

# Application of Multivariate Analysis in Mid-Infrared Spectroscopy as a Tool for the Evaluation of Waste Frying Oil Blends

Luciano Hocevar · Vitória R. B. Soares ·  
Fábio S. Oliveira · Maria Graças A. Korn ·  
Leonardo S. G. Teixeira

Received: 29 March 2011 / Revised: 6 October 2011 / Accepted: 26 October 2011 / Published online: 23 November 2011  
© AOCS 2011

**Abstract** Mid-infrared spectroscopy, in association with multivariate chemometric techniques, was employed for pattern recognition and the determination of the composition of waste frying oils (WFO); data are presented in terms of the percentage of soybean oil, palm oil and hydrogenated vegetable fat in frying oil blends. Principal component analysis (PCA) was performed using spectral data ( $3,000\text{--}600\text{ cm}^{-1}$ ) to discriminate between the samples containing 100% soybean oil, 100% palm oil, 100% hydrogenated vegetable fat groups and their blends. Additionally, the results indicated that partial least squares (PLS) models based on mid-infrared spectra were suitable as practical analytical methods for predicting the oil contents in WFO blends. PLS models were validated by a representative prediction set, and the root mean square errors of prediction (RMSEP) were 2.8, 4.7 and 5.5% for palm oil, soybean oil and hydrogenated vegetable

fat, respectively. The proposed methodology can be very useful for the rapid and low cost determination of waste frying oil composition while also aiding in decisions regarding the management of oil pretreatment and production routes for biodiesel production.

**Keywords** Waste frying oils · Principal component analysis · Partial least squares · Multivariate calibration

## Introduction

With a global focus on environmental awareness and increasing demands for energy, the concept of producing fuels from renewable resources that are environmentally acceptable has garnered significant attention in the research community. Biodiesel derived from vegetable oils or animal fats can be considered an alternative fuel source to fossil-based diesel [1, 2]. Biodiesel can be produced from a large variety of renewable lipid sources and has traditionally been obtained by alcoholysis of vegetable oils in the presence of an alkaline catalyst, methanol and heat.

A major concern regarding the production and commercialization of biodiesel is the high cost of raw materials, especially vegetable oil, a consideration that greatly limits widespread application. One way to reduce the cost of biodiesel is to employ low-quality feedstock, such as waste frying oils or fats, which are cheap, widely available and can be regarded as an attractive feedstock [3–5]. Additionally, the generation of waste oils and fats in many countries is significant and may result in environmental contamination if proper disposal methods are not employed [6].

Triacylglycerols are main components of vegetable oils and animal fats and have different physical and chemical properties. Fatty acid composition is the most important

---

L. Hocevar  
Escola Politécnica, Universidade Federal da Bahia,  
Av. Aristides Novis 2, Salvador, Bahia 40210-630, Brazil

V. R. B. Soares · M. G. A. Korn · L. S. G. Teixeira (✉)  
Instituto de Química, Universidade Federal da Bahia, Campus  
Universitário de Ondina, Salvador, Bahia 40170-115, Brazil  
e-mail: lsgt@ufba.br

V. R. B. Soares  
Instituto Federal de Educação, Ciência e Tecnologia Baiano,  
Km 2.5, BR 420, Santa Inês, Bahia 45320-000, Brazil

F. S. Oliveira  
Centro de Ciências da Saúde, Universidade Federal do  
Recôncavo da Bahia, Campo do Governo, Cajueiro,  
Santo Antônio de Jesus, Bahia 44574-490, Brazil

M. G. A. Korn · L. S. G. Teixeira  
INCT de Energia e Ambiente, Universidade Federal da Bahia,  
Salvador, Bahia 40170-115, Brazil

parameter and influences the properties of the oils, fats and the subsequently produced biodiesel [6].

Waste cooking oils are currently collected from large-scale food processing industries, restaurants and fast food shops. The waste oils consist of a mixture of residual frying oils and fats and often contain many impurities, such as free fatty acids, sterols and water. Free fatty acid and water contents negatively affect the alkaline catalyzed transesterification reaction and also interfere in the purification step due to the formation of soap. In addition, dependant on the source of the oil, the biodiesel produced will have a different composition of FAME and consequentially different physical and chemical properties [6].

