Talanta 88 (2012) 537–543

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

An automated approach for analysis of Fourier Transform Infrared (FTIR) spectra of edible oils

Siong Fong Sim*, Woei Ting

Universiti Malaysia Sarawak, Faculty of Resource Science and Technology, 94300 Kota Samarahan, Sarawak, Malaysia

ARTICLE INFO

Article history:
Received 9 September 2011
Received in revised form
10 November 2011
Accepted 10 November 2011
Available online 17 November 2011

Keywords:
Fourier Transform Infrared
Peak detection
Peak assignment
Peak matching
Edible oils

ABSTRACT

This paper reports a computational approach for analysis of FTIR spectra where peaks are detected, assigned and matched across samples to produce a peak table with rows corresponding to samples and columns to variables. The algorithm is applied on a dataset of 103 spectra of a broad range of edible oils for exploratory analysis and variable selection using Self Organising Maps (SOMs) and *t*-statistics, respectively. Analysis on the resultant peak table allows the underlying patterns and the discriminatory variables to be revealed. The algorithm is user-friendly; it involves a minimal number of tunable parameters and would be useful for analysis of a large and complicated FTIR dataset.

1. Introduction

Fourier Transform Infrared (FTIR) is a form of vibrational spectroscopy where infrared radiation is passed through a sample resulting in absorption of the radiation that stimulates vibrational motions. Fundamentally, a molecule is represented by a set of flexible and moving atoms where the atoms constantly oscillate around average positions. When the vibration of the atoms produces an oscillating electric field same as the frequency of the IR radiation, it gives rise to peaks (often referred to as bands) in vibrational spectrum. Each spectral peak is characterised by its frequencies and amplitude as a molecule only absorbs at frequencies corresponding to its molecular modes of vibration in the infrared region [1].

FTIR has been widely used as a routine diagnostic tool for qualitative and quantitative analyses. The major advantages of FTIR over other analytical methods are it is relatively fast, non-destructive and cost effective; in addition, only a small amount of sample is required [2]. FTIR in combination with various multivariate data analysis methods have allowed rapid evaluation of a large volume of spectral data for quality control [3], quantification [4] and pattern recognition [5–7]. Typically, this involves the profile of the entire spectrum or specific spectral regions containing relevant information [8]. In this paper, we report an automated computational approach for analysis of FTIR spectra and demonstrate the algorithm for the analysis of edible oils. Fundamentally, the algorithm finds peaks present in each sample and matches the peaks across samples to produce a peak table with rows corresponding to samples and columns to variables for multivariate data analyses. The algorithm is also designed to assign peaks according to frequencies suggesting the possible functional groups. This would speed up the peak assignment process which is often performed manually and serves as a guide for non-expert users in spectra interpretation. In fact, many automated data processing algorithms have been developed for various applications, i.e., GC–MS [9], GC–DMS [10], GC–GC [11], LC–MS [12], NMR [13] and HPLC [14]. They have been used for microbial, metabolomics and food study [15–20]. The method is particularly well-accepted when the data processing software available with the workstation can no longer cater the demand of the analyst, i.e., he/she wishes to look at hundreds of peaks in a complex matrix and compare the peaks across multiple samples. Hence, the algorithm offers an alternative to analyse large high dimensional FTIR datasets when interpreting complicated superimposed spectra becomes infeasible and peak integration becomes tedious and time consuming.

2. Materials and methods

2.1. Dataset

This dataset consists of 103 FTIR spectra of 8 types and 16 brands of edible oils. The samples represent a broad range of edible oils from olein palm oil, blended palm oil (consists of olein palm oil,

* Corresponding author. Tel.: +60 82 582995; fax: +60 82 583160.
E-mail addresses: sfsim@frst.unimas.my, siongfong@gmail.com (S.F. Sim).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.11.030

Table 1
The number of samples according to types and brands.

Type	Brand	No. of samples
Olein palm oil	A	15
	B	3
	C	3
	D	5
	E	5
	F	6
Blended cooking oil (olein palm oil, peanut and grains)	G	15
	H	3
	I	3
Sesame oil	J	6
Sunflower oil	K	6
	L	6
Sunflower and canola oil	M	9
Corn oil	N	6
Coconut oil	O	6
Canola oil	P	6
TOTAL		103

peanut oil and grain oil), sesame oil, sunflower oil, blended sunflower oil (consists of sunflower and canola oil), corn oil, coconut oil and canola oil. The number of samples according to types and brands are summarised in Table 1. The brands are labelled as *A* to *P*; these products are commonly available in the local retail market except brand *O* (*O* is the homemade crude coconut oil). The samples were obtained from different households and were analysed randomly using ATR-FTIR.

