm^{-1}$ 로 셋팅하고 16번 측정하여 이것을 평균하여 한 개의 스펙트럼으로 나타내도록 설정하였다. 이 때 1개의 샘플당 약 38 s의 측정시간이 소요되었다. 스펙트럼의 수집과 베이스 라인 수정 등은 Varian의 FT-IR Resolution Pro 프로그램을 이용하였다.

농도를 알고 있는 바이오 에탄올 혼합물을 만들기 위해 시약 급의 물질을 사용하여 표준샘플을 제조하였다. 이때 사용된 시약은 99.8% (v/v)메탄올, 99.9% (v/v)에탄올, 99.9% (v/v)벤젠, 그리고 초순수 증류수를 사용하였으며, 올레핀 계열의 불순물로는 98.0% (v/v) 1-Decene을 사용하였다.

에탄올 혼합물 성분 예측을 위한 검량모델 개발 및 평가를 위해서 Minitab 통계 소프트웨어 내 실험계획법 중 혼합물 설계를 이용하여 혼합물의 성분비를 결정하였다. 이때 5개 물질의 부피비(%)의 상한과 하한을 ASTM의 D4806에 근거하여 설정 하였다. 총 47개의 sample의 혼합물의 성분 함량이 결정되었고 이에 따라 47가지의 샘플을 제조하였다. 실제 성분비는 에탄올이 81~92%, 나머지 불순물의 성분비는 약 1~5%이다.

3.2. 실험결과

47개의 표준 샘플의 근적외선 분광분석 스펙트럼 전체를 Figure 5에 겹쳐 나타내었다. 바탕선 보정에도 불구하고 그림에서와 같이 연속 측정에 따라 여러 가지 기기 영향으로 인한 스펙트럼 바탕선의 이동이 눈에 띄며, 추가적인 보정이 필요한 것으로 판단된다. 따라서 앞에서 제시한 4 가지의 전처리 방법으로 이를 수행하였고 그 결과를 Figure 6~9에 나타내었다.

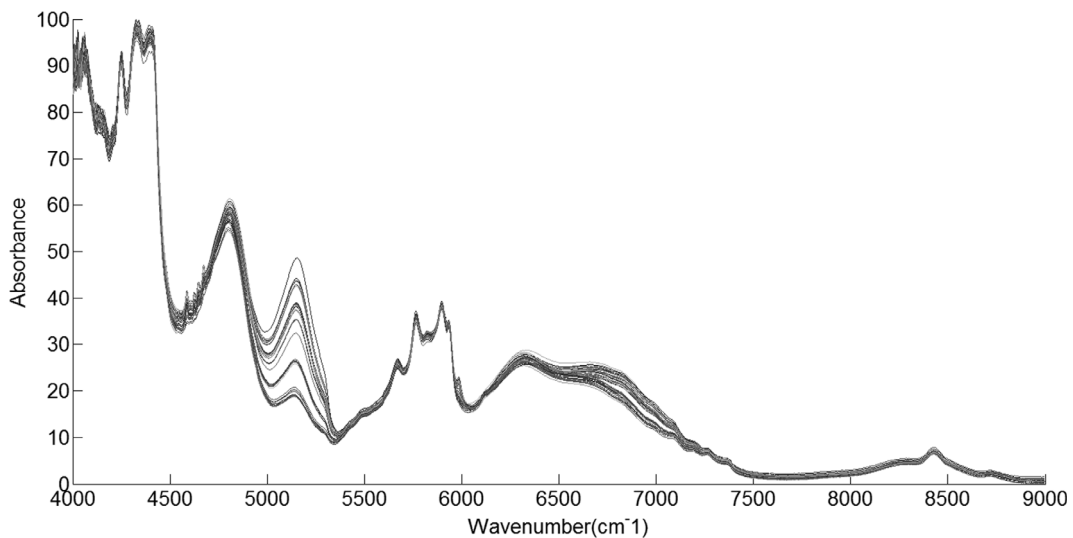


Figure 5. Raw NIR spectrum of standard samples.