Thus, if waste cooking oil were to be used as feedstock for biodiesel production, the oil composition must be considered. To this effect, a pretreatment strategy must be applied to the raw materials and the production route because the composition of the oil mixture will greatly affect the transesterification reaction and the resulting fuel quality.

Infrared spectroscopy has been used as an alternative, non-destructive analytical technique that allows the reliable determination of several properties without sample pretreatment [7]. The combination of infrared spectroscopy with various chemometric techniques provides a powerful tool for monitoring a variety of sample properties; this technique has seen increased interest based on its application in quality control [8]. The appealing aspects of methods that combine infrared spectroscopy with multivariate chemometric techniques has fostered its use in the analysis of virgin vegetable oil to determine: (1) the composition and the parameters of the oil [9], (2) to confirm claims concerning the geographic origin of the oils [10], and (3) to determine sample authenticity [11].

In the present study, we describe a technique based on mid-infrared spectroscopy (MIR) in association with the principal component analysis (PCA) method. This method was employed to build a screening model for the rapid supervised pattern recognition of waste frying oil (WFO) blends and aimed to find and interpret the hidden complex relationships between features in the data set. Additionally, partial least squares (PLS) multivariate calibration models were constructed to determine the content of each oil in the WFO samples. Waste frying oils consisting of soybean oil, palm oil and hydrogenated vegetable fat were employed in the preparation of the blends.

## Experimental

### Samples

WFO samples were collected from restaurants that used pure soybean oil (SO), palm oil (PO) or hydrogenated

vegetable fat (VF) for cooking. The WFO and the waste frying hydrogenated fats used in this study were kindly supplied by Renove—Recycling Company of Vegetable Oils (Salvador, Brazil). Any particulate matter present in the WFO was removed via filtration before analysis.

A group of 48 samples was prepared by combining different proportions of the WFO (SO, PO and VF). Each blend was prepared in triplicate using different WFO to vary the sample compositions. The experimental design is shown in Table 1.

### Apparatus and Data Pre-processing

A Perkin-Elmer Spectrum One spectrometer equipped with an attenuated total reflection (ATR) sampling accessory was employed to obtain spectra in the range of 3,000–600  $\text{cm}^{-1}$ . The ATR accessory was used in the absorbance mode and all spectra were recorded at  $20 \pm 1^\circ\text{C}$  with an average of 16 scans and a spectral resolution of 4  $\text{cm}^{-1}$ . Background spectra were obtained using a clean ATR cell accessory. After obtaining a spectrum, the cell was cleaned by successive treatments with hexane and acetone. One sample showed atypical absorption with a baseline offset; this spectrum was left out of all further chemometric analyses. All spectral data were exported to Unscrambler (version 9.1 from CAMO A/S). First derivative spectra were employed after smoothing by a Savitzky-Golay filter with a second order polynomial, using an 11-point window for PCA and a 21-point window

**Table 1** Composition of the samples prepared by mixing waste frying soybean oil, palm oil and hydrogenated vegetable fat

Sample	Hydrogenated vegetable fat (% v/v)	Soybean oil (% v/v)	Palm oil (% v/v)
1, 2, 3	100	0	0
4, 5, 6	0	100	0
7, 8, 9	0	0	100
10, 11, 12	33.3	33.3	33.3
13, 14, 15	80	10	10
16, 17, 18	10	80	10
19, 20, 21	10	10	80
22, 23, AS <sup>a</sup>	60	20	20
24, 25, 26	20	60	20
27, 28, 29	20	20	60
30, 31, 32	40	30	30
33, 34, 35	30	40	30
36, 37, 38	30	30	40
39, 40, 41	20	40	40
42, 43, 44	40	20	40
45, 46, 47	40	40	20

<sup>a</sup> Sample with atypical spectra

for PLS, to overcome the problem of systematic baseline drifts. First derivative spectra with Savitzky-Golay filter using a 21-point window were used in PLS since this window size provided lower RMSECV values. The variables were mean-centered before the modeling procedures.

