Multivariate Determination of Cloud Point in Palm Oil Using Partial Least Squares and Principal Component Regression Based on FTIR Spectroscopy

G. Setiowaty and Y.B. Che Man*

Department of Food Technology, Faculty of Food Science and Biotechnology, Universiti Putra Malaysia, UPM 43400, Serdang, Selangor D.E., Malaysia

ABSTRACT: A rapid FTIR spectroscopic method was developed for quantitative determination of the cloud point (CP) in palm oil samples. Calibration samples were prepared by blending randomized amounts of palm olein and palm stearin to produce a wide range of CP values ranging between 8.3 and 47.9°C. Both partial least squares (PLS) and principal component regression (PCR) calibration models for predicting CP were developed by using the FTIR spectral regions from 3000 to 2800 and 1800 to 1600 cm^{-1}. The prediction capabilities of these calibration models were evaluated by comparing their standard errors of prediction (SEP) in an independent prediction set consisting of 14 palm oil samples. The optimal model based on PLS in the spectral range 1800–1600 cm^{-1} produced lower SEP values (2.03°C) than those found with the PCR (2.31°C) method. FTIR in conjunction with PLS and PCR models was found to be a useful analytical tool for simple and rapid quantitative determination of CP in palm oil.

Paper no. J10534 in *JAOCS 81,* 7–11 (January 2004).

KEY WORDS: Chemometrics, cloud point, FTIR, palm oil, PCR, and PLS.

Palm oil, which consists mainly of TAG of palmitic and oleic acids, is semisolid at room temperature. Its composition and properties, however, can be modified by various processing techniques to produce a series of different products that can be tailored for specific uses. Among the major products of palm oil is a variety of oils and fats that are produced through dry fractionation, which separates palm oil into a liquid palm olein and a solid palm stearin by means of cooling (1). Refined, bleached, and deodorized (RBD) palm olein, commonly known as cooking oil, is composed of 46% saturated FA, 43% monounsaturated FA, and 11% PUFA (2). It is liquid at temperatures above 25°C but starts to crystallize and becomes cloudy at lower temperatures. Cloud is the visible suspension of stearin crystals normally seen in a container of palm olein stored at low temperature (<8°C). Cloud points (CP) were determined according to AOCS Method Cc 6-25, which is currently used for routine analysis (3). Although relatively simple, the AOCS method is labor-intensive and tedious.

In our laboratory, work is ongoing to develop rapid automatable methods for the analysis of palm oil and its products based on FTIR spectroscopy. This technology has several advantages over conventional dispersive-based IR spectroscopy. FTIR technology, which is based on interferometry, provides enhanced energy throughput and a better signal-to-noise ratio, and incorporates substantial computing, chemometric, and macroprogramming capabilities. Coupled with an attenuated total reflectance (ATR) sampling technique, FTIR provides a simple and convenient means of acquiring spectral information from a wide variety of samples, many of which are not readily amenable to IR analysis with the more common transmission sampling techniques such as IR window cells. For instance, window cells require that samples be predissolved before being injected into the cell. In addition, the presence of interference fringes, due to air bubbles in the cell, may mask the absorption bands of interest.

FTIR methods have been developed to measure TBARS (4–6), moisture content (7), iodine value (8), FFA (9,10), PV (11,12), anisidine value (13), and slip melting point (14). Chemometric approaches, such as partial least squares (PLS) and principal component regression (PCR), also have been applied in developing quantitative FTIR methods. The overall objectives for this study were (i) to develop a rapid FTIR method in conjunction with an ATR method to determine CP in palm oil with the aid of PLS and PCR techniques for multivariate analysis and (ii) to compare the prediction errors obtained by PLS and PCR.

MATERIALS AND METHODS

Materials. Palm oil products, such as RBD palm olein and RBD palm stearin, were obtained from a local refinery [Ngo Chiew Hong (M) Sdn. Bhd., Kuala Lumpur, Malaysia]. All chemicals used were of analytical grade.

Analytical methods. A set of 74 standards were prepared by blending palm olein and palm stearin (w/w) to obtain oil blends with a wide range of CP values. The CP of each oil was analyzed in duplicate using AOCS Standard Method Cc 6-25 in order to obtain the actual CP data (3).