2.2. Instrumentation

All spectra were obtained using a ThermoScientific FTIR spectrometer equipped with a removable ZnSe crystal controlled by OMNIC software (Thermo Nicolet Analytical Instruments, Madison, WI). An ATR-accessory was used with all spectra collected by co-addition of 32 scans at a resolution of 4 cm^{-1} in the range of 4000–650 cm^{-1}. The spectrum of each sample was ratioed against a fresh background spectrum recorded from the bare ATR crystal. Prior to collection of each background spectrum, the ATR crystal was cleaned with acetone to remove any residual.

2.3. Data analysis

2.3.1. Peak detection

The spectra in CSV format were converted to Matlab version 7.8 (The Mathworks, Inc., Natick, MA), each spectrum was represented as a vector of dimensions (6949×1) with a scanning rate of 0.482 cm^{-1}. The spectrum was baseline corrected using asymmetric least squares [21] prior to peak detection. Note that the algorithm is developed in-house in Matlab. The baseline corrected spectrum was first de-noised using soft heuristic thresholding and scaled noise option at level 4 by *sym8* wavelet [22]. The first derivative of the smoothed spectrum and the average absolute change of derivative over the frequencies, *h*, were then computed.

Based on the first derivative of the signal, the peak start, peak maxima and peak end are located according to the criteria described elsewhere [9–11]. Briefly, a peak start is recognised when the value of the first derivative is greater than 0 and is *f* times (where *f* refers to the noise factor) more than the calculated *h*. In this paper, the peak noise factor, *f*, is set at 3. The centre of the peak is identified at the zero-crossing point in the derivative where the signal crosses the *x*-axis going from positive to negative and the peak end is determined when the derivative of two subsequent points are above zero. Peaks that are too small, with a window of less than 20 scans (9.64 cm^{-1}), are rejected. This parameter is referred to as the peak filtering window, *w*. The peak area is calculated as the sum of all wavenumbers contributing to the peak. These are the typical criteria of peak detection; they are generally employed elsewhere yielding a promising efficiency (often with a correlation coefficient greater than 0.90) [9,10,14] therefore, in this paper, the efficiency of the peak detection algorithm will not be re-examined. The algorithm is also designed to allow characterisation of a peak whether it is a strong, medium or a weak peak according to the ratio of the intensity of the peak to the strongest peak. If the ratio is greater than 0.65, it is labelled as a strong peak; if the ratio falls between 0.35 and 0.65, it is a medium peak, otherwise a weak peak. Note that the value of the ratio ranges between 0 and 1. The algorithm reports the peaks detected in every individual sample including the positions (peak start, peak maxima, peak end), the peak area and peak characteristics.

2.3.2. Peak assignment

Peak assignment process can become tedious and time consuming when one is involved with a large number of samples consisting of many peaks. Fundamentally, a number of functional groups could possibly be associated to a peak found within a distinguished frequency range; Williams and Fleming [23] spent almost the entire chapter of the infrared method in their book describing the characteristic absorption frequencies of organic functional groups. It may not be too challenging for an expert user to identify the relevant functional groups if the number of samples is manageable. For a non-expert user, this can be difficult so the automated peak assignment algorithm can be used to improve the spectral interpretation process.

A database consisting of 168 peaks is created according to the group frequencies described in Williams and Fleming [23]; every peak is characterised by its frequency range, peak characteristic, i.e., strong, medium, weak and its corresponding functional group. For every peak detected, the algorithm would search through the library for the possible functional groups according to the frequency. If a peak is found within the frequency range of a documented functional group, the peak is assigned to the group. Peaks that cannot be assigned are denoted undefined. The algorithm yields a peak assignment table with possible functional groups designated to each peak. Note that a peak can be assigned to more than one functional group as the frequency range of several groups might have overlapped; for example, a peak at 3003 cm^{-1} has been assigned to three groups (=CH stretching, CH stretching of $CO–CH_3$ and intramolecular H bonded –OH), an analyst is required to interpret the peak according to prior knowledge on the sample and peak characteristics. In the peak assignment process, the peak characteristic is not taken into consideration although it can be an important criterion. This is because the peak characteristic often depends on the quality of the spectra for example, a poor sampling technique may result in changes in the entire absorption band distribution [24]. This could jeopardize the consistency of the peak assignment algorithm.

2.3.3. Peak matching

The peak matching algorithm is designed to match peaks across multiple samples producing a peak table $(N \times M)$ with rows corresponding to samples and columns to variables. Peaks detected are compared across samples; when a peak is examined, it is referred to as a target peak. The potential matching peaks are identified based on the peak matching window, defined as $\pm z$ scan number of the target peak where *z* is the tolerance. If a target peak is found at *b* scan number, the peak matching window ranges between $(b+z)$ and $(b-z)$. If no peaks are found matching the target peak, the target peak is identified as a unique peak. If more than one candidate matching peak is found in a sample, the candidate with its frequency closest to the target peak is chosen as the matching peak. Generally, the efficiency of the peak matching algorithm