### Chemometric Analysis

Principal component analysis is a tool in multivariate data analysis that can be used to simplify data by reducing the number of variables into a smaller number of orthogonal variables. The simplified variables are linear combinations of the original variables (wavenumber, in this case) and maximize the variability contained within them, thereby displaying most of the original variability in a smaller number of dimensions. A graphical representation of the pair-wise components allows the natural grouping of samples to be observed and indicates any similarity between the samples while also allowing different groups of samples to be identified [7, 12]. In the present study, PCA was used to identify the grouping tendencies between the WFO samples (SO, PO or VF) and blends containing the three waste cooking oils. Modeling power (MP) was used to indicate the important variables, i.e., the wavenumber range selection. The MP is a measure of the influence of a variable over a given model and can be used to study the relevance of a variable to indicate how much of the variable's variance was used to describe the model [13]. A preliminary PCA model was built using all variables of first derivative spectra. Then, the MP was obtained and the important variables were identified. Subsequently, another PCA model was built using relevant variables identified by MP and loading graph inspection.

Partial least-squares regression is a powerful chemometrics method for the analysis of mixtures to assess the degree of the relationship between a set of X-predictor variables and a set of Y-outcome variables by a linear multivariate model. The construction of a PLS model attempts to derive information from each source that is relevant to the relationship between the two blocks of variables [8]. In the present study, the concentrations of each oil in the WFO blends of the validation set were determined by PLS. All first derivatives spectral data was used for PLS calibration purposes. A model was constructed using a full cross-validation to define the number of retained latent variables, and an external set of samples was used to validate the model. The Unscrambler program was used for chemometrics calculations [12].

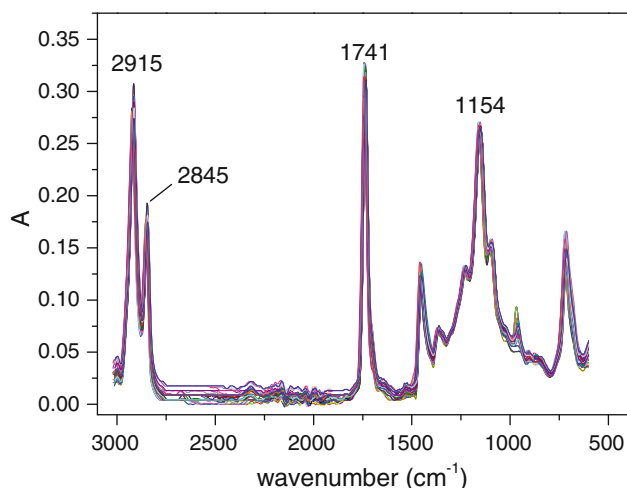
Thirty-four samples were used for the PLS calibration set. For the external validation set or the prediction samples, 10 samples were selected using the DUPLEX method [14] to provide representative sample and calibration sets. Three samples (24, 41 and 44) were detected at PLS cross-

validation as outliers by the anomalous values of residual Y-variance, residual X-variance and sample leverage. One sample showed atypical absorption with an offset baseline; this spectrum was excluded from all further chemometric analyses.

In the first step of the DUPLEX sample selection technique, the two points that are the furthest away from each other are selected for the calibration set. The two objects from the remaining points that are the furthest away from each other are included in the validation set. In the third step, the remaining point that is furthest away from the two points previously selected for the calibration set is included in that set. In the same way, a single point is selected for the validation set that is furthest from the existing points in that set. Following the same procedure, the points are alternately added to each set. These steps are repeated until the number of samples of the validation set, previously defined by the user, is reached. The remaining samples are included in the calibration set.

### Results and Discussion

MIR spectra of waste frying oils containing soybean oil, palm oil and hydrogenated vegetable fat as well as their blends are shown in Fig. 1. Differences between the spectra were not easily perceived via visual inspection; a multivariate statistical treatment was therefore, essential for pattern recognition of the blends in the effort to determine the oil content in each sample. Peaks with high-intensity absorption due to triacylglycerols, major components in vegetable oils, were detected at 2,915, 2,845, 1,741, and 1,154  $\text{cm}^{-1}$ . The peaks at 2,915 and 2,845  $\text{cm}^{-1}$  can be assigned to asymmetrical and symmetrical stretching



