

# 근적외선을 이용한 사과와 당도예측모델 개발과 비교<sup>+</sup>

Comparison of Statistical Methods for Calibrating Soluble Solid Measurement  
from NIR Interactance in Apples

이강진\*, W. R. Hruschka\*\*, J. A. Abbott\*\*, 노상하\*\*\*, Yud-Ren Chen\*\*

정회원

정회원

K. J. LEE

S. H. Noh

## 1. 서론

농산물의 선별작업의 관심 대상이었던 외부적인 요인, 즉 색택이나 형상, 크기 등을 판별하기 위한 기존의 연구 필요성이 점차 내부적인 품질요인까지 확대되고 있으며, 여러 연구분야중에서도 실용화가 가장 빠르다고 판단되는 근적외선분광법에 대한 관심은 이러한 요구와 더불어 과실의 내부 품질요인인 맛이나 부패, 치감 등의 판정을 위해 점차적으로 높아지고 있다.

1960년대에 개발된 근적외선 분광법의 역사는 그리 길지는 않지만 의약, 섬유, 화학 등의 공업분야에서는 공인된 정량정성법으로서 널리 사용되고 있으며, 농업분야에서도 분체나 액체 상태에서의 시료분석에 사용된 지 오래지만 산물상태의 고체시료의 분석에 응용된 역사는 십여년에 불과하다. 그러나 과실 등에 있어서 산물상태가 아닌 일부분을 절취하고 그 부분만을 측정하여 분석하던 것이 분석 기기의 발달과 광학기술의 발달로 완전한 산물상태의 시료를 측정할 수 있게 되어, 시료의 전처리과정에서 발생하던 화학적인 변화가 없으면 비파괴검사라고 규정하던 용어 자체도 물리적인 변화까지 배제할 수 있게 되는 기술적인 발전이 뒤따랐다. 이 연구가 먼저 시작된 외국에서는 이를 과실의 선별에 적용하여 당도 등의 내부 품질요인을 판정할 수 있는 과실 선별기도 출현하게 되었으나 아직까지 국내에서는 이와 관련된 기초적인 실험에 그치고 있는 실정이다.

본 연구에서는 과실의 내부 품질요인중 인간이 느끼는 맛과 관련된 인자인 당도를 객관적이고 과학적으로 판정할 수 있는 비파괴 과실 품질판정장치를 개발함에 있어서 기초실험에서 나타난 결과를 실제 장치를 개발하는데 적용할 수 있도록 현재 개발된 여러 센서들을 올바르게 선별할 수 있는 관점에서 출발하기 위해, 과실의 당도예측모델 개발에 사용되고 있는 여러 분석방법들을 이용하여 새로운 당도예측모델을 개발하고, 그 결과를 비교하였다.

\* 이 연구는 농촌진흥청 농업기계화연구소와 미국 농무성 농업연구청(USDA, ARS) 사이에 수행되고 있는 한미국제공동연구과제인 "비파괴 과실품질판정기술 개발(Non-destructive quality evaluation of fruits)"의 일환으로 수행된 것임.

\* 농촌진흥청 농업기계화연구소

\*\* U. S. A., USDA, ARS, NRI, Instrumentation & Sensing Laboratory

\*\*\* 서울대학교 농공학과 농업기계전공

## 2. 재료 및 방법

1996년 9월 4일에 수확된 Philadelphia산 사과(*gala*) 400개를 공시재료로 1996년 10월 26일부터 11월 8일까지 공시재료의 가시광선 및 근적외선 흡광스펙트럼(400~2498nm)과 가용성 고형물 함량(당도) 및 경도, 중량 등을 측정하였다.

측정된 흡광스펙트럼과 당도와의 상관관계를 구명하기 위하여 광을 이용한 정량정성분석에서 널리 사용되고 있는 다중회귀분석(Multiregression)을 비롯하여, 부분최소제곱회귀법(Partial Least Square Regression, 이하 PLS)과 신경회로망 기법(Artificial Neural Network, 이하 ANN)을 도입하여 최적의 파장을 선정하고 최적 당도예측모형을 개발한 후 이를 검증하고, 또한 각각의 기법에 대한 당도예측모형을 비교하여 과실의 비파괴 당도판정용 근적외선 센서개발에 활용하고자 하였다.

