Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 12
December 2023

Usability of NIR Hyperspectral Imaging for Evaluating Added Sugar Solution in Pineapple Juice

Achiraya Tantinantrakun [1], Anthony Keith Thompson [2], Sontisuk Teerachaichayut [3]

[1] Department of Food Science, School of Food-Industry, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Road, Ladkrabang, Bangkok 10520, Thailand.

[2] Department of Postharvest Technology, Cranfield University, College Road, Cranfield, Bedford, MK430AL, United Kingdom

[3] Department of Food Process Engineering, School of Food-Industry, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Road, Ladkrabang, Bangkok 10520, Thailand.

Abstract

Economically motivated adulteration of pineapple juice with sugar can be successfully detected by chemical analysis, but this is expensive and time consuming. Consequently, this research focused on testing whether the application of near-infrared (NIR) hyperspectral imaging could be successfully used to detect added sugar in pineapple juice. A dataset comprising 149 samples of pure concentrated pineapple juice and 149 samples of concentrated pineapple juice with added sugar solution in various concentrations, ranging from 0.5% to 99.5% w/w, was carried out. The data set of samples was divided into a calibration set and a prediction set for both quantitative and qualitative analyses. Partial least squares regression (PLSR) was employed for establishing the model for detecting the ratio of added sugar in the concentrated pineapple juice, achieving a correlation coefficient (R) of 0.985 with a root mean square error of prediction (RMSEP) of 4.971% w/w. Predictive images using colors was used to present added sugar in concentrated pineapple juice. The principal component analysis (PCA) showed the spectral information was able to be used to classify the group of pure concentrated pineapple juice and adulterated concentrated pineapple juice with the cumulative variance percentage of 98% for the first principal component (PC1) and the second principal component (PC2). Partial least squares-discriminant analysis (PLS-DA) was also employed for classifying the group of pure concentrated pineapple juice and adulterated concentrated pineapple juice, achieving an impressive total accuracy of 98.98% for the prediction set. It was concluded that this study demonstrated that NIR hyperspectral imaging could be used to detect added sugar adulteration of concentrated pineapple juice.

Keywords: quantitative, qualitative, spectra, calibration, prediction.

1. Introduction

Pineapple juice

Pineapple juice is consumed worldwide for its delicious flavour, but also for its health benefits, which include vitamin C, vitamins A, calcium and fibre . Economically motivated adulteration of fruit juices, including pineapple juice, with sugar can occur in many countries [6, 1]. Elliott [4] commented "I reckon there will be plenty of opportunities for diluting pineapple juice with sugary water or some other type of fruit juice fraud". Since consumers cannot detect adulterated pineapple juice visually, various analytical methods have been tested [6], but these methods are complicated. Examples of methods that have been reported for the detection of exogenous sugars in pineapple juice include liquid chromatography [6] and compound-specific stable hydrogen isotope analysis [1]. These methods are costly and time consuming. Hence, it is crucial to employ an effective assessment method that is fast, accurate and reliable to detect added sugar in pineapple juice, therefore near infrared hyperspectral imaging (NIR-HSI) was tested.

Near infrared hyperspectral imaging

NIR-HSI has been shown to be effective for assessing quality of samples by using spectral information in the wavelength between 750 nm and 2500 nm. Absorption vibration energy of molecular bonds such as C-H, O-H and N-H of samples, has been shown to provide spectra that can be related to the composition and properties of the sample . This is achieved by the spectra of every pixel within the region of interest (ROI) of each sample being used for developing the model to predict the qualities of samples both quantitatively and qualitatively. Predictive images can be created by the model to acquire predicted values of each pixel of samples, presenting by different colors that are related to variations of concentration. This approach helps visualize the distribution of qualities [1]. NIR-HSI has thus been employed to assess the qualities of different foods including: defects in green coffee , adulteration in tapioca starch , shelf life of cakes, moisture content in beef jerky , classify *Cinnamomum verum* and *C.cassia* sticks and predicting maturity index pineapple . Therefore, the objective of this research was to test whether NIR-HSI could be successfully used to detect added sugar solution in pineapple juice.