is governed by the peak matching window. If the window is too large, there is a greater chance of mismatching; if the window is too small, peaks are failed to be matched. In this paper, the peak matching window is set at ±8 scans (±3.856 cm^{-1}) yielding a peak table of dimensions (103 × 46). The parameters are often set according to user's experience. In this case, we evaluate the peak tables obtained using various settings of peak noise threshold and peak filtering window, the parameters are chosen when all observable peaks are detected with a minimal number of small peaks (possibly noise). For the peak matching window, we select the parameter that matches the peaks across samples appropriately. For example, a peak table indicates the presence of two variables at 2852 and 2853 cm^{-1}; these two peaks are possibly corresponding to the same functional group but have been identified as two different compounds suggesting inappropriate matching window size. For FTIR, the suitable peak matching window is very much depending on the application of the users. Sometimes, a small shift may indicate important information; for example in the study of cooking oil, an absorbance shift from 3003 to 3008 cm^{-1} implies an increase in oleic acid content. Therefore to find an appropriate window size, one is suggested to evaluate the peak table generated using different windows and determine the efficiency of the peak table, i.e., minimal mismatching and suspicious unmatched unique peaks which will be revealed on the peak assignment table.

2.4. Multivariate analysis

2.4.1. Preprocessing

The peak table was square rooted and standardised prior to multivariate data analyses. Square-rooting aims at reducing the influence of outlier measurements and standardisation ensures all peaks are weighed equally and on the same scale [25].

2.4.2. Self Organising Maps

Self Organising Maps (SOMs) is used to explore the underlying patterns of the dataset. The algorithm of SOM is described elsewhere [25–27]. Briefly, a map with predefined dimensions ($R \times S$) is initialised; the map consists of a total of U units ($R \times S = U$), each characterised by a ($1 \times M$) weight vector where the weight vector of each variable m is chosen randomly from a uniform distribution within the observed range of variable m. A sample is chosen randomly where the input pattern is compared to the weight vector of each map unit based on Euclidean distance. The most similar map unit is regarded as the best matching unit; the algorithm will update the weight vector of the best matching unit and its neighbouring units. This process is repeated for T iterations [25]. In this paper, we use a rectangular map with 300 hexagonal units of dimensions ($15 \times 20 = R \times S$); they are trained for 5000 iterations with a learning rate of 0.1. The interpretation of a SOM is similar to the score plot of Principal Component Analysis (PCA); similar samples are mapped close to each other and dissimilar apart.

2.4.3. t-Statistic

t-Statistic is employed to identify the discriminatory variables. It is a variable selection method used for comparison involving two classes; t-statistic evaluates whether the means for two classes are statistically different. The t-value for each variable, m is calculated according to the following equation [25].

$$t_m = \frac{\bar{x}_{mA} - \bar{x}_{mB}}{S_m \sqrt{(1/n_A) + (1/n_B)}}$$

$\bar{x}_{mA}$ and $\bar{x}_{mB}$ are the mean of each variable calculated for class A and B. S_m is the pooled standard deviation for each variable over the two classes. n_A, n_B are the number of samples in class A and B.

In this paper, t-statistic is performed on one-versus-all comparisons; if there are eight classes, the comparison is between class 1 and the rest, class 2 and the rest, class 3 and the rest, etc. for all classes. For each comparison, the variables are ranked according to the absolute t values; the variable with the highest t value is the most significant variable. The top 10 significant variables of each class are selected where the overall t-values of the variables are recalculated as the average of the t-values obtained over one-versus-all comparisons. For example, a variable is selected as the top 10 significant variables in 4 classes with t-values 2.8, 3.6, 5.5 and 4.2, respectively; the overall t-value is therefore 4.025 ((2.8 + 3.6 + 5.5 + 4.2)/4). The selected variables are then ranked according to the overall t-values. The rationale of this approach is to numerically reflect the contributions of the highly ranked variables over all comparisons.

3. Results and discussion

3.1. FTIR spectra

The FTIR spectra of eight different types of edible oils are plotted in Fig. 1; apparently, they appear very similar and one cannot easily interpret the differences between them. Generally, there are several large and consistent peaks at 2922, 2852, 1742, 1461, 1375, 1160 and 721 cm^{-1}. The spectra were subjected to the peak detection and matching algorithm ($z = \pm 8$ scans) yielding a peak table of dimensions (103 × 46). Fig. 2 shows the unique and matching peaks identified across samples in three spectral regions (3000–2750 cm^{-1}, 1800–1590 cm^{-1} and 1500–650 cm^{-1}). The resultant peak table was subjected to the peak assignment algorithm where each variable is designated to the possible functional groups. Table 2 summarises the major peaks found in the spectra of edible oils and the corresponding functional groups.