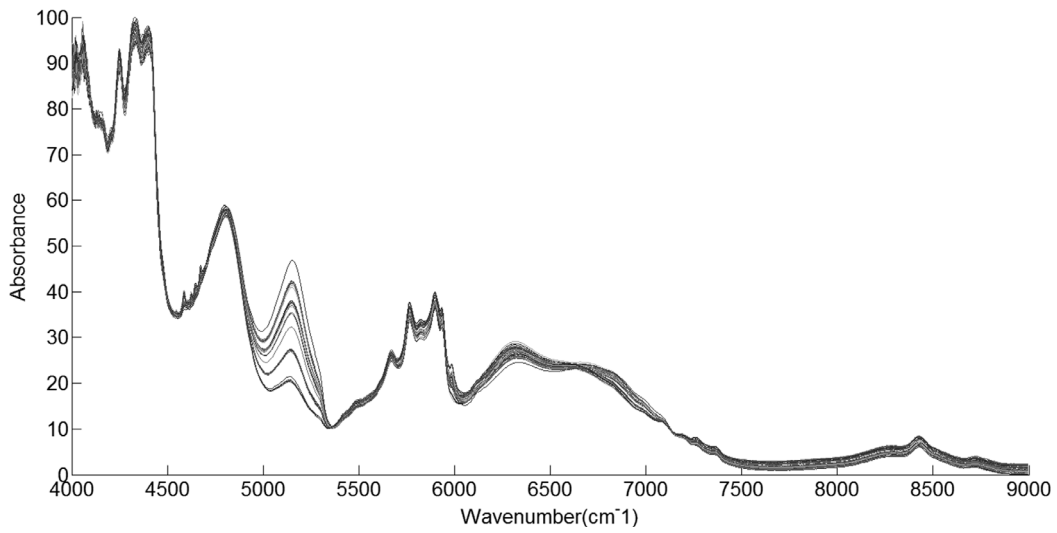


Figure 6. NIR Spectrum corrected by MSC.

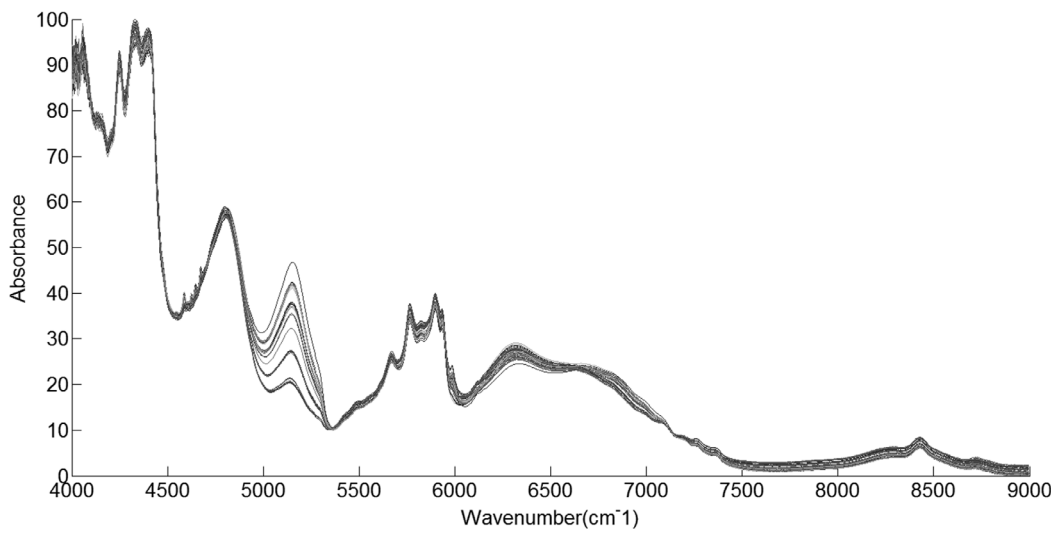


Figure 7. NIR Spectrum corrected by SNV.