2. Methodology

A. Sample preparation

A 60°Brix sugar solution was prepared, comprising fructose, glucose and sucrose in a ratio of 12:13:28, which was reported to be similar to the ratio in the natural pineapple juice [4]. Samples of natural 60°Brix concentrated pineapple juice from different sources of production were purchased from a pineapple processing factory located in Prachuap Khiri Khan in Thailand. Subsequently, samples of adulterated concentrated pineapple juice were prepared by adding a 60°Brix sugar solution to these samples, at levels ranging from 0.5% to 99.5% w/w to the natural 60°Brix concentrated pineapple juice. These samples were arranged into 2 groups, that was one group of pure concentrated pineapple juice and another group of adulterated concentrated pineapple juice.

B. Spectra acquisition

A NIR hyperspectral imaging system in wavelength range from 935 nm to 1720 nm (Specim, Spectral Imaging Ltd, Oulu, Finland) including the SPECIM FX camera, lamps, a tray conveyor and computer unit were employed for spectra acquisition (Fig 1). Each sample was placed in a transflectance container and scanned at a speed of 15 mm/s using the push broom line scanning technique, which moved and scanned the sample in every spatial line until all parts were covered. A dark reference was acquired when the scanner shutter was closed and the camera lens covered, and a white reference was acquired by scanning a Spectralon tile.

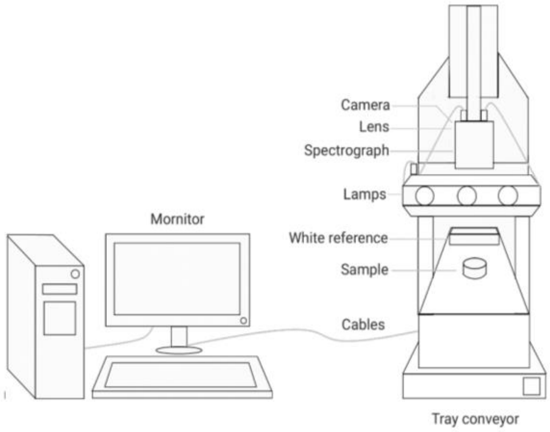


Fig 1. Schematic diagram of the NIR hyperspectral imaging system

C. Data analysis

For quantitative analysis, samples of each of the adulterated concentrated pineapple juice were tested in order to establish the calibration model using partial least squares regression (PLSR). PLSR is a powerful technique for handling data matrices with a large number of variables by extracting the essential information from the data matrices and generating more reliable models [1]. The spectral data were considered the independent variables while the percentage of added sugar was considered the dependent variables. The data were divided into two sets: a calibration set that was used for developing and cross validating the calibration model and a prediction set that was used for testing the calibration model. Spectral data were preprocessed using the following spectral pretreatment methods: Savitzky–Golay smoothing, first derivative, second derivative, multiplicative scatter correction (MSC) and standard normal variate (SNV) normalization, to obtain the optimal condition for developing the optimum model. The correlation coefficient and the root mean square error were used to evaluate the performance of the model.

For qualitative analysis, samples of undiluted concentrated pineapple juice, as it was obtained from the processing factory, were defined as 0 and samples of each of the adulterated concentrated pineapple juice were defined as 1. For classification analysis, the ability of classification using the spectral data from NIR-HSI was evaluated using principal component analysis (PCA). Also, the accuracy of classification was evaluated using partial least squares discriminant analysis (PLS-DA), which is a technique to reduce complexity by extracting latent variables from highdimensional spectral data for predicting sample variables. The percentage of correct predictions was used to evaluate the accuracy of discrimination. The Umbio Evince Hyper-Spectral Imaging Software (Prediktera Evince, version 2.7.5, Sweden), and the Unscramble X version 10.5.1 (CAMO, Osla, Norway) were used for analysis in this study.

3. Result and Discussion

A. Spectrum attributes

The spectral image from NIR-HSI of each sample in the transflectance container was investigated and the spectra data from every pixel of each sample were averaged and then used as its representative for analysis. The spectral data of pure concentrated pineapple juice samples within the range of 935 to 1720 nm (Fig 2a), the averaged spectrum (Fig 2b) and the spectra data of adulterated concentrated pineapple juice samples within the range of 935 to 1720 nm (Fig 2c) and the averaged spectrum (Fig 2d) all showed that the spectra of pure concentrated pineapple juice and spectra of adulterated concentrated pineapple juice were very similar. Prominent peaks appeared at around 1200 nm and 1410 nm, which are associated with the second overtone of O–H bonds, occurring at wavelengths of 1200 nm and 1410 nm, as well as the absorption at 1450 nm is associated the first overtone of O–H stretch in water and sugar components.

