ELSEVIER

Food Chemistry 67 (1999) 193–198

Food Chemistry

www.elsevier.com/locate/foodchem

Analytical, Nutritional and Clinical Methods Section

Multivariate calibration of Fourier transform infrared spectra in determining iodine value of palm oil products

Y.B. Che Man*, G. Setiowaty

Faculty of Food Science and Biotechnology, Universiti Putra Malaysia, 43400 UPM Serdang Selangor DE, Malaysia

Received 7 December 1998; received in revised form and accepted 1 March 1999

Abstract

The Fourier transform infrared (FT–IR) spectra of palm oil samples, in the range between 3025 and 2992 cm^{-1}, were used to compare different multivariate calibration techniques for quantitative iodine value (IV) determination. Forty-two spectra of palm oil with IV ranging between 53 and 65 were used to create calibration models based on partial least squares (PLS) and principle component regression (PCR) methods using different baseline types. The methods were compared with respect to the number of factors, coefficient of determination (R^2) and accuracy of estimation. The standard error of prediction (SEP) ratios were calculated to compare the prediction capabilities of these calibration methods. The calibration models generated the number of factors from 3 to 7, R^2 of 0.94443 to 0.98853, standard error of estimation (SEE) of 0.32 to 0.69 and SEP ratios of 1.43 to 13.48.

1. Introduction

Multivariate calibration methods have been used successfully by near-, mid-infrared and Raman spectrometry for quantitative analysis (Haaland, Higgins & Tallant, 1990; Martens & Naes, 1989). Common multivariate quantitative methods that have been applied include ordinary least-squares (classical least-square and multiple linear regression), partial least-squares (PLS) and principal component regression (PCR) methods (Marjoniemi, 1992). PCR and PLS are becoming more widely used as a growing number of researchers recognize the advantages of these methods over other methods.

Multivariate methods have also been used in food industry (Cadet, Bertrand, Robert, Maillot, Dieudonne & Rouch, 1990; Headley & Hardy, 1989; Liu, Espen & Adams, 1987; Robert, Bertrand, Devaux & Grappin, 1987; van Rooyen, Marais & Ellis, 1985). In palm oil research, some workers reported utilisation of multivariate analysis on processing data obtained by FTIR and NIR spectroscopies. Teo and Goh (1995) reported the use of the multiple linear regression (MLR) method for determining iodine value (IV), free fatty acid content (FFA) and saponification value. Gee (1995, 1996) has also successfully applied the PLS method to calculate IV and FFA content in palm oil. Che Man and Moh (1998) have utilized multivariate statistical methods such as MLR for quantitative NIR spectroscopy. Their works clearly demonstrate the applicability of multivariate methods to predict FFA content in palm oil. However, no previous work reported the application of two or more multivariate methods and comparison of them on the same data set.

In the present study, both PLS and PCR methods, and using different baseline types, are applied to the infrared spectra of a set of palm oil products to readily determine their IV without any preliminary, tedious and subjective spectral corrections. The use of multivariate calibration methods is also able to provide estimates of the precision of the analysis, and yields important qualitative and diagnostic information. Thus, multivariate calibration methods can provide better accuracy, precision, and significantly more information in considerably less time than previous data analysis methods. The aim of this study was to compare the standard error of prediction (SEP), coefficient of determination (R^2), standard error of estimation (SEE) and *F*-ratios from both methods.

2. Materials and methods

2.1. Sample preparation

A set of 21 palm oil products samples was obtained from four local palm oil refineries. These samples have been analyzed for their IV value according to standard

* Corresponding author. Tel.: +603-948-6101; fax: +603-942-3552.

0308-8146/99/$ - see front matter
PII: S0308-8146(99)00102-8

methods (AOCS, 1990) in their respective refineries' laboratories.