**Fig. 1** Raw spectra of waste frying oil from soybean oil, palm oil and hydrogenated fat and their blends

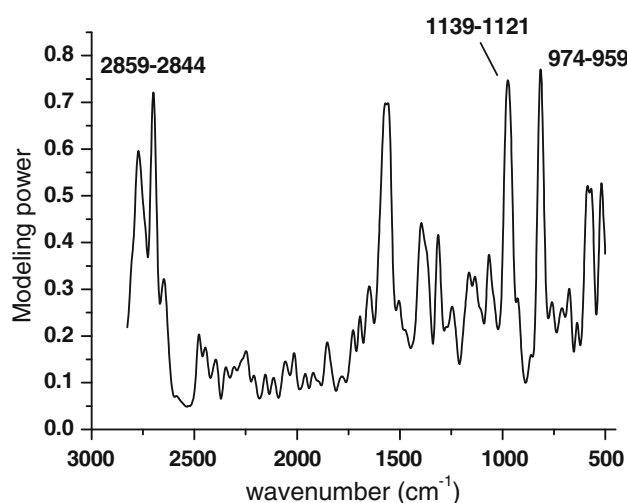
vibrations of C–H bonds, while the peak at  $1,741\text{ cm}^{-1}$  arose from C=O bond stretching. Stretching and bending of the C–O and  $-\text{CH}_2-$  accounted for the maximum at  $1,154\text{ cm}^{-1}$ . The region between 900 and  $1,300\text{ cm}^{-1}$  is known as the “fingerprint” region of vegetable oils, as previously described [10].

In this work, variables with modeling power averages of  $\geq 0.7$  were thought to be important for modeling. The closer the value was to 1, the more that variable was taken into account in the model, i.e., the wavenumbers with higher modeling power values were more significant with respect to the pattern recognition model. The modeling power result is shown in Fig. 2 and confirmed the importance of the three regions for building the PCA model: 2,859–2,844  $1,139$ – $1,121$  and  $974$ – $959\text{ cm}^{-1}$ . These regions correspond to 51 variables (spectral data). The important variables detected by modeling power criterion also showed higher loadings values (Fig. 3), which reinforce that these variables are really relevant for PCA model. A PCA model was calculated using the previous selected variable ranges of first derivative spectral data ( $2,859$ – $2,844$   $1,139$ – $1,121$  and  $974$ – $959\text{ cm}^{-1}$ ) and this model was used for further discussions about PCA results.

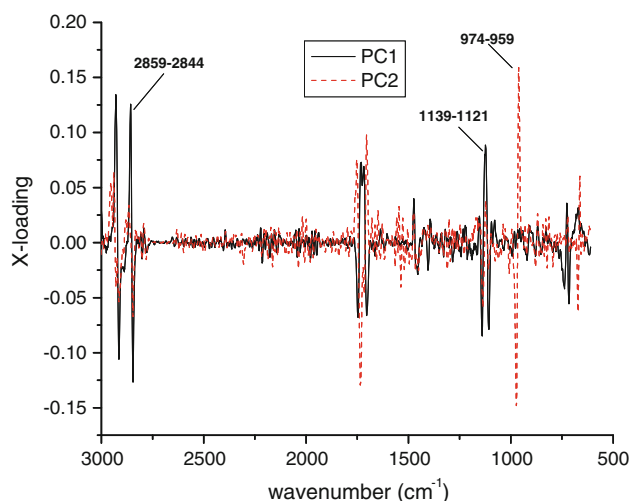
Figure 4 illustrates the PC1 versus the PC2 score and the loading plots, which are related to the sample and variable profiles, respectively. The first and second PCs accounted for 54.4 and 37.8%, respectively, of the variance in the data set. Observing the scores plot of the two components, PC1 versus PC2, samples from the pure soybean oil (SO), palm oil (PO) and hydrogenated vegetable fat (VF) were distinguished, which formed the vertices of a triangle. Mixtures containing primarily one type of oil were positioned next to their respective vertices. The samples with more homogeneous proportions of the oils were located at the center of the triangle.