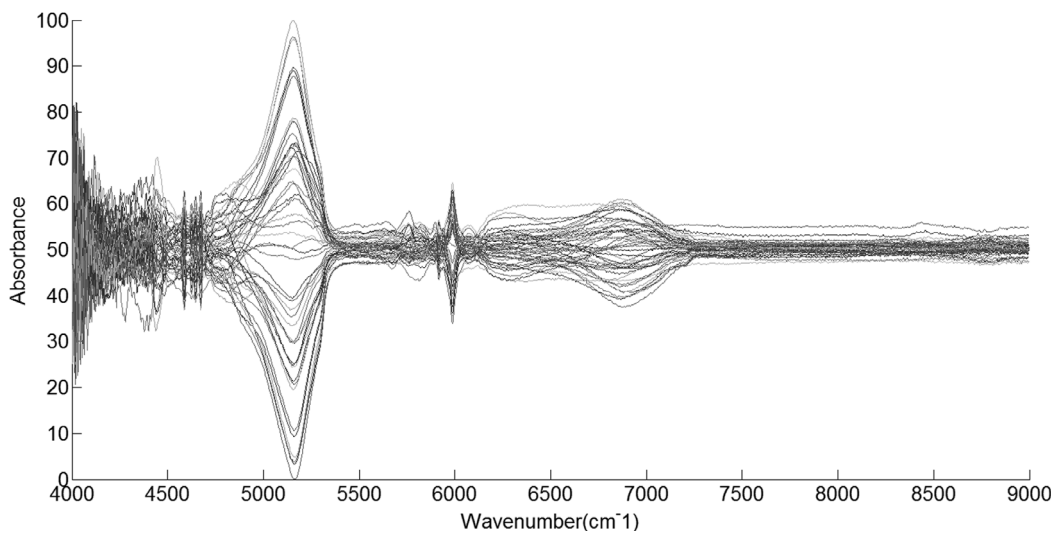


Figure 8. NIR Spectrum corrected by OSC.

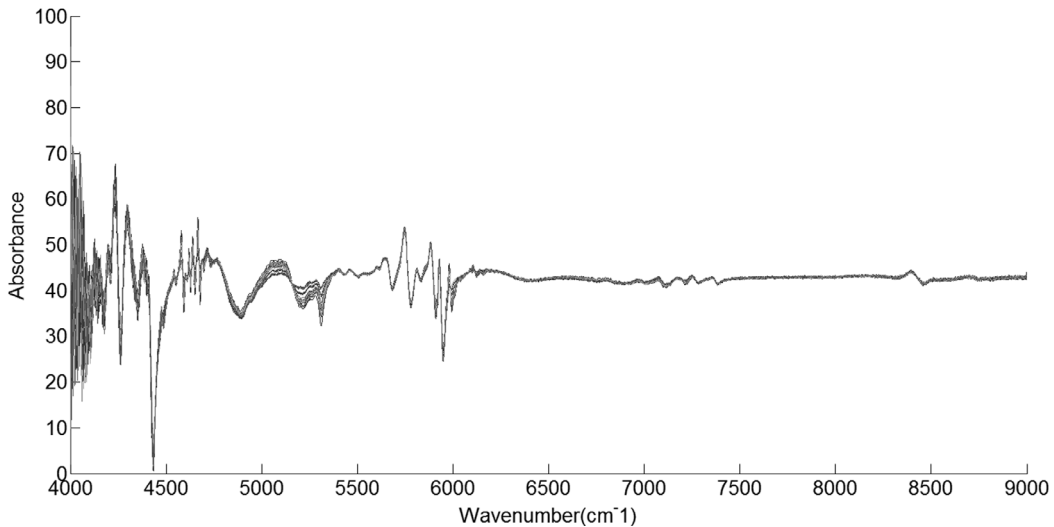


Figure 9. NIR Spectrum corrected by Savitzky-Golay preprocessing.

MSC나 SNV를 이용하여 보정을 한 경우 스펙트럼 형태가 모두 Raw 스펙트럼과 유사함을 알 수 있으며, Figure 5~7의 wavenumber 6,500~7,000 cm<sup>-1</sup> 영역을 비교해 볼 때, MSC와 SNV 모두 어느 정도 바탕선의 보정이 이루어졌음을 확인할 수 있다.

