ORIGINAL PAPER

Fourier Transform Near Infrared Spectroscopy as a Quality Control Tool for the Analysis of Lecithin and By-Products During Soybean Oil Processing

Hui Li · Mark Goulden · Robert Cocciardi · Jan Hughes

Abstract Fourier transform near infrared (FT-NIR) spectroscopy was used to analyze multiple measurement parameters in lecithin production samples and soybean oil refining by-products. For lecithin, partial least squares (PLS) calibration models were developed for acetone insolubles, acid value and moisture and leave-one-out cross validation of the calibration models yielded root mean square error of cross validation (RMSECV) values of 0.37%, 0.59 (mg KOH/g) and 0.050%, respectively. An independent test set consisting of 40% of the lecithin production samples were predicted from the PLS calibration models and a root mean square error of prediction (RMSEP) of 0.41%, 0.53 (mg KOH/g) and 0.056% were obtained for acetone insolubles, acid value and moisture, respectively. Comparison of FT-NIR predictions and corresponding reference method values of 10 lecithin samples using a two-tailed t test showed no significant difference at the $p = 0.05$ level. A set of 51 samples of soybean oil refining by-products, including acidulated soapstock, fatty acids and black oil, were used for developing PLS calibration models for measuring acid value, moisture and iodine value and leave-one-out cross validations for each model gave values for RMSECV of 6.59 (mg KOH/g), 0.046% and 0.42 (mg I_2/g), respectively. Overall, the results of this study demonstrate the suitability of FT-NIR spectroscopy for the routine analysis of lecithin production samples and soybean oil refining by-products for quality control purposes.

Introduction

Soybean oil is one of the least expensive vegetable oils and is commonly used for salad dressing and cooking oil or hydrogenated for use as margarine stock or frying oil [1]. Crude soybean oil needs to be processed to remove phosphatides, fatty acids and oxidation products, which negatively affect oil quality. Key steps in soybean oil processing include degumming, refining (alkali neutralization), bleaching and deodorization. During soybean oil processing, phospholipids are recovered from the degumming process to produce lecithin, an emulsifier commonly used as a food ingredient, while fatty acids are recovered by acidulating the soap stock produced during the refining process and are normally sold as fertilizer and feed ingredients [2].

A complete quality assessment of lecithin and free fatty acids during the production process is important because of their broad range of applications in various commercial products. At different steps during the production of lecithin and free fatty acids, various measurement parameters need to be evaluated for quality assurance. For lecithin production, percent acetone insolubles, moisture and acid value are common quality assessment parameters; acetone insolubles in particular are an important parameter for determining the amount of phosphatides in soy lecithin, while high levels of moisture may cause bacterial growth

H. Li (✉) · R. Cocciardi
Bruker Optics Inc., 19 Fortune Drive,
Billerica, MA 01821, USA
e-mail: hli@brukeroptics.com; hui.li@brukeroptics.com

M. Goulden · J. Hughes
CHS Inc., 2060 South Riverfront Drive,
Mankato, MN 56002, USA

and the acid value of lecithin is important for determining its blending properties in different types of food products. The acidulated soap stock obtained by soybean oil refining can be distilled to produce pure free fatty acids, leaving a black oil residue. During this process, the acid value is an important parameter for determining the yield of free fatty acids from the acidulated soap stock, which are often added back to lecithin to adjust its acid value, and the iodine value is important for measuring the degree of unsaturation of the fatty acids added back into lecithin.

Traditionally, these parameters are measured using time-consuming and tedious wet chemical analyses such as gravimetric, titrimetric and extraction methods that can take up to several hours per analysis. Manufacturers of soy lecithin and free fatty acids would significantly benefit from using a more rapid analytical technique that would allow at-line data to be obtained during the manufacturing process. Near-infrared (NIR) spectroscopy is a commonly used analytical technique for the quantitative analysis of liquids and solids, and several papers have reported on its use for the quality assessment of various food and agricultural products, such as seeds [3, 4], grains [4, 5], meat [6], dairy products [7] and edible oils [8–10]. Amongst NIR spectrometers, Fourier transform near-infrared (FT-NIR) spectrometers equipped with a cube-corner interferometer provide the highest performance with the advantages of acquiring NIR spectral data with high wavenumber accuracy and precision as well as being ideal for a production environment due to the stability of the cube-corner interferometer design [11, 12]. FT-NIR spectroscopy is, therefore, a rapid and non-destructive analysis technique that gives reproducible, accurate and precise results and, as such, could be considered as a suitable alternative for the analysis of soybean oil processing by-products.