The results illustrated by the PCA scores plot agreed with the ternary diagram, as was theoretically predicted for the three component mixtures. The pre-processed spectral data from the frying oil blends of unknown content of SO, PO or VF could be used to calculate their PC1 and PC2 score vectors. Comparing PC1 and PC2 score vectors of unknown samples with scores from known composition frying oil blends allows the identification of the most similar known sample blend. Since similar scores values mean similar compositions, by identifying a sample of known composition that is similar to an unknown sample, a first estimate of the unknown sample composition could be obtained.

Variables described by the loading graph that are located in the same quadrant of a sample group in the scores graph are the most important ones in describing a particular sample group. Thus, the loading graph indicates that the



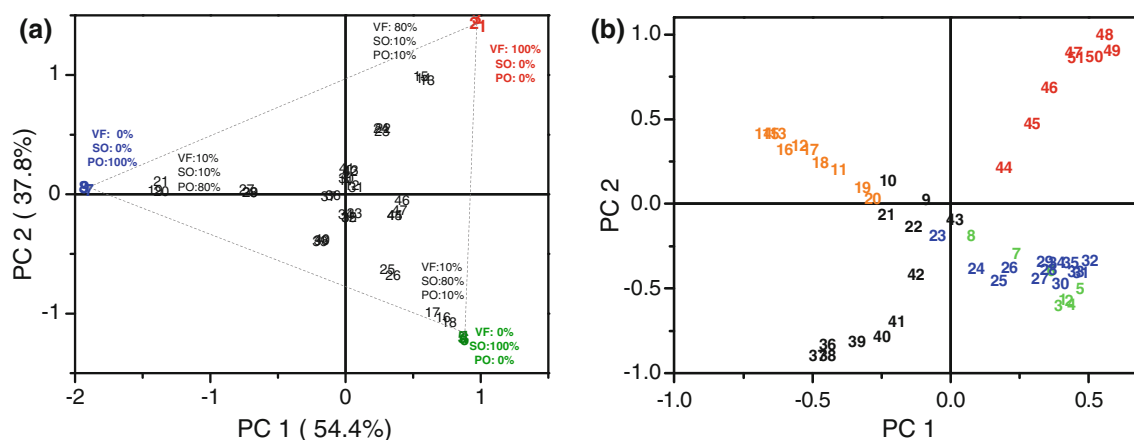
**Fig. 2** Modeling power of the MIR spectra ( $3,000$ – $600\text{ cm}^{-1}$ ) of the blends of soybean oil (SO), palm oil (PO) and hydrogenated vegetable fat (VF) in waste frying oils



**Fig. 3** Principal component analysis X-loading spectra: PC1 (solid line) and PC2 (broken line)

samples with high vegetable fat content were related to variables that ranged from 44–51 (wavenumbers  $966$ – $959\text{ cm}^{-1}$ ), while the samples with high soybean content were related to variables that ranged from 1–8 and 23–35 (wavenumbers  $2,859$ – $2,852$  and  $1,133$ – $1,121\text{ cm}^{-1}$ , respectively). Similarly, the samples with high palm oil content were related to variables that ranged from 11–16 and 17–20 (wavenumbers  $2,849$ – $2,844$  and  $1,139$ – $1,136\text{ cm}^{-1}$ , respectively). Remaining variables (9, 10, 21, 22, 36–43) were not related to a particular region of the loading graph, i.e., samples with high SO, VF or PO content ( $2,851$ ,  $2,850$ ,  $1,135$ ,  $1,134$  and  $979$ – $967\text{ cm}^{-1}$ , respectively).

To predict the oil content in WFO blends from the spectra data, multivariate calibration models were built by