2.2. Instrumentation

A Perkin–Elmer Model 1000 Paragon FT–IR spectrometer (Perkin–Elmer Instrument Corporation, USA), equipped with a room temperature deuterated triglycerine sulphate (DTGS) detector was employed to acquire the spectra by BaF_2 transmission cell having pathlength of 100 μm. The instrument was also equipped with a heated sample handling accessory set to 65°C by two thermostats. All samples were preheated to 65°C on the heater prior to loading to minimize temperature perturbations in the cell. Freshly distilled hexane was used to rinse and clean the cell to avoid oil build-up on the cell windows after running every sample. The FT–IR system was purged with pure nitrogen gas at a slow rate for 30 minutes to reduce interference of water vapour and carbon dioxide. The FT–IR system was interfaced with a DEC 5150 Venturis FX PC. Spectra recording was carried out by Perkin–Elmer 1.5 software and stored to disk under JCAMP file for further processing. A background spectrum was recorded with an air spectrum before every standard sample spectrum. The sample spectrum is automatically ratioed against the background spectrum before being displayed on the computer monitor and that spectrum is automatically stored in the computer hard disk. Each sample spectrum was of 16 scans at 4 cm^{-1} resolution and a gain of 2. The spectrum obtained is used for quantitative analysis by Nicolet Turbo Quant Analyst version 1.5 Calibration and Prediction package software (Nicolet Instrument Co., Madison, WI). A Calibration and quantitative analysis was performed using PLS and PCR methods. The prediction error sum of squares (PRESS) values were calculated for each factor with the 'leave-one-out' cross-validation to determine the optimal number of factors to be included in the calibration model.

2.3. Statistical multivariate methods

PCR and PLS are two multivariate full spectrum methods based on inverse modelling, commonly used for estimation of analyte concentration in multicomponent mixtures. Both PCR and PLS belong to the so-called indirect multivariate calibration methods, in which one or more components are selectively determined from data that are nonselective due to various kinds of interference. Only the concentrations of the modelled components need to be known, and the level of each interference has to vary sufficiently in the samples.

In PCR, the spectral variables in absorbance are first modeled by 'data compression' onto a few principal component regression factors that account for most of their variations. The chemical data are then modeled by multiple linear regression (MLR) on these regression factors.

In PLS, the chemical data are first used to identify a pattern in the spectral data that correlates with the chemical data. This procedure yields the spectrum of loading weights for spectral data for this factor. The level of this estimated pattern in each sample's spectral pattern is then estimated (the factor scores). These factor scores are used as regressors for modeling both the chemical data and the spectral data as accurately as possible. This step yields loadings for chemical data as well as another set of loading spectra used for modelling spectral data itself.

3. Results and discussion

The iodine values (IV) of the samples analyzed ranged from 53 to 65. Palm Oil Refiners Association of Malaysia (PORAM) standard specifications for IV of processed palm oil are as follows: palm oil, 50 to 55; palm olein, min. 56; and super olein, min. 60 (PORIM, 1995).

Fig. 1 illustrates a representative set of mid-IR spectra of 10 palm oil samples obtained with the FT–IR/transmission cell set up in the range of 4000 to 750 cm^{-1}. In general, these spectra appear almost similar to each other. IV, reflecting the degree of unsaturation, is known to have prominent absorption peaks at 3006, 1654 and 968 cm^{-1} due to the *cis* =CH stretching, *cis* C=C stretching and *trans* HC=CH bending, respectively (Guillen and Cabo, 1997). The strong absorption peaks at around 2900 to 2850, 1746 and 1465 cm^{-1} were due to CH stretching, C=O stretching and CH scissoring, respectively. In order to limit the analysis to the spectral region containing the most information related to IV and reduce interferences due to spectral contribution of other components, the region between 3025 and 2992 cm^{-1} (Fig. 2) was selected for analysis on the basis of positive correlation between absorbances and reference chemical IV. Around the 3006 cm^{-1} region, there were significant differences of the absorbances of these spectra. The high IV implies a high content of unsaturated fatty acids (*cis* and *trans*). It contributes a high absorbance value in the absorption region of *cis* c =CH stretching.

Thirty-two samples were used to develop the calibration and 10 samples were used as a validation set for the PLS and PCR quantitative methods. In order to develop the methods and to compare the results, we based our decisions on the following statistical tools: the prediction residual sum of squares (PRESS), the standard error of prediction (SEP), coefficient of determination (R^2), accuracy of estimation or the standard error of estimation (SEE), and SEP ratio (Geladi & Kowalski, 1986). As suggested by Haaland and Thomas in Luinge, de Koeijer and van der Maas (1993), the optimum model was the one with the smallest number of factors with a