OSC의 경우, Figure 8의 wavenumber 4,900~5,300 cm<sup>-1</sup>에서 확인할 수 있듯이 스펙트럼의 peak 부분이 상대적으로 강조됨을 확인할 수 있다. 하지만 각 파장에서의 스펙트럼이 강조되어, 바이오 에탄올의 성분 결정을 위한 peak 영역 뿐 만 아니라 각 실험의 불필요한 노이즈 영역 또한 두드러짐을 확인할 수 있다. 하지만 Figure 9의 Savitzky-Golay 전처리법의 경우, 성분 결정에 필요한 데이터가 각각의 농도에 따라 강조됨과 동시에 바탕선 보정이 이루어지며, 노이즈 성분에 대한 영역들은 전처리 결과 사라짐을 보였다.

여러 물질로 이루어진 혼합물의 경우 NIR 스펙트럼이 매우 복잡해서 IR의 경우와 같이 검량선을 만들기 위해 각 성분의 특정 peak를 선택할 수 없기 때문에, PLS 법을 이용하여 각각 전처리한 스펙트럼의 검량 모델을 만들고 전처리 효과를 정량적으로 분석하였다. 일반적으로 검량 모델의 성능측정은 결정계수 R<sup>2</sup>의 값을 가지고 가늠한다. ASTM 등의 표준에서는 그 값이 0.990 보다 같거나 클 것을 권장하므로 R<sup>2</sup>이 0.990 이상 일 경우 PLS의 인자 수를 이용하여 전처리 효과를 분석하였다.

Umetrics의 Simca-P v.8.0을 사용하여 5개의 조성 전체를 예측하는 모델과 각각의 조성 1개를 예측하는 모델 등 총 6개의 PLS 모델을 구하였으며, 이 때 2593개의 파장에서의 흡광도를 입력 변수 x로, 5개의 혼합비를 응답 변수 y로 각각 사용하였다. 각 전처리 방법의 검량식 모델링의 결과는 아래 Table 2 (PLS 잠재변수 수), 6 (R<sup>2</sup> 값)과 같다.

Table 2와 같이 SNV, MSC, OSC 전처리 방법은 Raw 스펙트럼의 결과와 비교했을 때 PLS 인자 수의 변화가 거의 없거나 일부 늘어나는 것을 볼 수 있는데, 이는 앞서 설명한 대로

불완전한 바탕선 보정(MSC와 SNV)과 노이즈의 증폭(OSC)에 의한 영향으로 인한 것으로 보인다. 그에 비해 Savitzky-Golay 전처리는 6가지 경우 모두 PLS 잠재변수 수를 최소화할 수 있는 효과적인 방법으로 보인다.

Table 3에서 보듯이 주성분인 에탄올은 물론, 메탄올, 물, 올레핀 그리고 벤젠 등 바이오에탄올 제조공정에서 발생하는 주요 불순물의 경우에도 모든 PLS 검량 모델은 정확한 농도 예측 값을 제공할 수 있음을 알 수 있다. 하지만 실제 연속 및 회분식 공정인 바이오에탄올 공정에서 실시간 측정용으로 사용할 경우에는, Savitzky-Golay 전처리법을 이용하여 만든 PLS 모델을 제외한 다른 모델들은 많은 수의 PLS 잠재변수가 필요하게 된다. 이럴 경우에 검량선을 모델링할 때는 물론, 공정의 상태 변화로 PLS 모델을 갱신할 필요가 있을 경우에도 상당한 노력이 필요하다. 이러한 관점에서 볼 때, 장기간의 실시간 측정용으로 사용될 경우에 Savitzky-Golay 전처리법을 이용하여 만든 PLS 모델이 더 선호된다고 할 수 있다. 이 모델을 사용하여 얻은 상관계수 R<sup>2</sup>의 값이 0.99 이상이므로 모든 성분에 대해 최종 제품의 품질관리를 위한 실험실 분석도 대체할 정도의 정확도를 보여주고 있다. 개별 성분별 모