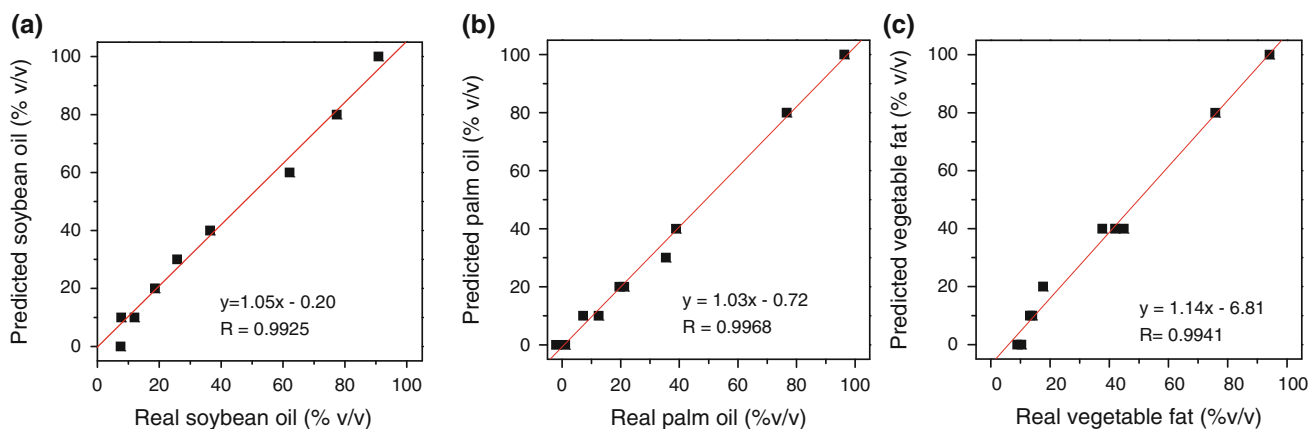


**Fig. 4** Principal component analysis: **a** scores, **b** loading graphs

**Table 2** Performance of the PLS models for determining the content of soybean oil (SO), palm oil (PO) and hydrogenated vegetable fat (VF) in waste frying oils

Type of oil or fat	No. LVs	RMSEP (% v/v)	RPD <sub>VAL</sub>	RMSECV (% v/v)	RPD <sub>CAL</sub>
VF	4	5.5	6.1	5.9	4.4
SO	3	4.7	7.4	4.6	5.6
PO	3	2.8	12.1	3.5	7.6

Number of retained latent variables (*no. LVs*), root mean square error for prediction (*RMSEP*), root mean square error of cross-validation (*RMSECV*), residual prediction deviation of validation (*RPD<sub>VAL</sub>*) and residual prediction deviation of calibration (*RPD<sub>CAL</sub>*)



**Fig. 5** Graphs of predicted versus real oil contents obtained by the PLS models for the validation set: **a** soybean oil, **b** palm oil, and **c** hydrogenated vegetable fat

PLS regression using different pre-processing strategies. PLS analysis was performed on the full spectral range. The best calibration results were obtained from the first derivative after smoothing by a Savitzky-Golay filter with a second order polynomial and a 21-point window. The root mean square error of prediction (RMSEP), the root mean square error of cross validation (RMSECV), the residual prediction deviation of calibration (RPD<sub>CAL</sub>) and validation (RPD<sub>VAL</sub>) [15], the number of retained latent variables

(*no. LVs*) and the percentage of explained variance are listed in Table 2.

The performance of the PLS model was evaluated using a validation set with ten samples. The results showed good agreement between the predicted and actual content of each oil in the WFO samples. As seen in Fig. 5, the slopes and intercepts of the regression lines are close to unity and zero, respectively, indicating low bias and low systematic regression errors.



The results obtained by PLS provided absolute prediction errors between 0.4% (v/v) and 10.3% (v/v). Considering the maximum absolute prediction errors and RMSEP values, the prediction quality of the models cannot be guaranteed for ranges below 5% (v/v) of frying oils, mainly for VF and SO. Despite the low number of calibration samples, the maximum number of latent variables retained was four; the number of samples in the calibration set was high enough to agree with the recommended minimum number of calibration samples [16].

RPD was used to measure precision and evaluate how well the calibration model predicts the validation set. The higher the RPD the more precisely are the data described by the calibration. According to Igne and Hurburgh Jr (2009), a model with an RPD value larger than 5 is considered good for quality control; with an RPD value larger than 6.5, the model can be used for process monitoring; and models with RPD higher than 8 can be used in any applications [15]. In this way, all calibration models obtained in this work can be used for quality control when considering  $RPD_{VAL}$  values.