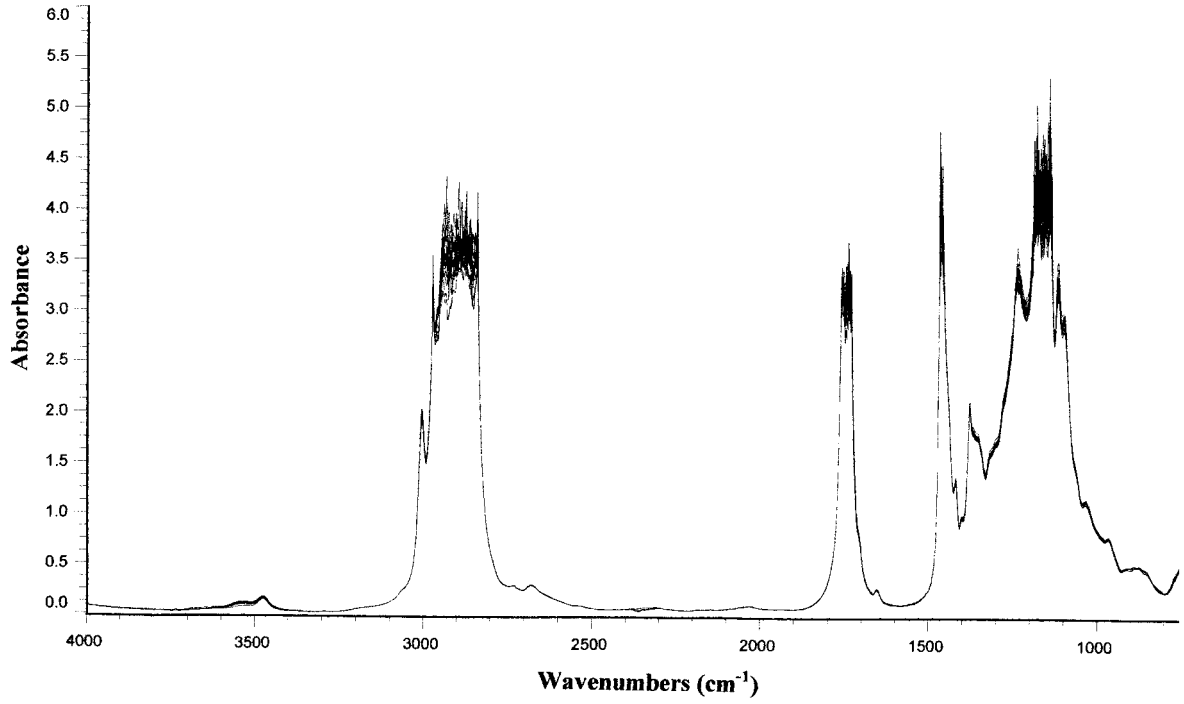

Fig. 1. Transmission/Fourier transform infrared spectra of palm oil at 4000–750 cm^{-1}.