Some recent work has been reported on using a dispersive NIR instrument for measuring acetone insolubles, moisture, acid value and phosphatidylcholine in crude, intermediate, final and de-oiled lecithin samples [13, 14]. However, the predictive accuracy of NIR spectroscopy in the analysis of lecithin during soybean oil processing has not been extensively investigated and no work has been reported on the quality control of other common by-products of soybean oil processing, such as acidulated soap stock and fatty acids. The objectives of this study were (1) to develop calibration models based on partial least squares (PLS) regression for quantitatively measuring acetone insolubles, acid value, and moisture in lecithin production samples using an FT-NIR spectrometer and (2) to investigate the feasibility of using an FT-NIR spectrometer for measuring acid value, iodine value and moisture in samples of acidulated soap stock, black oil and free fatty acids obtained as by-products from the soybean oil refining process.

Experimental Procedures

Samples and Reference Values

Lecithin and refining by-product samples were obtained from CHS Inc. (Mankato, MN). The sample set included 238 samples of lecithin, 18 acidulated soap stock samples, 4 fatty acid samples, 6 samples of black oil and 23 different blends of black oil and acidulated soap stock obtained from the manufacturing plant over a period of 14 months. On the basis of the availability of reference values; 238, 179 and 172 of the lecithin production samples were used for developing calibrations and validations for acid value, moisture and acetone insolubles, respectively. The by-product samples were grouped together ($n = 51$) into a second set for measuring acid value, iodine value and moisture in soybean oil refining by-products. Reference values for the calibration and validation samples were obtained for each measurement parameter using the following AOCS official methods: acetone insolubles (lecithin); Ja 4-46, acid value (lecithin); Ja 6-55, moisture (lecithin); Ja 2b-87, iodine value (by-products); Cd 1c-85, acid value (by-products); Cd 3d-63, moisture (by-products); Ca 2e-84. Each sample was measured in duplicate and the average value was used for NIR calibration and validation. An additional ten lecithin samples were measured in triplicate using both the FT-NIR spectrometer and reference methods for doing a statistical comparison of both techniques to determine if the values were significantly different. All samples were stored at room temperature in sealed containers prior to and after measurement.

FT-NIR Instrumentation and Measurement of Samples

All measurements were carried out using a Bruker Optics (Billerica, MA) Multi-Purpose Analyzer (MPA) FT-NIR spectrometer equipped with a cube-corner interferometer, tungsten–halogen NIR source and a thermally electrically cooled InGaAs detector. For FT-NIR sample measurement, samples were transferred into 8-mm glass vials (Klaus Ziemer, Langerwehe, Germany) and measured by placing them in a sample holder inserted into a temperature controlled (60 ± 1 °C) sample compartment. Samples were measured by scanning across the NIR spectral range (12,500–4,000 cm^{-1}) and by co-adding 16 scans at a resolution of 8 cm^{-1} using Bruker Optics' OPtical User Software (OPUS) package. The single channel spectra of the samples were ratioed against a background single channel spectrum collected prior to running samples, without any sample vial in the holder, and the absorbance spectrum of each sample was calculated and stored to disk. Each sample was measured in triplicate and the average spectrum was computed for further processing.

The operation, reproducibility and accuracy of the FT-NIR instrument was automatically checked on a daily basis using performance qualification (PQ) tests. The following tests were carried out on a regular basis: photometric accuracy, wavenumber accuracy, signal to noise, deviation from 100% line, interferogram amplitude and energy distribution. If the instrument passed the acceptance criteria for each test, then the instrument was considered "qualified" for running samples on that day.