Instrumentation. All spectra were collected on a Perkin-Elmer 1725X series FTIR spectrometer (PerkinElmer Corporation, Norwalk, CT) interfaced to a PerkinElmer model 7300 professional computer and run under Infrared Data Management System (IRDM) software. This instrument was equipped with IR source, sealed and desiccated interferometer, and room-temperature deuterated triglycine detector. An

*To whom correspondence should be addressed.
E-mail: yaakub@fsb.upm.edu.my

overhead ATR accessory was built into a dedicated sampling station. The accessory comprised transfer optics within a desiccated chamber sealed from the atmosphere by two potassium bromide windows. Through these windows the IR radiation could be directed into the detachable ATR element. The element was an 11-reflection zinc selenide crystal mounted into a plate with a shallow trough for sample containment. The crystal geometry was a 45º parallelogram with mirrored angled faces. Owing to the presence of the potassium bromide windows, the ATR plate could be removed for cleaning without allowing water vapor into the spectrometer.

Spectral acquisition. All samples were equilibrated to 65°C in a conventional oven (Memmert GmbH & Co. KG, Schwabach, Germany) and applied with a soft-tipped disposable pipette (1 mL) to ensure complete coverage to the top ATR element. All spectra were recorded from 4000 to 770 cm^{-1} at a resolution of 4 cm^{-1}. For each spectrum, 64 interferograms were co-added before Fourier transformation and zero-filled to give a data point spacing of approximately 2 cm^{-1} in the frequency domain. Normal Beer apodization was then employed. Each sample single-beam spectrum was ratioed out to a single-beam spectrum of the clean ATR plate collected under identical conditions and converted into absorbance units. The ATR plate was thoroughly cleaned between each sample by removing the previous sample with tissue and cleaning with hexane, distilled water, and finally acetone.

The spectral data were stored on disk for subsequent PLS and PCR calibration development. The Spectrum Quant+ (Version 4.1, PerkinElmer) software package was used for spectral analysis and PCR and PLS calibrations. Correlation and variance spectra were examined to determine the spectral regions in the calibration set where most of the changes took place. These regions were then explored for calibration development. Each calibration was assessed by using the "leave-one-out" cross-validation procedure and optimized in terms of the appropriate number of factors in the predicted residual error sum of squares (PRESS) test. The predicted values of the parameter in each sample were compared with the actual values of the parameter in this reference sample, and the PRESS was calculated. To select the number of factors in the PLS and PCR algorithms, in order to model the system without overfitting the concentration data, the PRESS was calculated in the same manner each time a new factor was added to the PLS or PCR models. The calibration was considered optimized when the PRESS was minimized. However, using the number of factors (H^*) that yields a minimum PRESS usually leads to some overfitting. A better criterion for selecting the number of factors involved a comparison of PRESS from models with fewer than h^* factors. The model selected is that with the fewest number of factors such that PRESS for that model was not significantly greater than PRESS from the model with h^* factors. The F statistic was used to make the significance determination. Haaland and Thomas (15) determined that an F-ratio probability of 0.75 is a good choice. We selected as the optimal number of factors for the PRESS value with the F-ratio probability of which drops below 0.75.

Data analysis. Seventy-four spectra of palm oil samples were used to assess the performance of the PLS and PCR techniques. The adequacy of the regression model for making predictions was determined by graphic analysis of the residuals and examining the differences between the reference and predicted CP values for systematic trends. The results of the 74 samples were split into two sets of data. The first set corresponded to 60 calibration samples, serving as the model-building data set, and the remaining 14 samples served as validation data, which were used to evaluate the predictive ability of the selected model. The second set from the same sample of data (74 samples) was used to build the calibration models and assessed their performance by a leave-one-out cross-validation step. Generally, the number of cases in the model-building set should be 6–10 times the number of factors (16). The performance of the different models was evaluated by comparing their standard error of calibration (SEC), the standard error of prediction (SEP), and the standard error of cross-validation (SECV) values.

RESULTS AND DISCUSSION

Analytical methods. The calibration standard analysis showed that the CP of the samples ranged from 8.3 to 47.9°C, with a mean of 28.12°C and SD of 11.79°C. This range covered almost all of the palm oil products specified by the Palm Oil Refiners Association of Malaysia (PORAM), i.e., RBD palm olein with minimum 8°C and RBD palm stearin with maximum 46°C (17).