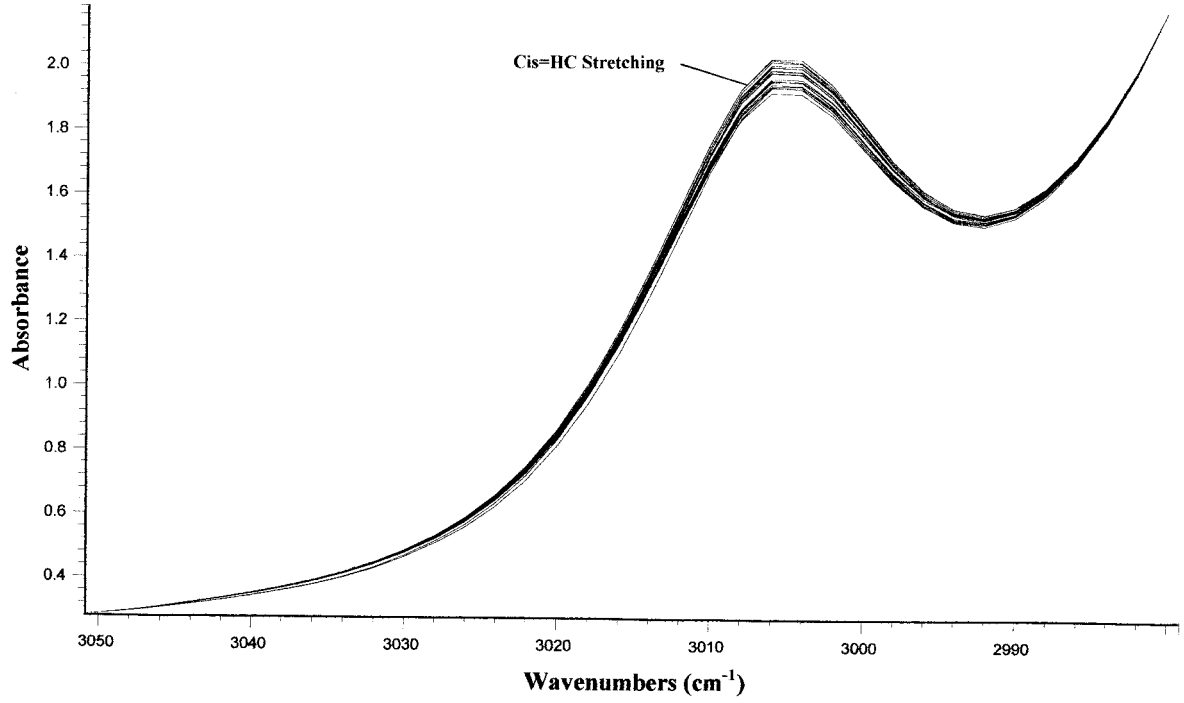

Fig. 2. Overlay spectra of palm oil at wavelength of 3050–2980 cm^{-1}.

PRESS not significantly differing from the minimum PRESS found. The standard error of prediction (SEP) was obtained from the root squares by dividing PRESS by the number of validation samples. The SEP ratio gives us the decision as it directly gives information on the ability of the method to predict the concentration of unknown samples. The accuracy of IV determined by PLS and PCR methods was affected by sample set. Poor sample set yields large error of estimation from the calibration model during regression. Poor sample set can be caused by inaccurate titration result (standard method) or poor spectra.

3.1. PLS

The PLS calibrations by cross-validation were carried out on the palm oil spectra in the calibration data set. The optimal number of factors indicated by the PRESS value were four for one point (method 3) and zero (method 1) baseline types and three for linear removed baseline type (method 2). PLS estimations obtained by multiplying the calibration coefficient vector and each spectrum in the calibration model are shown in Figs. 3–5. Accuracies of the estimations were good, with a standard error estimation (SEE) of 0.32 and coefficient of determination (R^2) of 0.98849 for method 1 and SEE of 0.34 and R^2 of 0.98615 for method 3 (Table 1). When estimations were made from method 2, the magnitudes of the errors increased. As seen in Table 1, this method has an SEE of 0.69 and R^2 of 0.94443.

3.2. PCR

Figs. 6–8 show the estimated IV of palm oil samples obtained using the PCR method at different baseline type. The standard deviation of the residuals between chemical and predicted IV, SEE, were 0.32 and 0.34 for method 4 and 6, respectively, while the R^2 were 0.98853 and 0.98541, using 4 and 6 factors, respectively. When predicted IV were regressed against chemical IV by method 5, IV estimates were not as accurate as those from other methods. Table 1 shows that SEE increased to 0.63, while R^2 decreased to 0.95330, using 7 factors.

To compare the predictive capabilities of all methods objectively, one should make a comparison between the ratio of SEP_1 and SEP_2 and the $F_{\alpha,m,m}$ statistic (F critical) (Luinge et al., 1993). The variable m represents the number of spectra used in the analyses. The 1-α defines the confidence level at which the null hypothesis can be

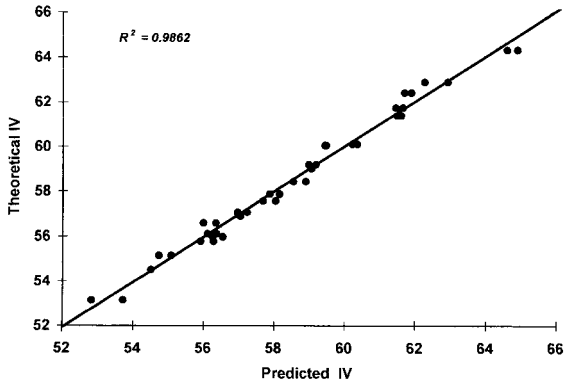

Fig. 5. Theoretical versus predicted iodine values calculated with PLS calibration method using one point baseline type at 3100 cm^{-1}.

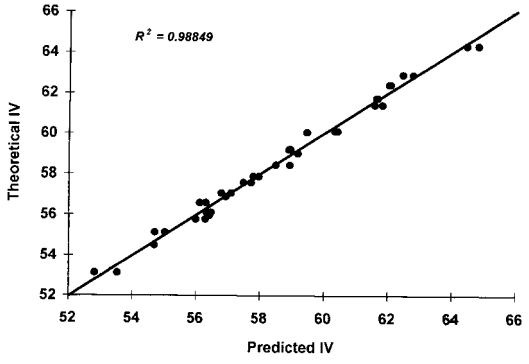

Fig. 3. Theoretical versus predicted iodine values calculated with PLS calibration method.