3.2. Self Organising Maps

Similar features are observed in the spectra of various types of edible oils; it is hard to deduce the differences between them manually. The peak table generated from the algorithm can be used for various multivariate data analyses. In this paper, we employ SOMs, a non-linear and unsupervised approach, to interpret the underlying patterns of the data. Fig. 3(a) illustrates the trained SOM of edible oils according to types where letters '*A* to *P*'

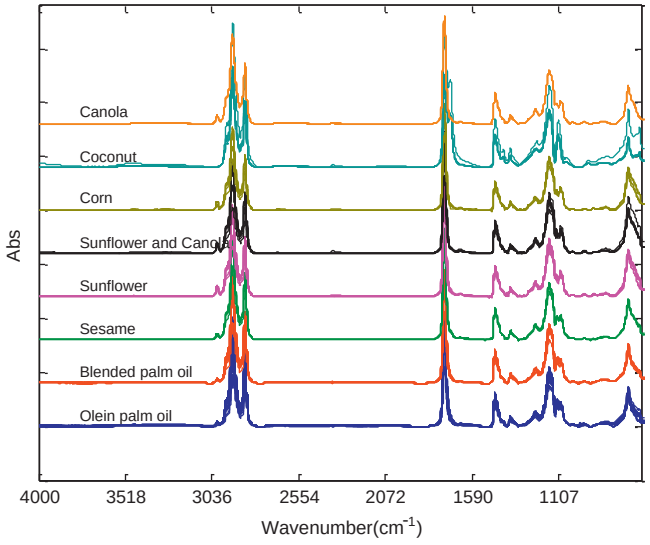

Fig. 1. The spectra of 103 samples of edible oils according to types.

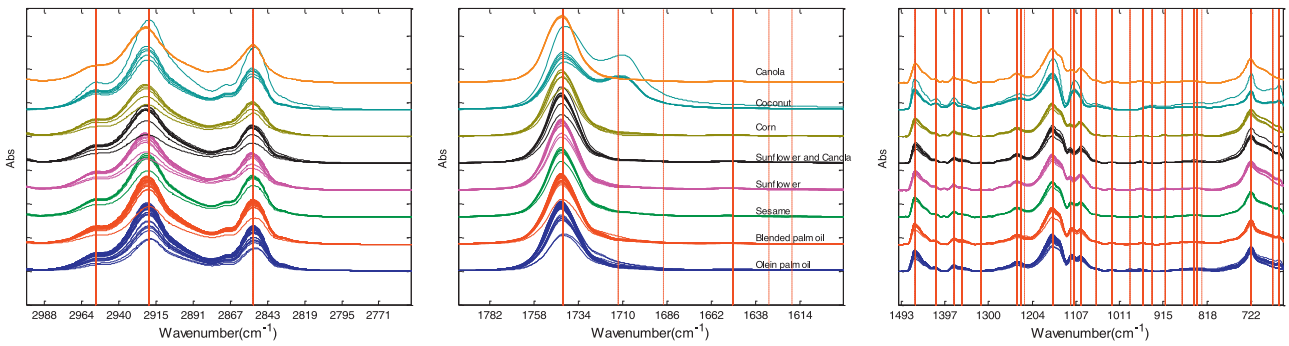

Fig. 2. Peaks identified in three specific regions (3000–2750 cm^{-1}, 1800–1590 cm^{-1} and 1500–650 cm^{-1}). The dashing lines are the unique and matching found using the automated approach.

Table 2
The major peaks found in the spectra of edible oils and the corresponding functional groups.

Wavenumber (cm^{-1})	Peak assignment	Intensity
3003	=CH stretching (sym)	Weak
2955, 2922	CH stretching (asym)	Strong
2852	CH stretching (sym)	Strong
1742	Ester C=O stretching of the triglycerides	Medium
1712	C=O, free fatty acid shoulder	Weak
1650	C=C stretching, *cis* RHC=CHR	Weak
1461	CH deformation (CH_2, CH_3)	Weak
1415	Rocking of =C–H bending, *cis* RHC=CHR	Weak
1158	Bending of CH_2 groups	Medium
1238, 1377	Bending of CH_2 groups	Weak
1027, 1062, 1095, 1110,1118	C–O stretching	Weak
958	C=C bending, *trans* RHC=CHR	Weak
721	C–H out-of-plane deformation	Medium
658, 912	C–H out-of-plane deformation	Weak

indicate the brands (refer to Table 1). Noticeably, blended palm oil consisting partly of palm oil shares the same cluster as the pure olein palm oil. The sunflower oil, corn oil, blended sunflower oil and canola oil on the other hand are mapped closely on the map; this suggests similar characteristics between them. The sesame oil and crude coconut oil are independently distinguishable. To verify the peak detection and matching algorithm has retrieved the correct information from the spectrum profile, we compare the trained SOM generated using the peak table (Fig. 3(a)) with the trained SOM produced based on the entire profile (Fig. 3(b)). Note that the matrix involving the entire spectrum profile is reasonably large (103×6949 = No. of samples $\times$ wavenumbers) therefore the training of SOM is computationally more intensive. The clustering patterns of SOMs produced using both approaches are comparable indicating the reliability of the algorithm.