### 가. 공시재료의 흡광스펙트럼 및 물성측정

사과의 가시광선 및 근적외선 흡광스펙트럼을 측정하기 위하여 NIR 6500 system을 이용한 실험장치를 그림 1과 같이 구성하였으며, 이 때 광화이버방식의 모듈을 사용하여 반사율을 측정한 후, 이를 다시 흡광도로 환산하여 사용하였다. 흡광스펙트럼은 400nm에서 2498nm까지 측정하였으나, 사용된 광화이버 모듈의 특성상 1900nm 이상에서는 광화이버에 의해 흡수되는 근적외선의 양이 증대되고 잡음이 증폭되어 출력되므로, 여기서 나오는 신호를 올바른 신호로 판단할 수 없어 1800nm를 넘는 파장은 분석시에 고려되지 아니하였다. 광화이버 모듈의 광 경로는 그림 2와 같다.

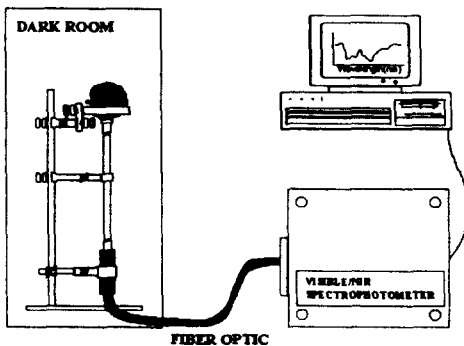


Fig 1. Schematic of visible/nir reflectance measurement system

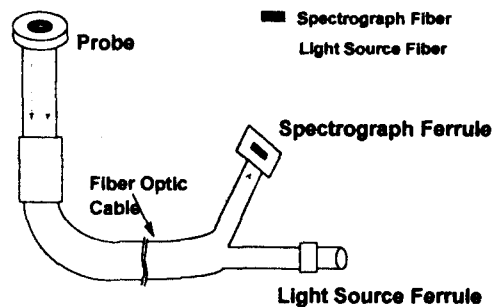


Fig 2. Fiber optic bifurcated cable

당도는 ATAGO제(Japan) 굴절계를 사용하여 측정하였고, 경도는 Universal Testing Machine을 이용하여 압축에 의한 방법으로 측정하여 가장 응력이 높은 지점을 채택하였다.

모든 실험은 사과와 마주하는 적도부위에서 이루어졌으며 한 사과당 두 개씩, 총 400개의 시료에 대해 800개의 데이터를 획득하였다. 당도판정모델의 개발을 위해 한 쪽 부위의 400개 데이터를 이용하였고, 이를 검증하기 위한 미지시료로서는 반대편 200개 데이터를 이용하였다.

#### 나. 당도판정모델의 개발을 위한 파장의 선정 및 분석

당도판정모델의 개발을 위해 전통적으로 이용되어온 다중회귀분석법을 비롯하여, 근적외선 분광분석분야에 최근에 이용되기 시작한 부분최소제곱회귀법(PLS) 및 신경회로망기법(ANN) 등을 도입하였다.

각각의 기법을 도입함에 있어 분석대상 파장역 및 전처리는 ① 400~1800nm,  $\log(1/R)$ , gap 8nm(176 파장) ② 400~1800nm,  $d^2\log(1/R)_{\text{gap}=30\text{nm}}$ , gap 8nm(176파장) ③ 800~1100nm,  $\log(1/R)$ , gap 8nm(38파장) ④ 800~1100nm,  $d^2\log(1/R)_{\text{gap}=30\text{nm}}$ , gap 8nm(38파장) 의 4가지 범주로 나누어 수행하였는데, 여기서 R 이란 반사율(Reflectance)을 말한다.

다중회귀분석을 위해서 SPAN(USDA, ARS, ISL)을 이용하였으며, 윈도우즈용 SAS 6.11로서 결과를 검증하였다. 또한 PLS 회귀법은 Galactic사의 Grams/32를, 신경회로망은 Neural Ware사의 NeuralWorks Professional II/PLUS를 이용하여 분석하였다.

#### 다. 다중회귀분석을 이용한 당도예측모델(다중회귀모델)의 조건

당도예측모델의 개발에 필요한 파장수를 증가시키면서 그 때의 실측치와 예측치 사이의 결정계수의 변화를 관찰하고, 결정계수( $R^2$ )의 값이 0.91이상이 될 때의 파장수를 이용한 것을 최적 모델로 하였다.