FT-NIR Calibration and Validation

All calibrations were developed by employing partial least squares (PLS) regression using the Bruker Optics' OPUS QUANT 2 software package. The optimal calibration model for each measurement parameter was selected based on the fit between the reference values and the predictions obtained by leave-one-out cross validation. Models that gave the lowest possible rank with the highest coefficient of determination (R^2), lowest root mean square error of cross validation (RMSECV) and highest value of residual prediction deviation (RPD) were chosen. Outliers were identified by examining scores plots and Mahalanobis distance versus residual plots as well as by calculating the F-value and F-probability for spectral residuals and concentration residuals. Sample spectra with an F-probability value >0.99 were determined to be outliers and removed from the data set. The lecithin calibrations for acetone insolubles, acid value and moisture were developed using 60% of the production samples and validated using the remaining 40% of the samples as an independent test set. For each measurement parameter NIR predicted values were compared to wet chemistry values and the coefficient of determination, root mean square error of prediction (RMSEP), bias and slope between actual and predicted values were calculated to assess the calibrations models' predictive capability.

Results and Discussion

FT-NIR Spectra of Lecithin and Refining By-Products

The first derivatives of the average spectra of the samples of lecithin, acidulated soapstock, black oil and fatty acids are displayed in Figs. 1 and 2 for the combination region (4,700–5,500 cm^{-1}) and first overtone region (6,000–7,800 cm^{-1}), respectively. In Fig. 1, there are differences in the intensity of the band at 5,300 cm^{-1} corresponding to the O–H combination band of water, illustrating the difference in moisture content between the different types of samples, with acidulated soap stock having the highest moisture content on average followed by lecithin, black oil
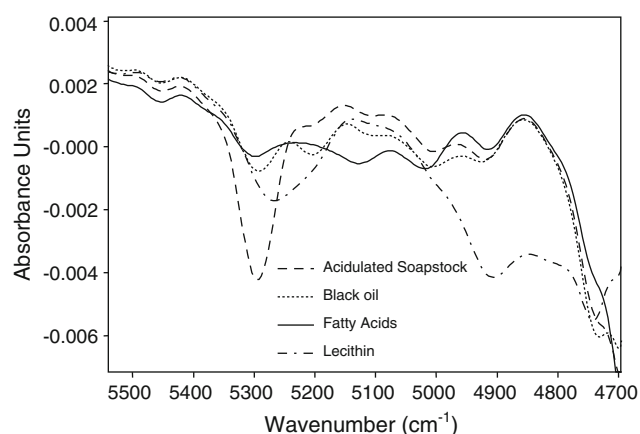

Fig. 1 First derivative of the average FT-NIR spectrum of acidulated soapstock, black oil, fatty acids and lecithin samples for the combination region over the spectral range between 5,500 and 4,700 cm^{-1}
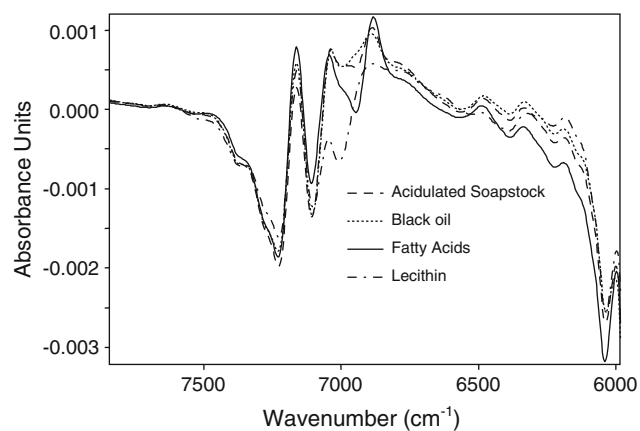