3.3. *Variable selection*

The discriminatory variables responsible for differences between spectrum profiles can be distinctively suggested by forming a peak table. In this paper, *t*-statistic is used to suggest the discriminatory variables; Fig. 4 illustrates the distribution of the discriminatory variables and the corresponding spectral regions (the discriminatory variables are marked with dashed lines). Generally, they can be grouped to represent three main functional groups, C=O, C–O stretching and double bonds. The peak at 1712 cm^{-1}, an indicative of the presence of free fatty acids, is distinctively observed in the crude coconut oil. This is likely because the coconut oil is unrefined (homemade) and it is common that unrefined oil contains some free fatty acids; the levels of free fatty acid will be reduced in refining [28]. Peaks corresponding to

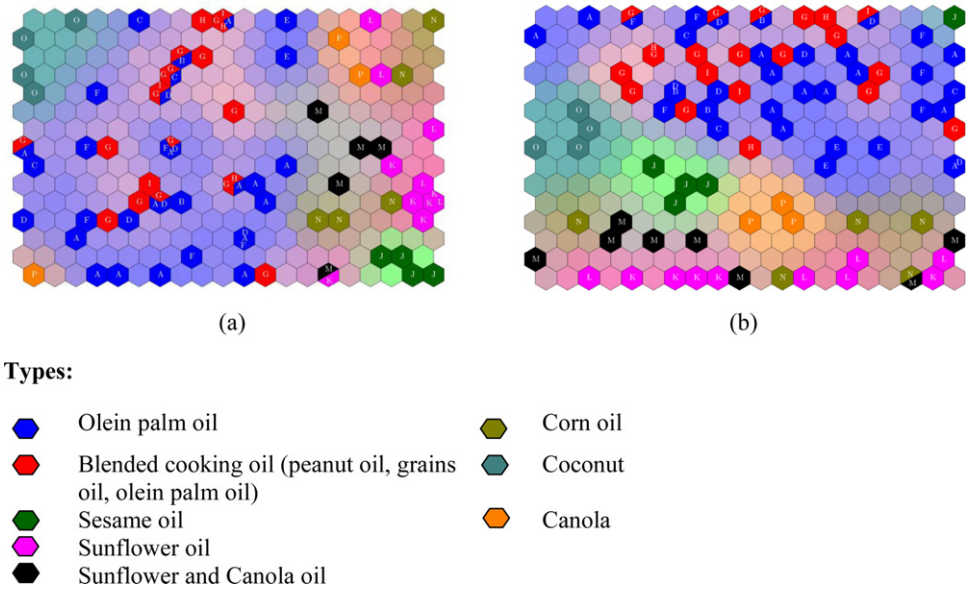

Fig. 3. (a) The trained SOM of edible oils using the peak table; and (b) the trained SOM of edible oils using the entire profile.