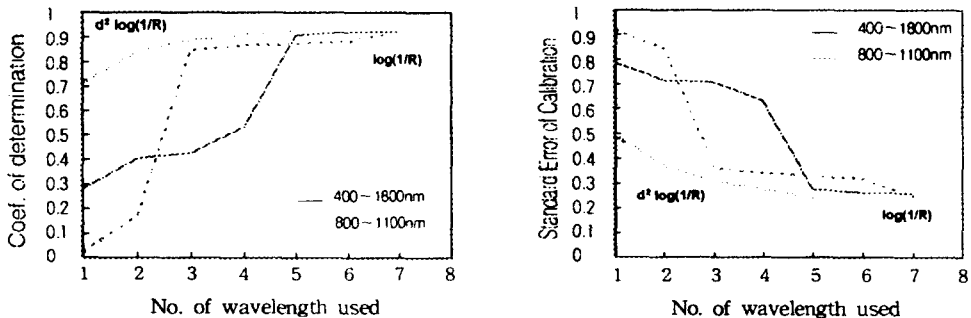


Fig 3. The change of coef. of determination and standard error of calibration(SEC) in calibration process by no. of wavelength used

라. PLS에 의한 당도예측모델(PLS 모델)의 조건

스펙트럼분석에 필요한 요인수를 증가시키면서 PRESS(Prediction Residual Error Sum of Squares)의 변화를 관찰하여 그 값이 수렴될 때의 요인수를 최적의 요인수로 보고, 이를 이용한 모델을 미지시료의 예측에 적용하였다.

마. ANN에 의한 당도예측모델(ANN 모델)의 조건

스펙트럼의 정량분석은 비선형적인 문제이므로 하나의 은닉층을 두었으며, 입력층과 출력층은 흡광스펙트럼과 당도로 하였다. 학습에 있어 전달함수는 시그모이드 함수를 사용하였으며 학습률(learning rate)과 모우멘텀(momentum)은 각각 0.5, 0.4로 하였다.

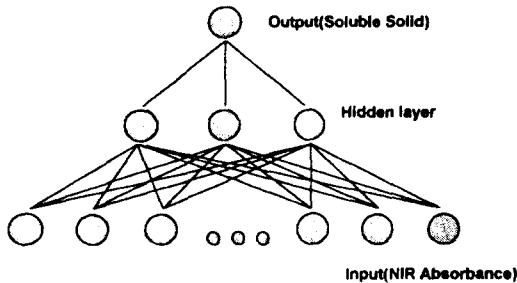


Fig 4. Architecture of ANN for calibrating Soluble Solid measurement

Table 1. The condition of ANN modeling

Wavelength range	input layer	hidden layer	output layer	Transfer function	RMS error
	data	no. of nodes			
400~1800nm	log(1/R)	177	13	sigmoidal activation function $Out = \frac{1}{1 + e^{-x}}$	0.0420
	$d^2 \log(1/R)$	177	7		0.0263
800~1100nm	log(1/R)	30	13		0.0390
	$d^2 \log(1/R)$	30	7		0.0277

learning rate : 0.5 momentum : 0.4

### 3. 결과 및 고찰

가. 사과와의 흡광스펙트럼 및 물성

400nm부터 1800nm까지 측정된 사과와의 흡광스펙트럼은 그림 5와 같으며, 당도, 중량, 경도 등의 측정 결과는 표 2와 같다.

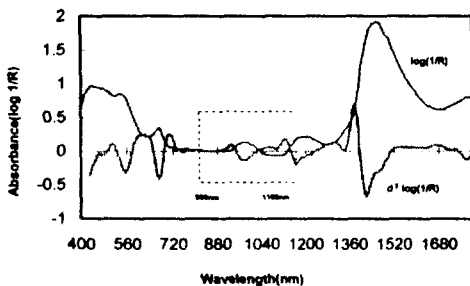


Fig 5. Visible/NIR absorbance of Apple

Table 2. Characteristics of Apples

공시(재료) : 사과 ( <i>Fallosidaphnia gala</i> )	당도(brix)			중량(g)			경도(n)		
	Max.	Min.	Ave.	Max.	Min.	Ave.	Max.	Min.	Ave.
Calibration Set (n = 400)	15.4	9.1	11.75	202.4	105.8	144.0	80.24	42.36	60.36
Prediction Set (n = 200)	14.4	8.4	11.76						