Fig. 2 First derivative of the average FT-NIR spectrum of acidulated soapstock, black oil, fatty acids and lecithin samples for the first overtone region over the spectral range between 8,000 and 6,000 cm^{-1}

and fatty acids. A band at approximately 4,900 cm^{-1} present in the average spectrum of lecithin but absent from the average spectra of the other samples is due to the absorption of various phosphatidyl compounds, such as phosphatidylcholine, phosphatidylethanolamine and phosphatidylinositol and hence is related to the acetone insolubles parameter. In Fig. 2, the band at 6,900 cm^{-1} in the first overtone region is due to the O–H first overtone of carboxylic acids, which is related to the acid value of the sample. This band is most intense in the average spectrum of fatty acids, followed by acidulated soapstock and black oil, which contains little or no fatty acids, while in the spectrum of lecithin this band is overlapped by another band at 7,000 cm^{-1} coming from the absorption of phosphatidyl compounds. In general, because the NIR region corresponds to the absorption of overtones and combinations of the fundamental frequency molecular vibrations,
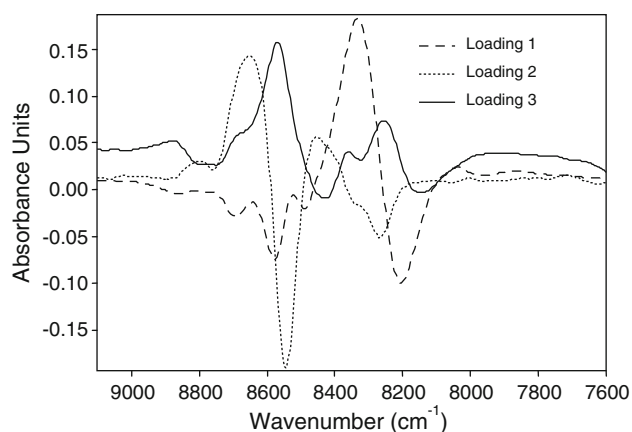
Fig. 3 Partial least squares regression loadings computed for iodine value for the soybean oil refining by-products spectral data set over the spectral range between 9,100 and 7,600 cm^{-1}

Fig. 4 Scores plot of principal component 1 and principal component 2 of 172 spectra of lecithin samples

NIR spectral features are broad and highly overlap. Therefore, chemometric methods such as principal component analysis (PCA) and PLS need to be employed to detect spectral differences by computing latent variables, also known as loading spectra, that are related to the component of interest. For example, due to significant band overlap, differences in iodine value are not evident by viewing the raw spectral data alone. Figure 3 illustrates the first three loading spectra obtained, when PLS was employed for measuring iodine value in the 51 samples of soybean oil refinery by-products in the region from 9,000 to 7,600 cm^{-1}. This corresponds to the second overtone of the C–H stretching band, the band most commonly used for NIR iodine value determinations [15]. The loading spectra indicate that there are spectral features in that region that can be correlated to differences in iodine value and that can be used in developing a PLS calibration model.