## Conclusions

Mid-infrared spectroscopy in association with PCA was successfully used to evaluate the patterns in various blends of waste frying oils consisting of soybean oil, palm oil and hydrogenated vegetable fat. Through this method, careful examination of the PCA results can be employed for a first estimate of the composition of oil blends. Additionally, partial least squares regression based on mid-infrared spectra was proven suitable to predict the content of each type of oil present in mixtures of waste frying oils. This methodology can be very useful for the rapid, reagent-free, low cost determination of the composition of waste cooking oil and can aid in decision making regarding the pretreatment of oils and the production routes for biodiesel.

**Acknowledgments** The authors would like to acknowledge the financial support received from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

## References

1. da Silva ML, Figueiredo AP, Cardoso AL, Natalino R, da Silva MJ (2011) Effect of water on the ethanolsis of waste cooking

- soybean oil using a tin(II) chloride catalyst. *J Am Oil Chem Soc* 88:1431–1437. doi:10.1007/s11746-011-1794-z
2. Kucek KT, César-Oliveira MAF, Wilhelm HM, Ramos LP (2007) Ethanolsis of refined soybean oil assisted by sodium and potassium hydroxides. *J Am Oil Chem Soc* 84:385–392. doi:10.1007/s11746-007-1048-2
3. Shu Q, Gao J, Nawaz Z, Liao Y, Wang D, Wang J (2010) Synthesis of biodiesel from waste vegetable oil with large amounts of free fatty acids using a carbon-based solid acid catalyst. *Appl Energ* 87:2589–2596. doi:10.1016/j.apenergy.2010.03.024
4. Tunc MF, Gurbuz H, Türkay SZ (2010) Non-evaporative solvent recovery step in deacidification of used frying oil as biodiesel feedstock by methanol extraction. *J Am Oil Chem Soc* 87:195–203. doi:10.1007/s11746-009-1488-y
5. Fröhlich A, Rice B (2009) Sources of methyl ester yield reduction in methanolysis of recycled vegetable oil. *J Am Oil Chem Soc* 86:269–275. doi:10.1007/s11746-008-1344-5
6. Lam MK, Lee KT, Mohamed AR (2010) Homogeneous, heterogeneous and enzymatic catalysis for transesterification of high free fatty acid oil (waste cooking oil) to biodiesel: a review. *Biotechnol Adv* 28:500–518. doi:10.1016/j.biotechadv.2010.03.002
7. Pimentel MF, Ribeiro GMGS, da Cruz RS, Stragevitch L, Pacheco Filho JGA, Teixeira LSG (2006) Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. *Microchem J* 82:201–206. doi:10.1016/j.microc.2006.01.019
8. Teixeira LSG, Oliveira FS, dos Santos HC, Cordeiro PWL, Almeida SQ (2008) Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline. *Fuel* 87:346–352. doi:10.1016/j.fuel.2007.05.016
9. Lerma-García MJ, Simó-Alfonso EF, Bendini A, Cerretani L (2011) Rapid evaluation of oxidised fatty acid concentration in virgin olive oil using Fourier-transform infrared spectroscopy and multiple linear regression. *Food Chem* 124:679–684. doi:10.1016/j.foodchem.2010.06.054
10. Hennessy S, Downey G, O'Donnell CP (2009) Confirmation of food origin claims by Fourier transform infrared spectroscopy and chemometrics: extra virgin olive oil from Liguria. *J Agric Food Chem* 57:1735–1741. doi:10.1021/jf803714g
11. Rohmana A, Mana YBC (2011) Application of Fourier transform infrared (FT-IR) spectroscopy combined with chemometrics for authentication of cod-liver oil. *Vib Spectrosc* 55:141–145. doi:10.1016/j.vibspec.2010.10.001
12. Esbensen KH (2002) Multivariate data analysis—in practice: an introduction to multivariate data analysis and experimental design, camo process AS, Norway
13. Liu F, He Y, Wang L (2008) Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Anal Chim Acta* 615:10–17. doi:10.1016/j.aca.2008.03.030
14. Snee RD (1977) Validation of regression models: methods and examples. *Technometrics* 33:75–100
15. Igne B, Hurburgh CR Jr (2010) Local chemometrics for samples and variables: optimizing calibration and standardization processes. *J Chemometrics* 24:75–86. doi:10.1002/cem.1274
16. American Society for Testing, Materials (2005) Standards practices for infrared multivariate quantitative analysis, ASTM E1655-05. DOI 10.1520/E1655-05