#### 나. 다중회귀모델에 의한 당도예측

전파장역을 대상으로 gap을 30nm로한 2차미분 전처리를 하여 784, 1056, 992, 808, 872nm의 5개 파장을 선택하였을 경우,  $R^2$ 는 0.93, SEC는 0.24bx이었으며 이를 이용하여 미지시료의 당도를 예측했을 때  $R^2$ 는 0.91, bias -0.02bx, SEP는 0.28bx로 나타났다. 미분처리를 하지 않은 경우는 미지 시료의 예측에 있어 전파장역에서보다 800nm에서 1100nm까지의 파장을 이용하여 예측한 경우에 오차를 더 줄일 수 있었다. 이 때 904, 880, 1032, 824, 1048, 1096nm의 7개 파장을 선택하였을 경우의  $R^2$ 는 0.93, SEC는 0.25bx로 나타났으며 이를 이용한 미지시료의 당도 예측에서도  $R^2$ 는 0.90, bias -0.03bx, SEP는 0.27bx로 양호하였다.

#### 다. PLS 모델에 의한 당도예측

PLS 모델에 있어서 400nm부터 1800nm까지의 전파장을 이용하여 분석한 것보다 파장의 일부를 선택하여 분석한 경우가 훨씬 양호하였으며, 2차미분의 파장전처리를 한 경우와 파장 전처리를 하지 않은 경우를 비교해 볼 때, 대체적으로 파장 전처리를 하지 않은 경우가 더 양호한 결과를 보였다. 800nm와 1100nm 사이의 파장역에서 8개의 요인을 선택하였을 경우,  $R^2$ 는 0.94, SEC는 0.23bx이었으며 이 모델을 이용하여 미지시료의 당도를 예측했을 때의  $R^2$ 는 0.92, bias -0.02bx, SEP는 0.25bx로 나타났는데, 이는 개발된 각 모델중 가장 우수한 결과였다.

#### 라. ANN 모델에 의한 당도예측

스펙트럼의 전처리를 하지 않은 모델에 있어 일부 파장역을 입력데이터로 하는 것보다 전 파장역의 흡광도를 입력데이터로 하는 것이 오히려 양호하였다. 또한 입력층인 흡광 스펙트럼의 전처리를 하지 않은 모델보다 2차미분을 한 모델이 모두 우수한 결과를 보였으며, 학습의 횟수도 감소시킬 수 있었다. 2차미분 전처리된 800nm에서 1100nm까지의 스펙트럼 데이터와 옵셋보정용 bias 등으로 구성된 39개의 입력층과 7개의 은닉층, 출력층을 당도로 하였을 때 당도예측모델의 개발을 위한 학습의 결과는  $R^2$  0.93, SEC 0.25bx로 나타났으며, 이를 이용한 미지시료의 당도 예측에 있어서의  $R^2$ 와 bias, SEP 각각은 0.90, -0.03bx, 0.29bx로 나타났다.

#### 마. 당도예측모델의 비교

각각의 당도예측모델의 개발 및 검증시의 결정계수, 표준오차는 표 3과 같다.

Table 3. The result of calibration and prediction using each model.

과 장 범 위			400~1800nm		800~1100nm	
과 장 전 처 리			log(1/R)	d'log(1/R)	log(1/R)	d'log(1/R)
입 력 데 이 터 수			176	176	38	38
모델별 조 건	Multiregression	선택과장수	.7	5	7	5
	PLS regression	최적요인수	20	19	8	9
	ANN	출력층 은닉층 입력층	1 13 177	1 13 177	1 7 39	1 7 39
당도예측 모델작성	R <sup>2</sup>	Multiregression	<b>0.9240</b>	<b>0.9335</b>	<b>0.9289</b>	<b>0.9306</b>
		PLS regression	0.8612	0.8611	<b>0.9380</b>	<b>0.9378</b>
		ANN	0.8193	<b>0.9119</b>	0.7795	<b>0.9348</b>
	SEC(bx)	Multiregression	0.2577	<b>0.2404</b>	<b>0.2492</b>	<b>0.2457</b>
		PLS regression	0.3459	0.3379	<b>0.2304</b>	<b>0.2307</b>
		ANN	0.5575	0.3687	0.4649	<b>0.2491</b>
당도예측모델을 이용한 미지시료의 예측결과	R <sup>2</sup>	Multiregression	0.8935	<b>0.9065</b>	<b>0.9021</b>	0.8992
		PLS regression	0.8860	0.8338	<b>0.9236</b>	<b>0.9177</b>
		ANN	0.8243	0.8856	0.7507	0.8992
	bias(bx)	Multiregression	<b>-0.0190</b>	-0.0220	-0.0330	-0.0230
		PLS regression	<b>0.0172</b>	-0.0206	<b>-0.0198</b>	<b>-0.0089</b>
		ANN	-0.0352	-0.0274	-0.0593	-0.0265
	SEP(bx)	Multiregression	0.2980	<b>0.2800</b>	<b>0.2730</b>	0.2930
		PLS regression	0.3101	0.3729	<b>0.2527</b>	<b>0.2634</b>
		ANN	0.4906	0.3091	0.4622	0.2903