Calibration Model Development and Outlier Detection

All calibrations for the various parameters of the different types of samples were developed using PLS regression. In order to develop robust and suitable PLS calibration models, outliers needed to be detected and removed from the calibration data set, as some samples may be contaminated, have a different composition or were not measured properly. Spectral outliers can be detected by using PCA and by comparing the Mahalanobis distance of each sample spectrum in the data set relative to the spectra of the other calibration samples. In Fig. 4, a plot of the scores values of principal components 1 and 2 for 172 sample spectra used for the analysis of acetone insolubles in lecithin indicates that three of the spectra were different from the rest of the data set and were thus flagged as spectral outliers. PLS regression of the remaining calibration spectra was then
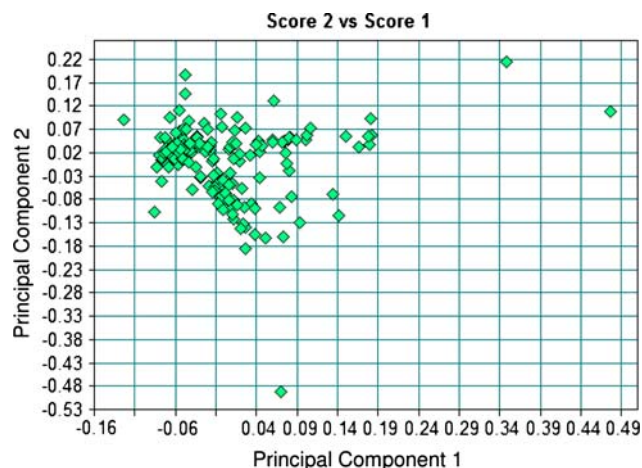
carried out using different spectral regions and preprocessing techniques, and each model was evaluated by leave-one-out cross validation. Further outliers were determined by calculating the F-value and F-probability of the spectral and concentration residuals for each sample in the data set after carrying out a leave-one-out cross validation. A high F-value and an F-probability above 0.99 means that the spectrum of that sample does not fit with the rest of the data set and should be removed from the calibration. Overall, for acetone insolubles in lecithin, among 172 sample spectra collected, four sample spectra were found to have a high F-value and F-probability >0.99 for spectral residuals and 23 sample spectra were found to have a high F-value and F-probability >0.99 for concentration residuals. This process of removing outliers to obtain robust leave-one-out cross validation results was carried out in developing PLS calibration models for all measurement parameters.

After all outliers were removed, 60% of the samples were randomly chosen for developing the final PLS calibration models for each measurement parameter. For each PLS calibration, the regions and preprocessing method selected that gave the lowest prediction error with the lowest rank was chosen as the most robust model. Tables 1 and 2 list the measurement regions, data preprocessing method, rank, coefficient of determination (R^2), root mean square error of cross validation (RMSECV), residual prediction deviation (RPD) and range of concentration values for the PLS calibration models developed for each measurement parameter in lecithin and soybean oil refinery by-products, respectively. The regions ultimately selected for the calibration models included the corresponding components' observed NIR spectral band intensity changes in Figs. 1 and 2. For example, a band at approximately 6,900 cm^{-1} was shown to be related to acid value and the

region used for the acid value calibration ranged from 8,871 to 6,137 cm^{-1}. For calibrations for moisture, the first overtone of the OH stretching band at approximately 7,100 cm^{-1} proved to give the best results while using the OH combination band did not generate quite as good results perhaps due to greater variability of the baseline in that region. For lecithin, the best results were obtained for acid value, with an R^2 of 0.968 and an RPD value of 5.61, followed by acetone insolubles with an R^2 of 0.948 and an RPD value of 4.38. Generally, the RPD, the ratio of the standard deviation (SD$_{ref}$) of the reference values in the data set relative to the standard error of prediction (SEP), is a better indicator of the performance of a calibration model than the R^2 value or RMSECV. The R^2 value may be misleading because high R^2 values can be obtained for calibrations with a large dynamic concentration range while still having a high prediction error and low R^2 values can be obtained for calibrations with narrower concentration ranges while still giving good quality predictions. The RMSECV gives a good indication of the prediction error but cannot be used as an overall measure of the quality of a calibration model because its value depends on the units of measurement. In contrast, because the RPD is a ratio of standard deviations, it is independent of the units of measurement, and in general calibrations with an RPD >5 are considered good for quality control purposes and those with an RPD >3 are good for screening purposes. The lower R^2 and the RPD value for moisture (0.855 and 2.63) in lecithin can be explained by the fact that the standard deviation of the reference values relative to the mean reference value is small compared to the standard deviation for accuracy of the reference method. In this case, the quality of the PLS model is limited by the accuracy of the standard reference method, which, by consequence gives higher SEP values and lower RPD and R^2 values.