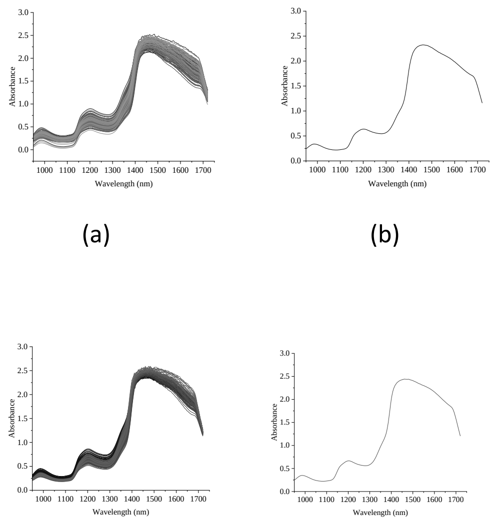


Fig 2. Spectra of pure concentrated pineapple juice from NIR-HSI (a), averaged spectrum of pure concentrated pineapple juice (b), all spectral of the adulterated concentrated pineapple juice (c) and averaged spectrum of adulterated concentrated pineapple juice (d)

Spectral images of the sugars in solution (fructose, glucose and sucrose) (Fig 3) show that each sugar exhibited a spectral feature with its peaks, as reported by Workman and Weyer [16], as 1410 nm for sucrose, 1450 nm for fructose and 1560 nm for glucose.

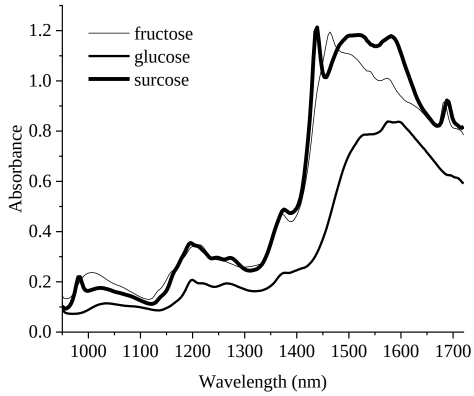


Fig 3. Spectra of fructose, glucose and sucrose.

B. Data analysis

Quantitative analysis

Partial least squares regression

149 samples of adulterated concentrated pineapple juice were carried out for quantitative analysis, with the calibration set containing 100 samples and the prediction set containing 49 samples. The characteristics of the percentage of added sugar solution in concentrated pineapple juice of samples (Table I) show that the range of the values in the prediction set was covered by those of values in the calibration set. Also, their standard deviations of both sets were similar.

Table I. Added sugar solution in concentrated pineapple juice in the calibration set and the prediction set.

Set	Number of samples	Range (% w/w)	Mean (% w/w)	SD (% w/w)
Calibration	100	0.50-99.50	50.46	29.06
Prediction	49	2.00-98.50	50.48	28.61

SD = standard deviation

The results of spectral pretreatment methods for developing the models using PLSR in the calibration set (Table II) show that the smoothing combined with SNV spectral pretreatment method gave the lowest RMSECV (5.014% w/w) and the highest Rcv (0.985). Therefore, the smoothing combined with SNV spectral pretreatment method was selected for developing the calibration model in this study.

Table II. PLSR results by spectral pretreatment methods in the calibration set.

PLSR	a	b	c	d	e	f	g
LV	4	6	4	4	5	5	5
R_{cv}	0.974	0.983	0.975	0.916	0.980	0.981	0.985
RMSECV (%w/w)	6.558	5.251	6.440	11.596	5.760	5.549	5.014

PLSR = partial least squares regression

LV = latent variable

RMSECV = root mean square error of cross validation

R = correlation coefficient

a = original spectrum

b = Savitzky–Golay Smoothing

c = 1st Derivative

d = 2nd Derivative

e = multiplicative scatter correction

f = standard normal variate

g = Smoothing and SNV

The calibration model that was developed, using smoothing combined with SNV pretreatment method was performed on both the calibration set and the prediction set (Table III) and gave the good accuracy of PLSR model for both the calibration and prediction sets, with an R value of 0.988 and 0.985, and an RMSE of 4.389% w/w and 4.971% w/w respectively.