#### 4. 요약 및 결론

본 연구는 과실의 내부품질요인을 비파괴적으로 판정하기 위한 장치 개발에 필요한 광센서를 선별하기 위하여, 현재 이용되고 있는 여러 가지 분석적 방법을 도입하여 적절한 당도 예측모델을 개발함과 동시에 각각의 모델을 비교검토하기 위해 실시되었다. 연구결과를 요약하면 다음과 같다.

가. 다중회귀분석에 의한 모델개발에 있어서 과장의 전처리를 하지 않은 경우 7개, 2차미분전처리를 한 경우 5개의 과장만을 이용하였을 때 예측치와 실측치사이의 결정계수를 0.90 이상으로 증가시킬 수 있었으며, 2차미분 전처리를 한 예측모델을 이용하여 미지 시료를 예측했을 때의 결정계수와 바이어스, SEP 각각은 0.9, -0.02bx, 0.28bx였다.

나. PLS 회귀법에 의해 개발된 모델은 전과장을 이용하기 보다 선택된 일부 과장과 과장전처리를 하지 않음으로서 예측의 정확성을 증진시킬 수 있었다. 800nm에서 1100nm까지의 과장을 이용하여 개발된 모델의 예측에 있어서의 결정계수는 0.92였으며, 이 때의 바이어스와 SEP는 -0.02bx와 0.25bx로 개발된 각 모델중 가장 우수한 결과를 나타냈다.

다. 신경회로망에 의한 모델개발에 있어서는 2차미분 전처리를 해야만 정확성을 높일 수 있었다. 800nm에서 1100nm까지의 일부 파장역에서 파장 전처리를 하지 않은 모델의 미지시료 예측시의 결정계수와 SEP는 0.75, 0.46bx였지만 2차미분 전처리를 한 모델의 결정계수와 SEP는 0.90, 0.29bx로 훨씬 양호한 결과를 보였다.

라. 과실 비파괴 품질판정기를 개발할 경우, 다중회귀모델을 적용한다면 5개 내지 7개 이상의 광다이오드와 대역필터를 이용해야하고, ANN 모델이나 PLS 모델을 적용할 경우엔 광다이오드어레이 센서를 선별해서 사용해야 한다. 본 연구의 결과로 볼 때 여러 개의 광다이오드와 필터를 사용하여 품질판정기를 개발하는 것 보다, 2차미분 등의 전처리가 필요하지 않은 PLS모델을 적용하고 적절한 광다이오드어레이 센서를 선별한다면 시작기의 개발에 소요되는 시간과 노력을 훨씬 절감할 수 있을 것으로 판단되었다.

## 5. 참고문헌

1. David, C. S. 1992. Near Infrared Analysis of Soluble Solids in Peaches. ASAE paper No. 92-7056.
2. Kawano, S., H. Watanabe and M. Iwamoto. 1992. Determination of sugar content in intact peaches by near infrared spectroscopy with fiber optics in interactance mode. J. Jpn. Soc. Hort. Sci. 61(2):445-451.
3. 西村 功ら. 1990. 品質評價技術の實用化. 青果物の貯藏・流通の新技术.
4. 市川 哲生ら. 1992. 近赤外重回歸分析装置とその應用. 鳥律評論 別刷. 49(1・2):11-18
5. 木村美紀夫. 1993. 近赤外線による果實非破壞検査システムの取構り組みと課題. 果實日本. 48(2):31-35.
6. 河野澄夫ら. 1989. 近赤外分光分析法による温州ミカンの糖度, 酸度の非破壞計測. 園藝學會雜紙. 58(別 2):604-605
7. 天間毅ら. 1990. 近赤外分光法によるリンゴ糖度の測定. 第6回 非破壞計測シンポジウム. 98-102.
8. 김형욱. 1995. 근적외선 분광분석법에 의한 과실의 비파괴 당도 측정. 서울대학교 석사학위논문.
9. 최규홍. 1994. 비파괴검사기술의 농업적 이용현황과 전망. 주변첨단기술의 농업기계 분야 이용현황과 전망 심포지엄. 농업기계화연구소. 31-52.