Validation of Lecithin Calibrations

An independent test set validation was carried out for each calibration by using the remaining 40% of samples not used in the PLS calibration model. Table 3 lists the range of concentration values, value for R^2, RMSEP, bias and slope between actual and predicted values for acetone insolubles, acid value and moisture content in lecithin. The results in Table 3 indicate that there was a relatively good correlation between the reference values and FT-NIR predictions with no bias between the methods, values for the slope of the regression line >0.9 for all parameters, and RMSEP values comparable to their respective RMSECV values in Table 1. The plots of actual versus predicted values for the independent test set validation for acetone insolubles, acid value and moisture are illustrated in Figs. 5, 6 and 7, respectively. Overall the results showed relatively good predictions for the test set validation indicating that the calibration models are robust. The feasibility of using the PLS calibrations for quality control of lecithin was further evaluated by comparing FT-NIR prediction results to reference data on a total of ten lecithin samples measured in triplicate using both FT-NIR and corresponding reference methods. The results were compared using a paired two-tailed t test at the $p = 0.05$ significance level to determine whether they were significantly different. Values for t of 1.81, 1.95 and 0.45 were obtained for acetone insolubles, acid value and moisture, respectively, which is below the critical t value of 2.23 at the $p = 0.05$ significance level. Based on these results, it was determined that the FT-NIR predictions are comparable to their corresponding reference methods and therefore the FT-NIR based method for lecithin analysis is suitable for implementation in routine quality control.

Feasibility of PLS Calibration Models for Soybean Oil Refining By-Products

The feasibility of developing PLS calibration models for predicting acid value, moisture and iodine value in soybean oil refining by-products was assessed using a set of 51 samples of acidulated soap stock, black oil, fatty acids and various blends of acidulated soap stock and black oil, which were blended directly in production tanks so as to approximate real production samples. Including the blend samples into the calibration set provided a more uniform distribution

Table 1 Spectral regions, data preprocessing methods, rank (# of factors), range, R^2, RMSECV and RPD for PLS regression models for acetone insolubles, acid value and moisture in lecithin

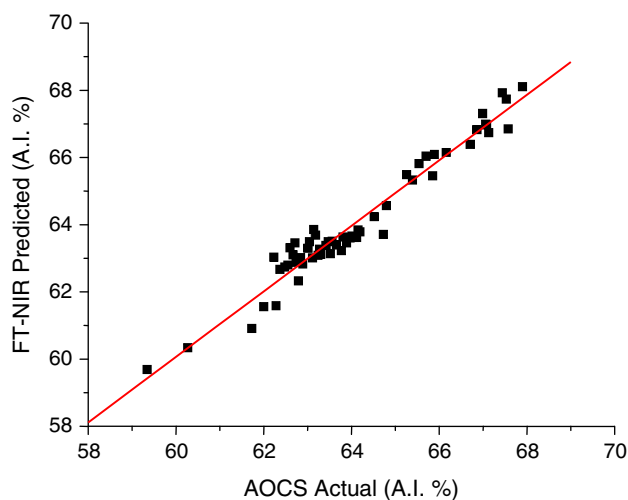| Measurement parameter	Spectral region(s) (cm^{-1})	Data preprocessing	Range	Rank	R^2	RMSECV	RPD
Acetone insol. (%)	8995–6098	1st derivative + VN	59.2–68.0	5	0.948	0.374	4.38
	5450–4494						
Acid value (mg KOH/g)	8871–6137	2nd derivative	18.1–29.8	6	0.968	0.588	5.61
Moisture (%)	9165–6098	1st derivative	0.17–0.78	2	0.855	0.0503	2.63

VN vector normalization

Table 2 Spectral regions, data preprocessing methods, rank (# of factors), range, R^2, RMSECV and RPD for PLS regression models for iodine value, acid value and moisture in soybean oil refining by-products