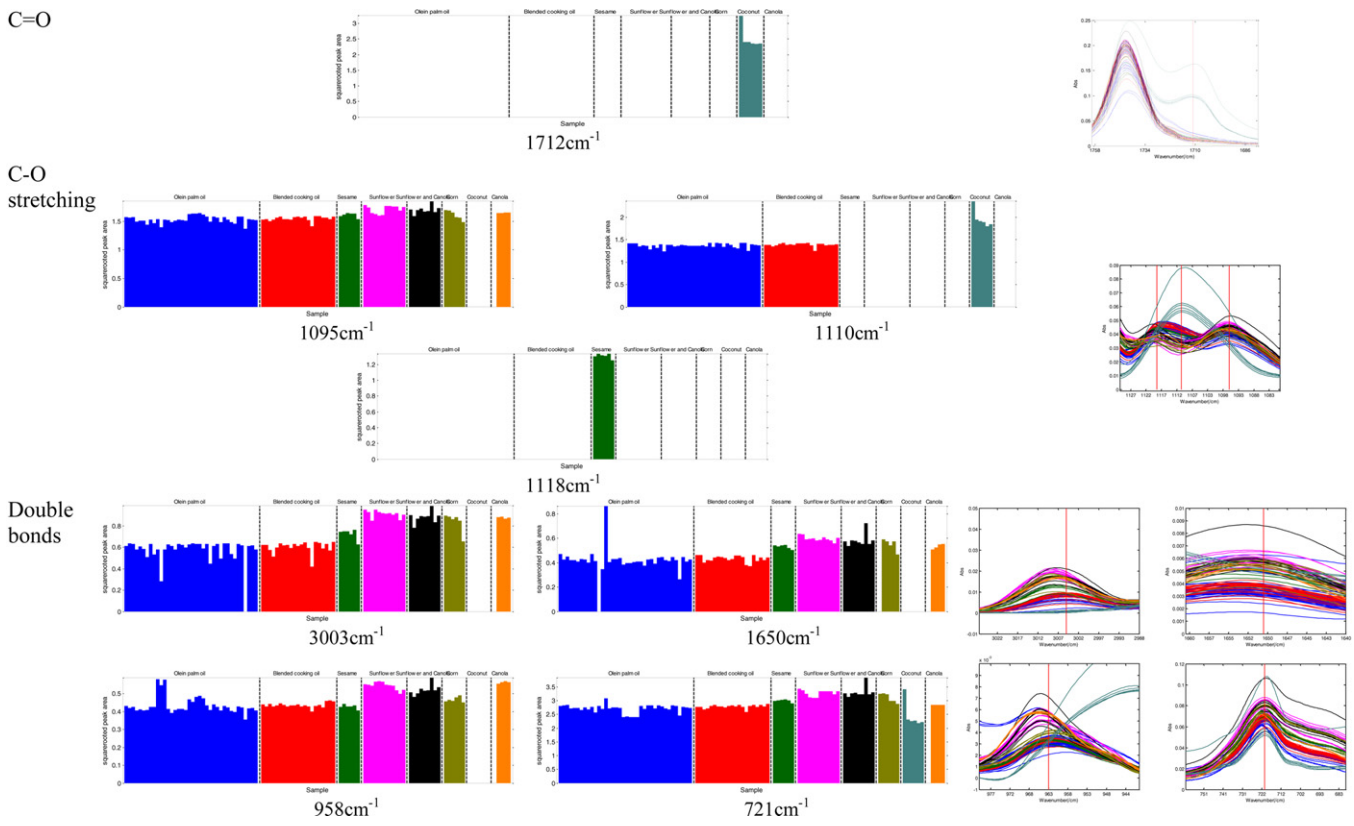

Fig. 4. The distributions of the discriminatory variables and the corresponding spectral regions.

C–O stretching of ester groups at 1095, 1110 and 1118 cm^{-1} are invariably found across samples. This vibration consists of two asymmetric coupled vibrations C–C(=O)–O and O–C–C with the former being more important [29,30]. As observed, the peak at 1095 cm^{-1} is detected in almost all samples except the coconut oil. The peak at 1110 cm^{-1} however is predominantly found in palm oil and coconut oil (saturated oil), this peak is shifted to higher wavenumber at 1118 cm^{-1} in sesame oil. Peaks corresponding to unsaturation at 3003, 1650, 958 and 721 cm^{-1} also exhibit noticeable variations across samples. They have been assigned to =C–H stretching of *cis* RHC=CHR and methylene rocking vibration of straight chain paraffins with the out of plane vibration of *cis*-disubstituted olefins. The observed variation (as shown in Fig. 4) reflects the fatty acid compositions of various edible oils. Typically, palm oil contains higher amount of saturated fatty acids, pure sunflower oil contains higher amount of polyunsaturated fatty acids whilst canola oil is richer in monounsaturated fatty acids [8]. This observation is corroborated by the fatty acid compositions suggested elsewhere; Table 3 summarises the fatty acid compositions of some common edible oils reported in Ref. [31]. The unsaturation percentage decreases in the following order: sunflower oil, corn oil, sesame oil, canola oil, palm oil and coconut oil; this explains the observed absorbance values of peaks at 3003, 1650, 958 and 721 cm^{-1}. Sunflower oil with a larger proportion of polyunsaturated groups has higher absorbance values than those with a lower amount. A weak absorption band at 958 cm^{-1} is found in almost all edible oils except the coconut oil. This peak is attributed to the presence of trans-*olefins* where polyunsaturated and monounsaturated oil, i.e., sunflower and canola oil indicate higher intensities; the *trans*-olefins may be produced during the heating of oils in the purification process [32].

The spectral interpretation of FTIR can be very sensitive to peak shifting; peaks occurring at very close wavenumbers may provide important information for example, a peak at 3009 cm^{-1} was found in the spectra of vegetable oils but for extra virgin olive oil, the peak was shifted to 3005 cm^{-1}. This indicates extra virgin olive oil contains higher proportion of oleic acyl groups whilst vegetable oils are richer in linoleic acyl groups [33,34]. To illustrate the ability of the algorithm in retrieving information relating to peak shifting, we reduce the peak matching window, z, to ± 5 scans (2.41 cm^{-1}). The resultant peak table was subjected to SOMs; no significant differences were observed between SOMs produced using the peak matching parameter of $z = \pm 5$ (SOM not shown) and $z = \pm 8$. With a lower peak matching parameter, two different peaks with fairly

Table 3
The typical fatty acid compositions of some common edible oils (reported in [31]).