Table 1
Standard error of prediction (SEP), number of factors, correlation coefficient (R^2) and accuracy of IV of palm oil products obtained by PLS and PCR methods

No.	Method	Baseline type	Factor	SEP	R^2	Accuracy
1.	PLS	None	4	0.042	0.98849	0.32
2.	PLS	Linear removed	3	0.566	0.94443	0.69
3.	PLS	One point (3100)	4	0.114	0.98615	0.34
4.	PCR	None	5	0.042	0.98853	0.32
5.	PCR	Linear removed	7	0.180	0.95330	0.63
6.	PCR	One point (3100)	6	0.126	0.98541	0.34

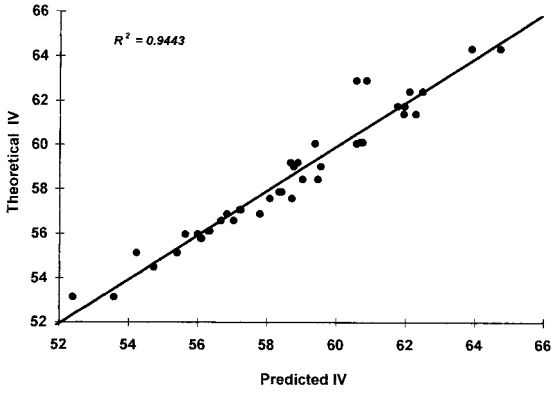

Fig. 4. Theoretical versus predicted iodine values calculated with PLS calibration method using linear removed baseline type.

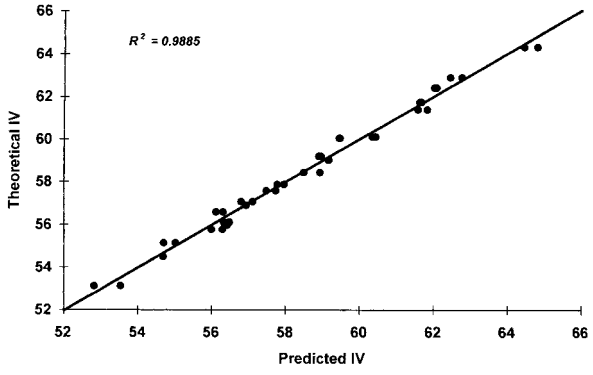

Fig. 6. Theoretical versus predicted iodine values calculated with PCR calibration method.

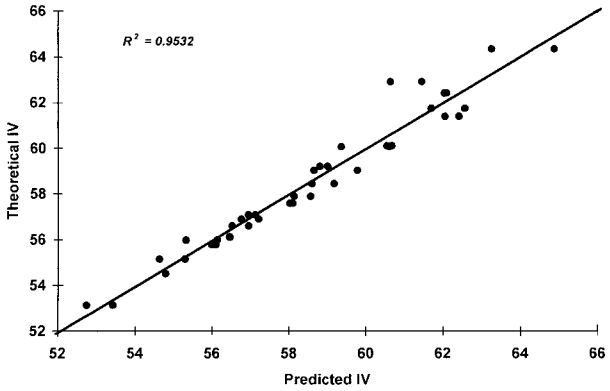

Fig. 7. Theoretical versus predicted iodine values calculated with PCR calibration method using linear removed baseline type.

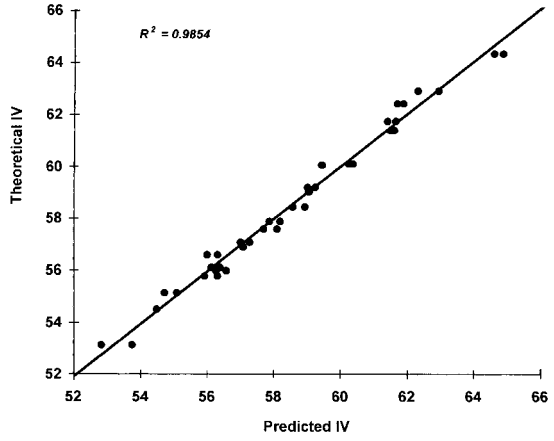

Fig. 8. Theoretical versus predicted iodine values calculated with PCR calibration method using one point baseline type at 3100 cm^{-1}.