Spectral analysis. Figure 1 presents offset FTIR absorbance spectra for 10 samples having CP values of 8.3, 12.4, 17.2, 21.8, 26.2, 30.6, 36, 40.4, 43.6, and 47.9°C. In the spectra shown, the only noticeable differences are in the CH_2/CH_3 stretching vibration (3000–2850 cm^{-1}) (18) and the ester linkage carbonyl band (1750–1730 cm^{-1}) (19). One way of assessing the representativeness of the calibration subset is by examining the variance spectrum, which is indicative of the spectral variability relative to the mean spectrum. The variance spectrum (Fig. 2) shows the spectral regions where spectral variations exist to a significant degree. Although the magnitudes of the variations are small, they are readily measured, and this information provides a starting point for chemometric analysis. A correlation spectrum was generated from the variance spectrum by using the CP data as an aid in searching for useful spectral regions that could be used to develop the predictive models (20). Although many combinations and permutations of wavelength regions can provide reasonable results (of positive correlation between absorbances and reference CP values), the regions determined to be optimal for the palm oil samples were the CH_2/CH_3 stretching vibration (3000–2800 cm^{-1}) and the ester linkage C=O stretching vibration (1800–1600 cm^{-1}).

Statistical multivariate models. Figure 3 shows the variations in PRESS values obtained with a different number of factors included in the PLS calibration model created with the use of the calibration subset in the spectral regions tested. The

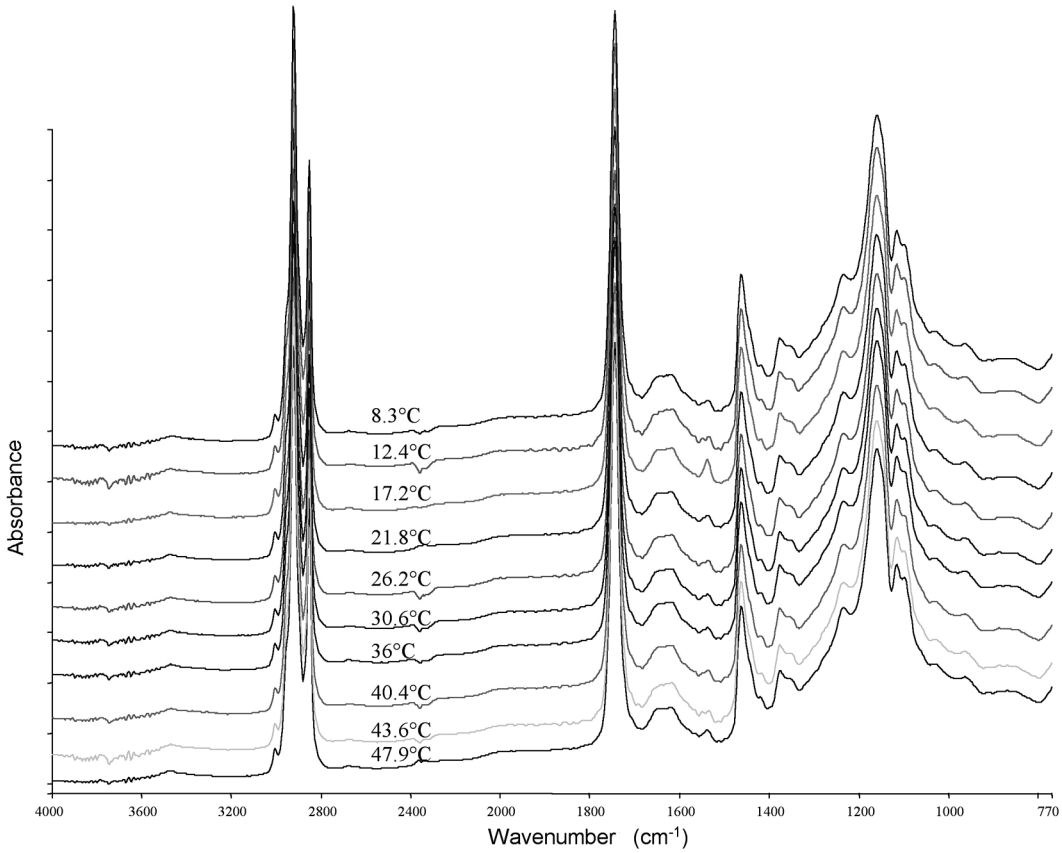

FIG. 1. Representative FTIR spectra of palm oil blend samples obtained from palm olein/palm stearin blends having cloud point values between 8.3 and 47.9°C.