Table III. Performance of the PLSR model from samples in the calibration set and the prediction set.

Set	Pre treatment	LV	N	R	RMSE (%w/w)
C	Smoothing and SNV	5	100	0.988	4.389
P	Smoothing and SNV	5	49	0.985	4.971

N = number of samples

LV = latent variable

C = calibration set

P = prediction set

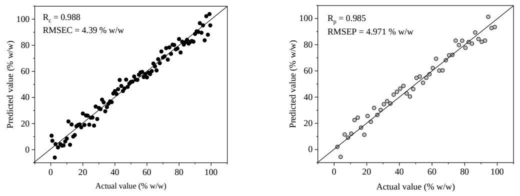
R = correlation coefficient

RMSE = root mean square error

Smoothing = Savitzky–Golay smoothing

SNV = standard normal variate

The scatter plots obtained by the PLSR model for the calibration set (Fig 4a) and prediction set (Fig 4b) indicate that the PLSR model was accurate since the plots of predicted values versus actual values are closely to 45-degree line, as previously reported by Williams [17]. These results indicate a good potential of NIR-HSI for use in predicting percentage of added sugar solution in concentrated pineapple juice.



(a) (b)

Fig 4. The scatter plot of predicted values versus actual values using the PLSR model for the calibration set (a) and the prediction set (b).

Visualization of adulteration

Spectral images, acquired from NIR-HSI, were used to create predictive images using the calibration model, by predicting the percentage of added sugar solution in concentrated pineapple juice in every pixel. Interpretation from the percentage of added sugar solution to colors was done by image processing. The linear color scale showed the colors varied, and this variation was related to the percentage of added sugar solution where 0% (blue) up to 100% (red). The visualization of adulterated concentrated pineapple juice based on the percentage of added sugar solution (Fig 5) shows that adulterated concentrated pineapple juice could be detected by predictive images. It could be applied for grading systems by visualization.

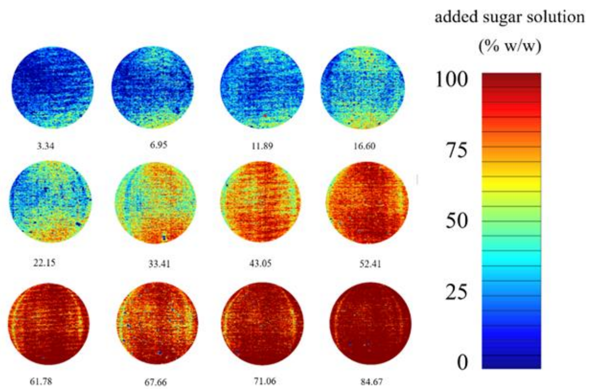


Fig 5. Predictive images of adulterated concentrated pineapple juice.

Qualitative analysis

Principal component analysis

The score plot of PC1 and PC2 using spectral data of pure concentrated pineapple juice and the adulterated concentrated pineapple juice shows clear separation of two clusters with the cumulative variance percentage of 98% (Figure 6). The variation for PC1 and PC2 were 74% and 24%, respectively, indicating that spectral information of samples could be used to separate pure concentrated pineapple juice and the adulterated concentrated pineapple juice.

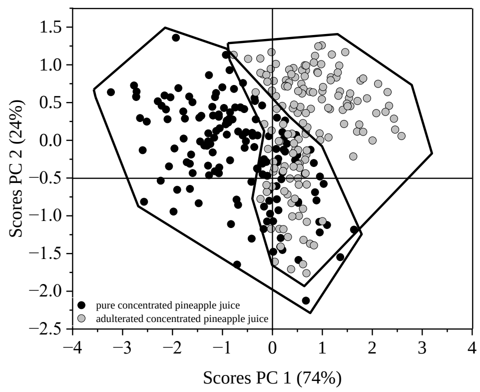


Fig 6. PCA of the spectral information of pure concentrated pineapple juice and the adulterated concentrated pineapple juice.