Table 2
SEP ratios of different methods and *F* critical at 95% confidence level

Methods	SEP ratio	*F* critical
PLS		2.99
Baseline types		
Linear removed and none (2 and 1)	13.48	
Linear removed and one point (2 and 3)	4.96	
One point and none (3 and 1)	2.71	
PCR		
Baseline types		
Linear removed and none (5 and 4)	4.29	
Linear removed and one point (5 and 6)	1.43	
One point and none (6 and 4)	3.00	

rejected. This hypothesis states that neither method provides significantly better estimations. From Table 2, we can observe that the SEP ratio between method 2 and 1; method 2 and 3 are larger than *F* critical (2.99). PLS methods 1 and 3 clearly provide more accurate estimations than method 2 and both methods are not significantly different in predicting IV of palm oil samples. For PCR, it appears that method 4 can give the predicted IV more accurately than others. A comparison of SEP results from method 3 and 1, and method 1 and 4 reveals that the prediction error variances are not significantly different.

From this study, it can be concluded that the multivariate methods, PCR and PLS, are important in predicting iodine value of palm oil products. Applying the PCR method by zero baseline type and PLS method by one point and zero baseline types is superior to transmission data and gives similar results, either predicted IVs or SEP.

Acknowledgements

The authors thank Universiti Putra Malaysia for its financial support (IRPA No. 03-02-04-048) of this research and Perkin–Elmer Instrument (M) Sdn. Bhd. for providing the spectrometer used in this study.

References

AOCS (1990). *Official and tentative methods of the American Oil Chemists' Society* (15th ed.). Champaign, IL: American Oil Chemist Society.

Cadet, F., Bertrand, D., Robert, P., Maillot, J., Dieudonne, J., & Rouch, C. (1990). Quantitative determination of sugar cane sucrose by multidimensional statistical analysis of their mid infrared attenuated total reflectance spectra. *Applied Spectroscopy, 45*, 166–170.

Che Man, Y. B., & Moh, M. H. (1998). Determination of free fatty acids in palm oil by near-infrared reflectance spectroscopy. *Journal of the American Oil Chemists Society, 75*, 537–562.

Gee, P. T. (1995). Iodine value determination by FT-IR spectroscopy. *Malaysian Oil Science and Tecnology, 4*, 182–185.

Gee, P. T. (1996) Superior quality control by FT-IR spectroscopy. In *Proc. 1996 PORIM International Palm Oil Congress*, (pp. 13–17), Kuala Lumpur.

Geladi, P., & Kowalski, B. R. (1986). An example of 2-block predictive partial least-squares regression with simulated data. *Analytical Chimica Acta, 185*, 19–32.

Guillen, M. D., & Cabo, N. (1997). Infrared spectroscopy in the study of edible oils and fats. *Journal of the Science of Food and Agriculture, 75*, 1–11.

Haaland, D. M., Higgins, K. L., & Tallant, D. R. (1990). Multivariate calibration of carbon raman spectra for quantitative determination of peak temperature history. *Vibrational Spectroscopy, 1*, 35–40.

Headley, L. M., & Hardy, J. K. (1989). Classification of whiskies by principal component analysis. *Journal of Food Science, 54*, 1351–1354.

Lui, X., Espen, P. V., & Adams, F (1987). Classification of Chinese tea samples according to origin and quality by principal component techniques. *Analytical Chimica Acta, 200*, 421–430.

Luinge, H. J., de Koeijer, J. A., & van der Maas, J. H. (1993). Application of an automated interpretation system for infrared spectra, part II. Characterization of aromatic sulphone copolymers. *Vibrational Spectroscopy, 4*, 301–308.

Marjoniemi, M. (1992). Calibration of chromium tanning solution with different regression methods. *Applied Spectroscopy , 46*, 1908–1911.

Martens, H., & Naes, T. (1989). *Multivariate calibration*. New York: Wiley.

PORIM (1995). *Pocketbook of Palm Oil Uses*, (3rd ed.) Malaysia: Palm Oil Research Institute of Malaysia, Ministry of Primary Industries.

Robert, P., Bertrand, D., Devaux, M. F., & Grappin, R. (1987). Multivariate analysis applied to near infrared spectra of milk. *Analytical Chemistry, 59*, 2187–2191.

Teo, W. B., & Goh, E. M. (1995). Determination of the total unsaturation, free fatty acid content and saponification value in palm oil by Fourier transform infrared spectroscopy. *Malaysian Oil Science and Technology, 4*, 78–81.

van Rooyen, P. C., Marais, J., & Ellis, L. P. (1985). Multivariate analysis of fermentation flavour profiles of selected south african white wines. *Development of Food Science, 10*, 359–385.