calibration models with 4–8 factors yielded the optimal PRESS values (117.93–788.87°C) and were therefore selected as the optimal number of factors for this technique. A similar analysis with PCR indicated that 6–10 factors were required to obtain the minimum PRESS values (114.6–488.03°C). PLS and PCR are two multivariate whole-spectrum methods based on inverse modeling, commonly used for estimation of analyte concentration in multicomponent mixtures. These chemometric methods

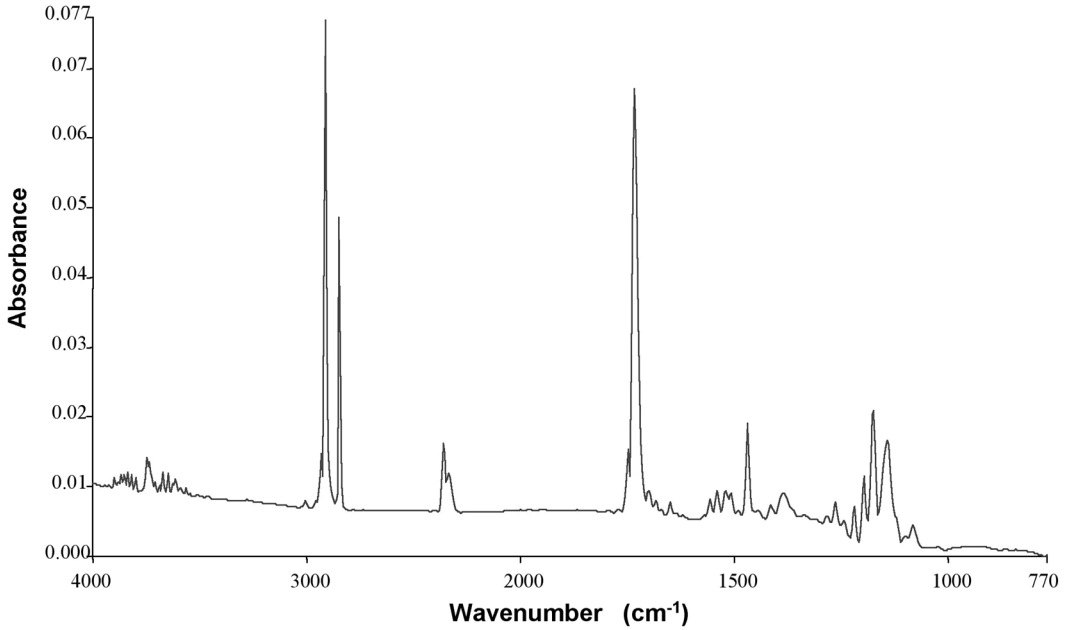

FIG. 2. The variance spectrum obtained for 60 palm oil samples, illustrating the regions where spectral variation occurs relative to the overall mean spectrum.

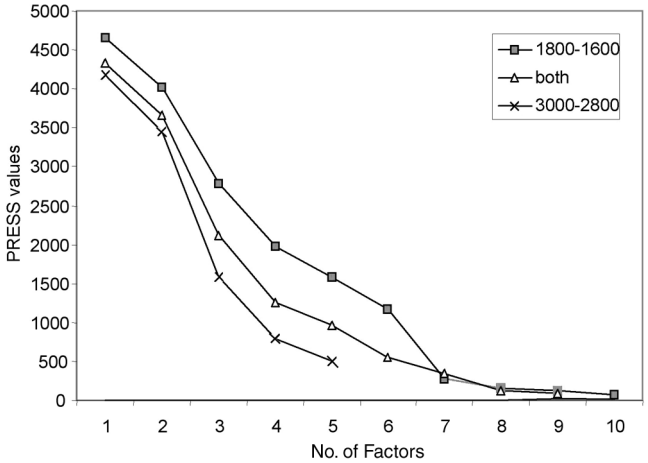

FIG. 3. Variation in predicted residual error sum of squares (PRESS) values with the number of partial least squares (PLS) factors used in the calibration models in the spectral regions of 3000–2800 cm^{-1}, 1800–1600 cm^{-1}, and both regions combined.