	Saturated fatty acids (SFA)	Monounsaturated fatty acids (MUFA)	Polyunsaturated fatty acids (PUFA)
Olein palm oil	45.3	41.6	8.3
Sesame	14.0	42.0	45.0
Sunflower	11.9	20.2	63.0
Corn	12.7	24.7	57.8
Coconut	85.2	6.6	1.7
Canola	5.3	54.3	24.8

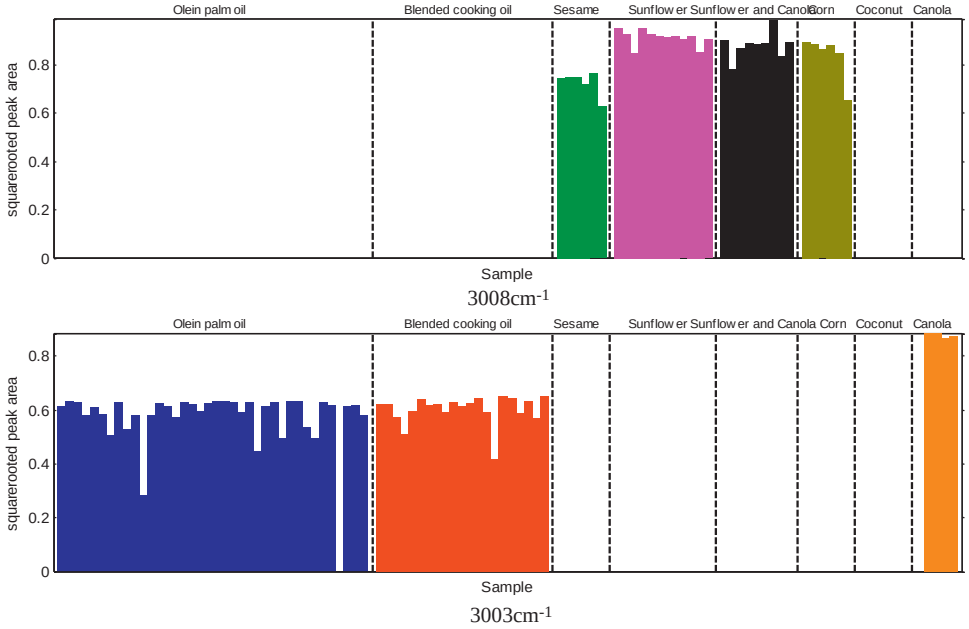

Fig. 5. The peak shift encountered at ~3003 cm^{-1}.

close wavenumbers at 3008 and 3003 cm^{-1} are identified (Fig. 5). Sunflower, blended sunflower, corn oil and sesame oil show a maximum absorbance at 3008 cm^{-1}; the absorbance is shifted to 3003 cm^{-1} for palm oil, blended palm oil and canola oil. This suggests that sunflower oil, corn oil and sesame oil contain a lower percentage of oleic acid followed by canola oil and palm oil [33].

4. Conclusion

The algorithm is implemented for the analysis of edible oils; the information extracted agrees well with the observations published elsewhere [6,8,32–36]. In addition, when the peak detection method is compared to the full spectrum approach using SOMs, similar inferences are drawn. This automated algorithm is user-friendly; it involves a minimal number of tunable parameters (peak noise factor, peak filtering window and peak matching window) therefore it can be easily dealt with by non-expert users. Expert users who wish to mine the information relating to peak shifting can carefully reduce the size of peak matching window. On the other hand, if one is looking at contaminants which often appear as small peaks, the peak noise factor and peak filtering window can be set to a lower value. The algorithm extracts information by forming a peak table; this allows specific peaks to be identified. It is an advantageous feature of peak detection methods over the entire profile.

The computational approach for analysis instrumental output is not a new subject; it has been employed in other applications however it is largely unexplored in FTIR. FTIR is known to be a rapid technique; we could easily collect spectra of hundreds of samples rendering manual spectral interpretation infeasible. The automated algorithm would close the gap between the rate at which spectra data is generated and the rate it can be understood. This algorithm would complement the advantage of FTIR as a rapid technique for routine chemical analysis especially when a large dataset is involved. Nevertheless, it is important to note that the automated peak detection method is not always perfect as it is parameter dependent. For example, if the peak noise threshold is set too high, small peaks could possibly be missed; if it is set too low, undesirable peaks attributable to noise will be identified as signals.

Acknowledgements

The authors would like to thank the colleagues from Universiti Malaysia Sarawak and KTA (Sarawak) Sdn Bhd for providing the cooking oil samples.