Partial least squares-discriminant analysis

149 samples of pure concentrated pineapple juice and 149 samples of adulterated concentrated pineapple juice were used for classification analysis, giving a total of 298 samples. These were divided into the calibration set (n=200) and the prediction set (n=98). The same number of pure concentrated pineapple juice samples (n=100) and adulterated concentrated pineapple juice samples (n=100) was arranged in the calibration set. Additionally, the same number of pure concentrated pineapple juice samples (n=49) and adulterated concentrated pineapple juice samples (n=49) were arranged in the prediction set.

Spectral data of samples in the calibration set were preprocessed using the spectral pretreatment methods described above, giving the PLS-DA of the various spectral pretreatment methods. The smoothing spectral pretreatment method gave the most accurate results with a total accuracy of 92.00% (Table IV). Therefore, the smoothing spectral pretreatment method was selected for PLS-DA in this study.

Table IV. Partial least squares-discriminant analysis (PLS-DA) results by spectral pretreatment methods in the calibration set.

Pretreatment methods	LV	Pure concentrated pineapple juice (0)(100 samples)		Adulterated concentrated pineapple juice (1)(100 samples)		Total accuracy (200 samples)		Total accuracy (%)
		Correct	Incorrect	Correct	Incorrect	Correct	Incorrect	
Original	5	88/100	12/100	90/100	10/100	178/200	22/200	89.00
Smoothing	11	93/100	7/100	91/100	9/100	184/200	16/200	92.00
1st Derivative	5	92/100	8/100	89/100	11/100	181/200	19/200	90.50
2nd Derivative	2	78/100	22/100	80/100	20/100	158/200	42/200	79.00
MSC	5	90/100	10/100	91/100	9/100	181/200	19/200	90.50
SNV	5	88/100	12/100	90/100	10/100	178/200	22/200	89.00

LV = Latent variable

Smoothing = Savitzky–Golay smoothing

1st Derivative = first derivative

2nd Derivative = second derivative

MSC = multiplicative scatter correction

SNV = standard normal variate

PLS-DA was performed using the classification model that had been developed using the smoothing pretreatment method. PLS-DA results of the calibration set, and the prediction set (Table V) show good results for classification for both the calibration and prediction sets with the total accuracy of 96.0% and 98.98% respectively.

Table V. Partial least squares-discriminant analysis (PLS-DA) results by samples in the calibration set and the prediction set.

Set	LV	Pretreatment method	Groups of samples			Total accuracy (%)
C 200 samples	11	Smoothing	0	correct	96/100	96.00
				incorrect	4/100	
			1	correct	96/100	
				incorrect	4/100	
P 98 samples	11	Smoothing	0	correct	48/49	98.98
				incorrect	1/49	
			1	correct	49/49	

	incorrect	0/49

C = calibration set

P = prediction set

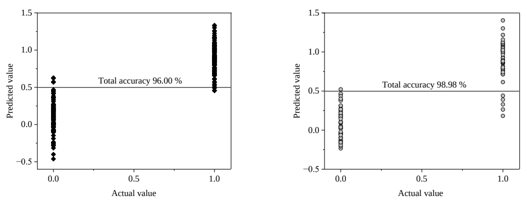
LV = latent variable

0 = pure concentrated pineapple juice

1 = adulterated concentrated pineapple juice

A scatter plot of actual and predicted values of the pure concentrated pineapple juice samples (0) and the adulterated concentrated pineapple juice samples (1) (Fig 7) Show the cut off value of 0.5, which was used to evaluate the accuracy of classification. Using PLS-DA, the total accuracy of classification by cross validation in the calibration set was 96.0% while the total accuracy of classification by testing the samples in the prediction set was 98.98%. These results

showed the potential of NIR-HSI for use in differentiating between groups of pure concentrated pineapple juice and adulterated concentrated pineapple juice.



(a) (b)

Fig 7. The scatter plot of predicted values versus actual values by PLS-DA for the calibration set (A) and the prediction set (B).

4. Conclusion

The tests showed that near-infrared hyperspectral imaging, in the range of 935-1720 nm, when smoothing combined with standard normal variate normalization pretreatment using partial least squares regression could be successfully used to predict added sugar adulterating pineapple juice. This method gave a predictive correlation coefficient of 0.985 and a root mean square error of prediction of 4.971% w/w. Smoothing pretreatment was used for establishing the classification model for comparing groups of the pure concentrated pineapple juice with concentrated pineapple juice adulterated with added sugar, using partial least squares-discriminant analysis, achieving an overall accuracy of 98.98%. It was concluded that the results showed that NIR-HSI could be applied for use in detecting sugar adulteration in concentrated pineapple juice.