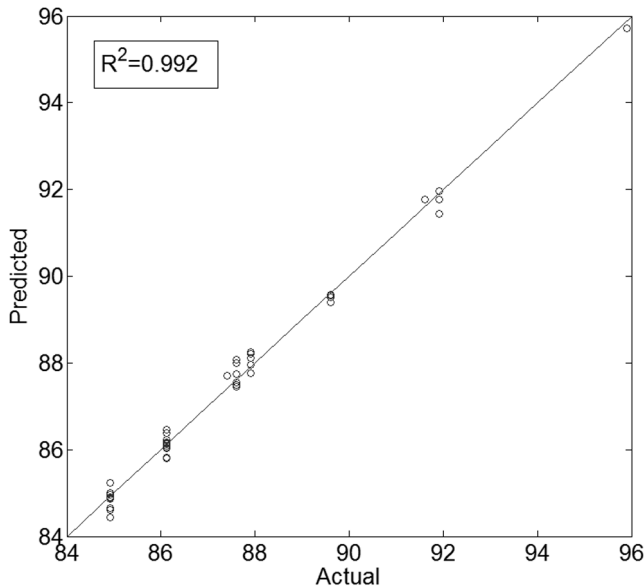
Table 2. Regression results in terms of the number of latent variables in PLS models

Component	The number of latent variables in PLS models				
	Raw	SNV	MSC	OSC	Savitzky-Golay
All component	27	27	21	20	12
Ethanol	14	13	13	13	5
Methanol	14	13	13	12	5
Water	13	12	12	10	5
Olefin	13	14	14	13	5
Benzene	13	14	14	12	5

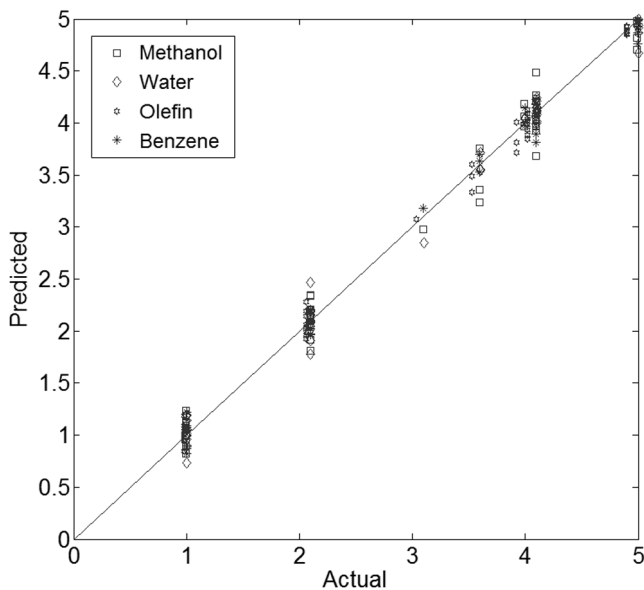


**Table 3.** Regression results in terms of  $R^2$  values

Component	$R^2$ value				
	Raw	SNV	MSC	OSC	Savitzky Golay
All component	0.990	0.992	0.993	0.991	0.992
Ethanol	0.998	0.993	0.993	0.993	0.993
Methanol	0.993	0.992	0.991	0.991	0.992
Water	0.991	0.994	0.993	0.992	0.990
Olefin	0.997	0.991	0.993	0.991	0.991
Benzene	0.991	0.995	0.997	0.994	0.995



**Figure 10.** Predicted versus actual ethanol concentration in standard samples (using Savitzky-Golay preprocessing).



**Figure 11.** Predicted versus actual impurity concentrations in standard samples (using Savitzky-Golay preprocessing).

텔을 이용하여 Figure 10, 11에 각 성분별 실제 값 대비 예측 값의 그래프를 나타내었다.

#### 4. 결론

본 논문에서 근적외선 분광분석법과 화학계량학을 이용하여 설계한 분석 시스템은 바이오 에탄올의 주성분인 에탄올의 농도는 물론, 제조공정 중 발생할 수 있는 여러 불순물들의 농도 또한 정확히 예측 가능함을 보였다. 또한 스펙트럼 데이터를 이용한 데이터 전처리의 효과 및 최적 전처리 방법을 규명하였다. Savitzky-Golay 전처리를 이용한 스펙트럼으로 검량 모델링을 할 경우 모델 안정성이 향상되고 효과적인 공정관리가 가능함을 보였다.