References

[1] C. Sacksteder, B.A. Barry, J. Phycol. 37 (2001) 197–199.
[2] R. Davis, L.J. Mauer, Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology, Formatex, Badajoz, 2010.
[3] D.W. Lachenmeier, Food Chem. 101 (2007) 825–832.
[4] I.F. Duarte, A. Barros, I. Delgadillo, C. Almeida, A.M. Gil, J. Agric. Food Chem. 50 (2002) 3104–3111.
[5] E. Nioroge, S.R. Alty, M.R. Gani, M. Alkatib, Proc. IEEE Eng. Med. Biol. Soc. 1 (2006) 5338–5341.
[6] D.A. Rusak, L.M. Brown, S.D. Martin, J. Chem. Educ. 80 (2003) 541.
[7] L. Mariey, J.P. Signolle, C. Amiel, J. Travert, Vib. Spectrosc. 26 (2001) 151–159.
[8] S.T.H. Sherazi, M.Y. Talpur, S.A. Mahesar, A.A. Kandhro, S. Arain, Talanta 80 (2009) 600–606.
[9] S.J. Dixon, R.G. Brereton, H.A. Soini, M.V. Novotny, D.J. Penn, J. Chemometr. 20 (2007) 325–340.
[10] S.F. Sim, P. Rearden, C. Kanchagar, C. Sassetti, J. Trevejo, R.G. Brereton, Anal. Chem. 83 (2011) 1537–1546.
[11] S. Peter, G. Vivó-Truyols, P.J. Marriott, P.J. Schoenmakers, J. Chromatogr. A 1156 (2007) 14–24.
[12] J.Q. Zhang, W. Haskins, Genomics 11 (Suppl. 3) (2010) S8.
[13] Q.N. Van, A.J. Shaka, J. Magn. Reson. 132 (1998) 154–158.
[14] M.J. Fredricksson, P. Petersson, B.O. Axelsson, D. Bylund, J. Sep. Sci. 32 (2009) 3906–3918.
[15] S.F. Sim, V. Sági-Kiss, R.G. Brereton, Talanta 83 (2011) 1269–1278.
[16] K. Wongravee, N. Heinrich, M. Holmboe, M.L. Schaefer, R.R. Reed, J. Trevejo, R.G. Brereton, Anal. Chem. 81 (2009) 5204–5217.
[17] S.J. Dixon, N. Heinrich, M. Holmboe, M.L. Schaefer, R.R. Reed, J. Trevejo, R.G. Brereton, Chemometr. Intell. Lab. Syst. 99 (2009) 111–120.
[18] K. Wongravee, G.R. Lloyd, J. Hall, M.E. Holmboe, M.L. Schaefer, R.R. Reed, J. Trevejo, R.G. Brereton, Metabolomics 5 (2009) 387–406.
[19] S. Zomer, S.J. Dixon, Y. Xu, S.P. Jensen, H. Wang, C.V. Lanyon, A.G. O'Donnell, A.S. Clare, L.M. Gosling, D.J. Penn, R.G. Brereton, Analyst 134 (2009) 114–123.
[20] S.J. Dixon, N. Heinrich, M. Holmboe, M.L. Schaefer, R.R. Reed, J. Trevejo, R.G. Brereton, J. Chemometr. 23 (2009) 19–31.
[21] H.F.M. Boelens, R.J. Dijkstra, P.H.C. Eilers, F. Fitzpatrick, J.A. Westerhuis, J. Chromatogr. A 1057 (2004) 21–30.
[22] D.L. Donoho, IEEE Trans. Inform. Theory 42 (1995) 613–627.
[23] D.H. Williams, I. Fleming, Spectroscopic Methods in Organic Chemistry, fifth ed., McGraw-Hill, London, 1995.
[24] J. Coates, Encyclopaedia of Analytical Chemistry, Wiley, Chichester, 2000.
[25] R.G. Brereton, Chemometrics for Pattern Recognition, Wiley, Chichester, 2009.
[26] T. Kohonen, Self Organising Maps, third ed., Springer, Berlin, 2001.
[27] G.R. Lloyd, R.G. Brereton, J.C. Duncan, Analyst 133 (2008) 1046–1059.

[28] S. Johnson, N. Saikia, Technical Report, Centre for Science & Environment, Pollution Monitoring Laboratory, New Delhi, India, 2009.
[29] R.M. Silverstein, G.C. Blaser, T.C. Morril, Spectrometric Identification of Organic Compounds, Wiley, New York, 1974, pp. 73–119.
[30] M.D. Guillén, N. Cabo, J. Am. Oil Chem. Soc. 74 (1997) 1281–1286.
[31] Food Standards Agency, McCance & Widdowson's, The composition of Foods, Royal Society of Chemistry, Cambridge, 2002.
[32] S. Yoshida, H. Yoshida, Biopolymers 70 (2003) 604–613.
[33] N. Vlachos, Y. Skopelitis, M. Psaroudaki, V. Konstantinidou, A. Chatsilazarou, E. Tegou, Anal. Chim. Acta 573–574 (2006) 459–465.
[34] M.D. Guillén, I. Carton, E. Goicoichea, P.S. Uriarte, J. Agric. Food Chem. 56 (2008) 9072–9079.
[35] M.A. Moharam, L.M. Abbas, Afr. J. Micro. Res. 4 (2010) 1921–1927.
[36] R.S. Farag, M.A.S. El-Agaimy, B.S. Abd El-Hakeem, Food Nutr. Sci. 1 (2010) 24–29.