Acknowledgement

This work is supported by King Mongkut's Institute of Technology Ladkrabang. The authors are grateful to Prof. Dr. Panmanas Sirisomboon for providing technical assistance.

References

[1] Shahidi, F. and C. Alasalvar, Handbook of functional beverages and human health. 1st ed. ed. 2016.

[2] Dasenaki, M.E. and N.S. Thomaidis, "Quality and authenticity control of fruit juices-a review". Molecules, 2019. 24(6).10.3390/molecules24061014

[3] Kelly, S.D., et al., "Detection of exogenous sugars in pineapple juice using compound-specific stable hydrogen isotope analysis". NPJ Sci Food, 2021. 5(1): p. 10.10.1038/s41538-021-00092-5

[4] Elliott, C., Pineapples set off food fraud alarm bells, in Food Safety. 2016.

[5] Pudełko, A., et al., "Application of FT-NIR spectroscopy and NIR hyperspectral imaging to predict nitrogen and organic carbon contents in mine soils". Measurement, 2020. 164: p. 108117.https://doi.org/10.1016/j.measurement.2020.108117

[6] Kucha, C.T., et al., "Assessment of Intramuscular Fat Quality in Pork Using Hyperspectral Imaging". Food Engineering Reviews, 2021. 13(1): p. 274-289.10.1007/s12393-020-09246-9

[7] Oliveri, P., et al., "An innovative multivariate strategy for HSI-NIR images to automatically detect defects in green coffee". Talanta, 2019. 199: p. 270-276.https://doi.org/10.1016/j.talanta.2019.02.049

[8] Khamsopha, D., S. Woranitta, and S. Teerachaichayut, "Utilizing near infrared hyperspectral imaging for quantitatively

predicting adulteration in tapioca starch". Food Control, 2021. 123: p. 107781.https://doi.org/10.1016/j.foodcont.2020.107781

[9] Sricharoonratana, M., A.K. Thompson, and S. Teerachaichayut, "Use of near infrared hyperspectral imaging as a nondestructive method of determining and classifying shelf life of cakes". LWT, 2021. 136: p. 110369.https://doi.org/10.1016/j.lwt.2020.110369

[10] Achata, E.M., et al., "Development of NIR-HSI and chemometrics process analytical technology for drying of beef jerky". Innovative Food Science & Emerging Technologies, 2021. 69: p. 102611.https://doi.org/10.1016/j.ifset.2021.102611

[11] Cruz-Tirado, J.P., et al., "Rapid and non-destructive cinnamon authentication by NIR-hyperspectral imaging and classification chemometrics tools". Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2023. 289: p. 122226.https://doi.org/10.1016/j.saa.2022.122226

[12] Tantinantrakun, A., et al., "Nondestructive evaluation of SW-NIRS and NIR-HSI for predicting the maturity index of intact pineapples". Postharvest Biology and Technology, 2023. 195: p. 112141.https://doi.org/10.1016/j.postharvbio.2022.112141

[13] Elkins, E.R., et al., "Characterization of Commercially Produced Pineapple Juice Concentrate". Journal of Food Composition and Analysis, 1997. 10(4): p. 285-298.https://doi.org/10.1006/jfca.1997.0547

[14] Asante, E.A., et al., "Detection and assessment of nitrogen effect on cold tolerance for tea by hyperspectral reflectance with PLSR, PCR, and LM models". Information Processing in Agriculture, 2021. 8(1): p. 96-104.https://doi.org/10.1016/j.inpa.2020.03.001

[15] Quan, N.M., et al., "Species and geographical origin authenticity of green coffee beans using UV–VIS spectroscopy and PLS–DA prediction model". Food Chemistry Advances, 2023. 2: p. 100281.https://doi.org/10.1016/j.focha.2023.100281

[16] Workman, J. and L. Weyer, Practical guide and spectral atlas for interpretive near-infrared. 2012: CRC.

[17] Williams, P.C., Implementation of Near-Infrared Technology. 2nd Edition ed. 2001, American Association of Cereal Chemists. St. Paul: Near-Infrared Technology in the Agricultural and Food Industries.