본 연구를 기반으로 설계한 분광분석을 이용한 바이오연료 실시간 모니터링 시스템은 ASTM 등의 표준에서 제시하는 기존의 분석화학 방법보다 빠른, 약 1분 내로 측정 가능하기 때문에 측정값에 따라 제조공정 중 조업조건 등을 바뀌어도, 공정을 항상 최적의 상태로 운전이 가능하게 할 수 있다. 또한 본 논문에서는 광섬유 probe를 사용하여 분석이 가능함을 보였으므로 분광분석 장치가 원거리에도 in-line으로 운영이 가능하기 때문에 예측값을 직접 피드백 제어에 사용하여 실시간 조업조건의 변경이 가능하다.

그리고 한 가지 분석 장비로 하나의 성분만을 측정할 수 있는 기존의 분석방법과는 달리, 하나의 분석 시스템으로 여러 성분을 동시에 측정이 가능하기 때문에 분석에 관련된 제반 비용의 절감도 가능하다는 장점이 있다.

#### 감사

이 논문은 농림수산식품부의 지원과 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(2010-0003056)이며, 이에 감사드립니다.

#### 참고문헌

1. Kim, H. J., and Min, S. K., "United Nations Framework Convention on Climatic Change and Korean's Countermeasures," Samsung Economics Research Institute-CEP Information, 294, May, 2001.
2. Kim, H. N., Kang, H. C., and Jung, H. S., "Necessity and Expected Effect of Green Technology Agreement," Samsung Economics Research Institute-Issue Paper, June, 2010.
3. Kim, H., Lee, E. S., Kim, W., Suh, D. J., and Ahn, B. S., "Material and Heat Balances of Bioethanol Production Process by Concentrated Acid Saccharification Process from Lignocellulosic Biomass," *Clean Technology*, **17**(2), 156-165 (2011).
4. ASTM D4806-10, American Society for Testing & Materials, 2010.
5. Ahn, J. W., Woo, Y. h., and Kim, H. J., "Determination of Water Content in Alcohol by Portable Near Infrared (NIR)

- System,” *Analytical Science & Technology*, **16**(2), 95-101 (2003).
6. Cho, N. W., Jeon, S. M., Kang, S. H., and Lee, D. H., “A Study of Identification Test Method for Fire Resistive Coatings in Near-Infrared Spectroscopy,” *Journal of Korean Institute of Fire Science and Engineering*, **11**, 192-197 (2009).
  7. Chung, H. C., and Kim, H. J., “Near-Infrared Spectroscopy: Principles,” *Analytical Science & Technology*, **13**(1), 138-151 (2000).
  8. Kim, S. Y., Lee, B. S., Chung, C. B., and Choi, S. H., “Multivariate Statistical Analysis Approach to Predict the Reactor Properties and the Product Quality of a Direct Esterification Reactor for PET Synthesis,” *Journal of Control, Automation, and Systems Engineering*, **11**(6), 551-552 (2005).
  9. Ferrer, A., Aguado, D., Vidal-Puig, S., Prats, J. M., and Zarzo, M., “PLS: A Versatile Tool for Industrial Process Improvement and Optimization,” *Appl. Stoch. Model. Bus.*, **24**, 551-557 (2008).
  10. Ryu, D. S., Hwang, L. G., and Noh, S. H., “Preprocessing Techniques on VIS/NIR Spectral Data for Non-Destructive Quality Evaluation of Fruits,” Proceedings the KSAM 2000 Winter Conference, The Korean Society of the Agricultural machinery, 451-456 (2000).
  11. Wold, S., Antti, H., Lindgren, F., and Ohman, J., “Orthogonal Signal Correction of Near-infrared Spectra,” *Chemo. Intell. Lab. Sys.*, **44**(1-2), 175-185 (1998).
  12. Cho, J. H., “Enhancement of Signal-to-noise Ratio of Biological Signal by Factor Analysis and Comparison to Those by a Conventional Method,” *SWU J. Pharm. Sci.*, **22**, 38-46 (2005).