can effectively model the relationship between IR absorbances and concentrations of an unknown component, without knowledge of the concentrations of the other constituents in the sample and in spite of the nearly linear relationship between absorbances at differences wavelengths. Both PLS and PCR multivariate methods consist of two steps: (i) spectral data compression and (ii) linear regression of the compressed spectral features on the concentration of the analyte of interest. In PCR, no concentration information is used for the first step, and the spectral variations are compressed into orthogonal scores (21). Unlike PCR, PLS models used the CP data information to calculate scores and therefore required fewer factors for the optimal model.

To assess the adequacy of the regression models for making predictions, graphic analysis of the residuals was performed. Figure 4 shows the concentration residuals plotted against the predicted CP values of the prediction set for an eight-factor PLS model created with the use of the calibration subset in the spectral range of 1800–1600 cm^{-1}. This plot

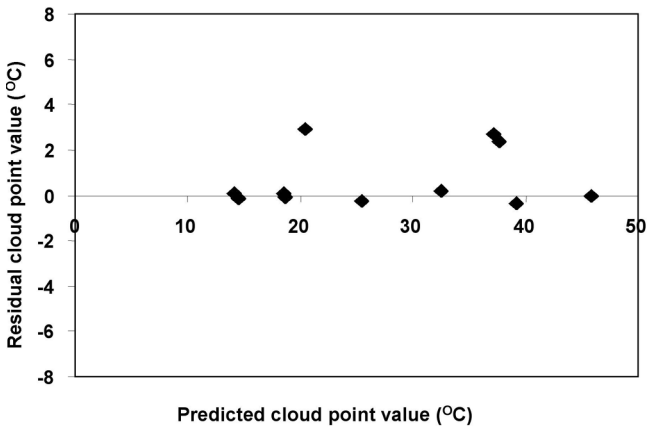

FIG. 4. Distribution of concentration residuals with the predicted cloud point values for the prediction set obtained using eight PLS factors (3000–2800 cm^{-1}). For abbreviation see Figure 3.

TABLE 1
Error Predictions Obtained with Different PLS and PCR Models and Applied to the Prediction Subset on the Chosen Spectral Ranges

Spectral ranges	SEC_{PLS} (°C)[a]	SEP_{PLS} (°C)[b]	SEC_{PCR} (°C)	SEP_{PCR} (°C)
3000–2800 cm^{-1}	3.41 (4)[c]	3.63	2.19 (6)	2.79
1800–1600 cm^{-1}	1.56 (8)	2.03	1.23 (10)	2.31
3000-2800 and 1800–1600 cm^{-1}	1.33 (8)	3.85	2.45 (8)	2.85

[a]SEC, standard error of calibration.
[b]SEP, standard error of prediction.
[c]Numbers in parentheses show the factor number for the calibration model built. PLS, partial least squares; PCR, principal component regression.

showed that all samples clustered closely around the zero line, whereas three samples show a residual greater than 2°C.

Comparison of models. Table 1 summarizes the SEC and SEP results obtained with different PLS and PCR models. It seemed that the prediction errors obtained with cross-validation (SECV) of PLS and PCR (Table 2) in the complete sample population were smaller than those obtained with models using only a calibration subset to predict concentrations in the independent prediction set. Therefore, we concluded that the model prediction performance given by the two methods was similar and that this was a good sign of model robustness. The optimal PLS analysis performed in the spectral region of 1800–1600 cm^{-1} was slightly better than those applied for either the 3000–2800 cm^{-1} region or both used simultaneously. The application on that region improved the prediction capability by 10 to 30%. The prediction capabilities of PLS and PCR could be compared by calculating the square of their SEP ratio into the F critical value. With this approach, the optimal CP prediction with PLS was not significantly better than PCR in the region of 3000–2800 cm^{-1} (F-ratio = 1.69), 1800–1600 cm^{-1} (F-ratio = 1.29), and both regions (F-ratio = 1.82) at the 95% confidence level (F-critical = 2.48).