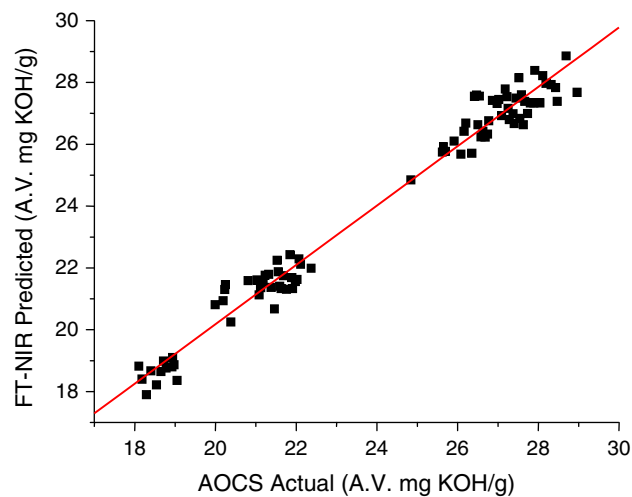| Measurement parameter	Spectral region(s) (cm^{-1})	Data preprocessing	Range	Rank	R^2	RMSECV	RPD
Iodine value (mg I$_2$/g)	9157–7498	1st derivative + VN	126.8–133.9	2	0.914	0.418	3.40
Acid value (mg KOH/g)	7502–6314 5446–4509	1st derivative + VN	21.7–210.9	2	0.979	6.59	6.95
Moisture (%)	7502–6098	None	0.01–1.64	5	0.988	0.0456	8.71

VN vector normalization

Table 3 Comparison of results between wet chemical analysis and FT-NIR predictions of an independent test set validation for the determination of acetone insolubles, acid value and moisture in lecithin

Measurement parameter	Range	R^2	RMSEP	Bias	Slope
Acetone insolubles (%)	59.3–67.9	0.953	0.405	0.0392	0.975
Acid value (mg KOH/g)	18.1–29.0	0.977	0.529	−0.0145	0.960
Moisture (%)	0.17–0.78	0.826	0.0561	−0.00898	0.925

Fig. 6 Plot of FT-NIR predicted values versus AOCS reference method values for acid value in lecithin for the independent test set validation

Fig. 5 Plot of FT-NIR predicted values versus AOCS reference method values for acetone insolubles in lecithin for the independent test set validation
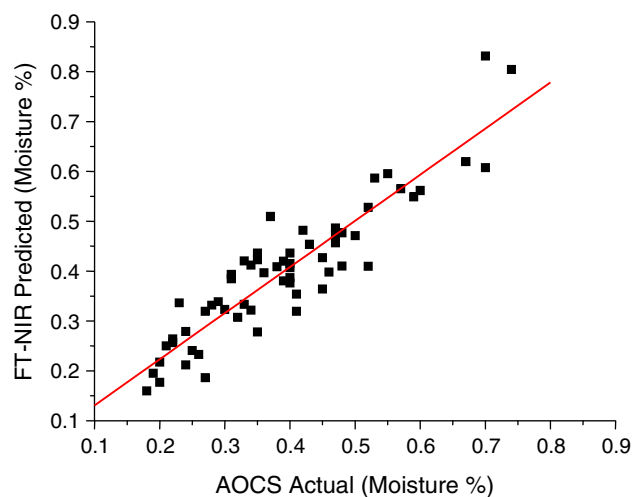
Fig. 7 Plot of FT-NIR predicted values versus AOCS reference method values for moisture in lecithin for the independent test set validation

of concentration values for acid value and moisture across the dynamic range of concentration values for these two parameters. Statistical analysis results for soybean oil refinery by-products calibrations are listed in Table 2. Good results were obtained for moisture and acid value with an R^2 of 0.988 and 0.979 and RPD values of 8.71 and 6.95, respectively, with slightly poorer results obtained for iodine value with an R^2 and RPD of 0.914 and 3.40, respectively. For acid values, the RMSECV was much higher for soybean oil refinery by-products, 6.59 (mg KOH/g), as compared to

lecithin, 0.588 (mg KOH/g), which can be explained by the much larger variation in the types of samples used in the soybean oil refinery by-products calibration and by the much larger variation in the range of acid values in the soybean oil

refinery by-products data set (21.7–210.9 mg KOH/g) relative to the lecithin data set (18.1–29.8 mg KOH/g). The concentration range for moisture values for the soybean oil refining by-products data set and lecithin data set were more similar and therefore gave closer RMSECV values of 0.046 and 0.050%, respectively. Overall, these results demonstrated the feasibility of employing FT-NIR spectroscopy to measure acid value, moisture and iodine value in soybean oil refining by-products.