Figure 5 shows the correlation between predicted CP in palm oil blend samples from the validation set obtained with the optimal PLS model applied in the 1800–1600 cm^{-1} region and the actual CP data obtained by AOCS Method Cc 6-25 (3). The ideal predictions were represented by a straight line that had a slope of 1 and passed through the origin. The deviation of the predictions from the ideal values could be expressed statistically in terms of SE (1.994) and R^2 (0.972).

Results from this study indicated that FTIR spectroscopy was a useful technique for measuring palm oil parameters

TABLE 2
Error Predictions Obtained from the Cross-Validation Step with Different PLS and PCR Models and Applied to All Palm Oil Blends in the Chosen Spectral Ranges

Spectral range	$SECV_{PLS}$ (°C)[a]	$SECV_{PCR}$ (°C)
3000–2800 cm^{-1}	3.15 (5)[b]	2.25 (8)
1800–1600 cm^{-1}	1.80 (8)	1.30 (11)
3000–2800 and 180–1600 cm^{-1}	1.18 (10)	1.28 (11)

[a]SECV, standard error of cross-validation.
[b]Numbers in parentheses show the factor number for the calibration model built. For other abbreviations see Table 1.

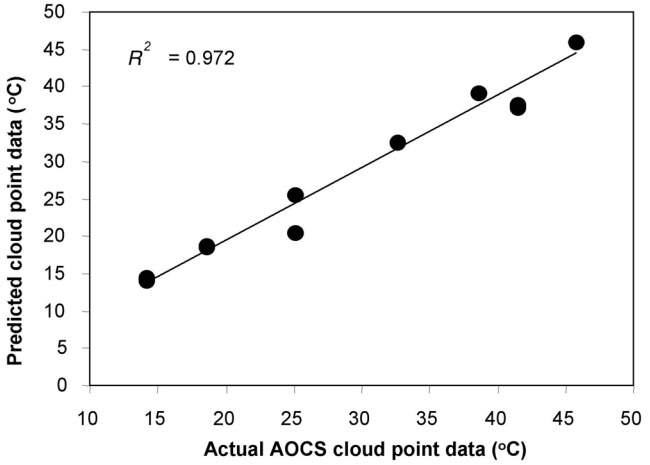

FIG. 5. Graph showing the correlation between the predicted cloud point of palm oil samples using the optimal PLS model and actual cloud point data. For abbreviation see Figure 3.

such as CP. The total analysis took less than 5 min once the spectrometer was precalibrated. Compared to the chemical method, FTIR spectroscopy was capable of measuring 100 samples in a day, so the cost of labor can be reduced dramatically.

The FTIR method developed, which was simple, fast, nondestructive, and safe, was based on the use of the FTIR absorbance spectrum in the spectral regions of 3000–2800 cm^{-1} and 1800–1600 cm^{-1}, respectively. The calibration models obtained from PLS and PCR gave high R^2 and low SEC. Those calibration models could be used for CP prediction for palm oil with low error predictions.

ACKNOWLEDGMENTS

The authors thank the Faculty of Food Science and Biotechnology, Universiti Putra Malaysia, for funding this research.