This study demonstrates the feasibility of employing an FT-NIR spectrometer for measuring key parameters during lecithin production and soybean oil refining. Calibrating the FT-NIR instrument requires a data set of samples that represent the expected variability of the components and parameters of interest. More robust calibrations are obtained by using real production samples, and variability in the sample data set can be introduced by blending samples directly in the production tank. Consideration must be taken as to how reference data is acquired and minimizing variability and errors in the reference data will improve FT-NIR calibrations. Both spectral outliers and concentration outliers must be removed from the calibration data set to obtain robust calibration models for each parameter of interest. Once the calibrations are in place, FT-NIR is a rapid and easy to use method for the quality control of lecithin and soybean oil refining by-products in a production environment. This can lead to time and cost savings as multiple parameters can be measured quickly and simultaneously, eliminating the need to use several costly wet chemical analysis methods. In addition, because the results for sample analysis can be obtained almost immediately after drawing the sample from the production line, operators can diagnose problems quickly before production is compromised. Overall, using FT-NIR spectroscopy as a quality control tool for lecithin production and soybean oil refining can greatly benefit producers.

References

1. Hammond EG, Johnson LA, Su C, Wang T, White PJ (2005) Soybean oil. In: Shahidi F (ed) Bailey's industrial oil and fat products, 6th edn. Wiley, New York, pp 577–653
2. Mounts TL, Khym FP (1980) Refining. In: Erickson DR, Pryde EH, Brekke OL, Mounts TL, Falb RA (eds) Handbook of soy oil processing and utilization. American Soybean Association & American Oil Chemists' Society, St. Louis, pp 89–103
3. Kaffka KJ, Norris KH, Peredi J, Balogh A (1982) Attempts to determine oil, protein, water and fiber content in sunflower seeds by the NIR technique. Acta Aliment 11(3):253–269
4. Williams P (2007) Grains and seeds. In: Ozaki Y, McClure FW, Christy AA (eds) Near-infrared spectroscopy in food science and technology. Wiley, Hoboken, pp 165–217
5. Ciurczak EW (1995) Use of near-infrared spectroscopy in cereal products. Food Test Anal 1(2):35–39
6. Nadai BT, Mihalyi-Kengyel V (1984) Investigations of different equations predicting moisture, fat and protein content of raw meat by NIR-technique. Acta Aliment 13(4):343–353
7. Frankhuizen R (2008) NIR analysis of dairy products. Pract Spectrosc 35:415–437
8. Azizian H, Kramer JKG (2005) A rapid method for the quantification of fatty acids in fats and oils with emphasis on trans fatty acids using Fourier transform near infrared spectroscopy (FT-NIR). Lipids 40(8):855–867
9. Li H, Van de Voort FR, Sedman J, Ismail AA (1999) Rapid determination of *cis* and *trans* content, iodine value, and saponification number of edible oils by Fourier transform near-infrared spectroscopy. JAOCS 76(4):491–497
10. Armenta S, Garrigues S, De la Guardia M (2007) Determination of edible oil parameters by near infrared spectrometry. Anal Chim Acta 596(2):330–337
11. White RL, Coffey PJ, Covey JP, Mattson DR (1983) High-performance FTIR using a corner cube interferometer. Am Lab 15(11):90–93
12. White RL (1985) Performance of an FT-IR with a cube-corner interferometer. Appl Spectrosc 39(2):320–326
13. Rathjen T, Lange R, Steinhart H (1998) Analysis of lecithin by near-infrared spectroscopy. Fett/Lipid 100(8):358–363
14. Lange R, Schipmann K (2000) Application of NIR spectroscopy in lecithin production and manufacturing. Lipid Technol 1:4–17
15. Cox R, Lebrasseur J, Michiels E, Buijs H, Li H, van de Voort FR, Ismail AA, Sedman J (2000) Determination of iodine value with Fourier transform-near infrared based global calibration with disposable vials: an international collaborative study. JAOCS 77(12):1229–1234