REFERENCES

1. Basiron, Y., Palm Oil, in *Bailey's Industrial Oil and Fat Products*, 5th edn., edited by Y.H. Hui, John Wiley & Sons, New York, 1996, Vol. 2, pp. 271–375.
2. Gunstone, F.D., *The Lipid Handbook*, Chapman & Hall, New York, 1986.
3. *Official Methods and Recommended Practices of the American Oil Chemists' Society,* 4th edn., edited by D. Firestone, American Oil Chemists' Society, Champaign, 1989, Method Cc 6-25.
4. Mirghani, M.E.S., Y.B. Che Man, S. Jinap, B.S. Baharin, and J. Bakar, Rapid Method for Determination of Malondialdehyde as Secondary Oxidation Product in Palm Olein System by FTIR Spectroscopy, *Phytochem. Anal. 13*:195–201 (2002).
5. Mirghani, M.E.S., Y.B. Che Man, S. Jinap, B.S. Baharin, and J. Bakar, Multivariate Calibration of Fourier Transform Infrared Spectra for Determining Thiobarbituric Acid-Reactive Substance Content in Palm Oil, *J. Am. Oil Chem. Soc. 78*:1127–1131 (2001).
6. Setiowaty, G., and Y.B. Che Man, A Rapid Fourier Transform Infrared Spectroscopic Method for the Determination of 2-TBARS in Palm Olein, *Food Chem. 81*:147–154 (2003).
7. Che Man, Y.B., and M.E.S. Mirghani, Determination of Moisture Content in Crude Palm Oil by Fourier Transform Infrared Spectroscopy, *J. Am. Oil Chem. Soc. 77*:631–637 (2000).
8. Che Man Y.B, G. Setiowaty, and F.R. van de Voort, Determination of Iodine Value of Palm Oil by Fourier Transform Infrared Spectroscopy, *Ibid. 76*:693–700 (1999).
9. Che Man, Y.B, and M.H. Moh, Determination of Free Fatty Acids in Palm Oil by Near-Infrared Reflectance Spectroscopy, *Ibid. 75*:557–564 (1998).
10. Che Man, Y.B., and G. Setiowaty, Application of Fourier Transform Infrared Spectroscopy to Determine Free Fatty Acid Contents in Palm Olein, *Food Chem. 66*:109–114 (1999).
11. Setiowaty, G., Y.B. Che Man, S. Jinap, and M.H. Moh, Quantitative Determination of Peroxide Value in Thermally Oxidized Palm Olein by Fourier Transform Infrared Spectroscopy, *Phytochem. Anal. 11*:74–78 (2000).
12. Moh, M.H., Y.B. Che Man, F.R. van de Voort, and W.J.W. Abdullah, Determination of Peroxide Value in Thermally Oxidized Crude Palm Oil by Near Infrared Spectroscopy, *J. Am. Oil Chem. Soc. 76*:19–23 (1998).
13. Che Man, Y.B., and G. Setiowaty, Determination of Anisidine Value in Thermally Oxidized Palm Olein by Fourier Transform Infrared Spectroscopy, *Ibid. 76*:243–247 (1999).
14. Setiowaty, G., and Y.B. Che Man, Determination of Slip Melting Point in Palm Oil Blends by Partial Least Squares and Principal Component Regression Modelling of FTIR Spectroscopic Data, *Ibid. 79*:1081–1084 (2002).
15. Haaland, D.M., and E.V. Thomas, Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information, *Anal. Chem. 60*:1193–1202 (1988).
16. Neter, J., W. Wasserman, and M.H. Kutner, *Applied Linear Statistical Models*, 3rd edn., Irwin, Homewood, IL, 1990.
17. Pantzaris, T.P., *Pocketbook of Palm Oil Uses*, 3rd edn., Palm Oil Research Institute of Malaysia, Kuala Lumpur, 1995.
18. Guillen, M.D., and N. Cabo, Relationships Between the Composition of Edible Oils and Lard and the Ratio of the Absorbance of Specific Bands of Their Fourier Transform Infrared Spectra. Role of Some Bands of the Fingerprint Region, *J. Agric. Food Chem. 46*:1788–1793 (1998).
19. van de Voort, F.R., K.P. Memon, J. Sedman, and A.A. Ismail, Determination of Solid Fat Index by Fourier Transform Infrared Spectroscopy, *J. Am. Oil Chem. Soc*. 73:411–416 (1996).
20. Fuller, M.P., G.L. Ritter, and C.S. Draper, Partial Least Squares Quantitative Analysis of Infrared Spectroscopic Data. Part I. Algorithm Implementation, *Appl. Spectrosc*. *42*:217–227 (1988).
21. Bhandare, P., Y. Mendelson, R.A. Peura, G. Janatsch, J.D. Kruse-Jarres, R. Marbach, and H.M. Heise, Multivariate Determination of Glucose in Whole Blood Using Partial Least Squares and Artificial Neural Networks Based on Mid Infrared Spectroscopy, *Ibid. 47*:1214–1221 (1993).

[Received January 6, 2003; accepted September 9, 2003]