ENGINEERING

Evaluation of benzene residue in edible oils using Fourier transform infrared (FTIR) spectroscopy

Ritu Joshi[1], Byoung-Kwan Cho[1,*], Santosh Lohumi[1], Rahul Joshi[1], Jayoung Lee[1], Hoonsoo Lee[2], Changyeun Mo[3,*]

[1]Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, Daejeon 34134, Korea

[2]Department of Biosystems Engineering, College of Agriculture, Life & Environment Science, Chungbuk National University, Cheongju 28644, Korea

[3]Department of Biosystems Engineering, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon 24341, Korea

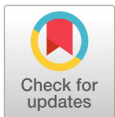*Corresponding author: chobk@cnu.ac.kr; cymo76@gmail.com

Abstract

The use of food grade hexane (FGH) for edible oil extraction is responsible for the presence of benzene in the crude oil. Benzene is a Group 1 carcinogen and could pose a serious threat to the health of consumer. However, its detection still depends on classical methods using chromatography which requires a rapid non-destructive detection method. Hence, the aim of this study was to investigate the feasibility of using Fourier transform infrared (FTIR) spectroscopy combined with multivariate analysis to detect and quantify the benzene residue in edible oil (sesame and cottonseed oil). Oil samples were adulterated with varying quantities of benzene, and their FTIR spectra were acquired with an attenuated total reflectance (ATR) method. Optimal variables for a partial least-squares regression (PLSR) model were selected using the variable importance in projection (VIP) and the selectivity ratio (SR) methods. The developed PLS models with whole variables and the VIP- and SR-selected variables were validated against an independent data set which resulted in R^2 values of 0.95, 0.96, and 0.95 and standard error of prediction (SEP) values of 38.5, 33.7, and 41.7 mg/L, respectively. The proposed technique of FTIR combined with multivariate analysis and variable selection methods can detect benzene residuals in edible oils with the advantages of being fast and simple and thus, can replace the conventional methods used for the same purpose.

Keywords: benzene adulteration, edible oils, food safety, FTIR spectroscopy, rapid measurement

Introduction

Edible oils are subjected to various processes to ensure their suitability for human consumption. Edible oils have a critical effect on the taste and mouth-feel of foods, whilst enhancing the nutritive

value of the food (Hong et al., 2018). Sesame oil have a mild odour, pleasurable taste and as such used as a natural salad oil requiring little or no gear up. Sesame oil is described to be the high nutritive and biological values as well as excellent quality taste (Park et al., 2013). It is widely used as a cooking oil, in shortening and margarine, as a soap fat, in pharmaceuticals and work as reactionary for insecticides (Budowski and Markely, 1951). While, cottonseed oil has long been considered to be a good vegetable oil for frying, in part because it tends to impart a toasted aroma to fried products (Dowd et al., 2010). Chemical oil extraction, which uses a solvent during oil extraction, is popular commercial procedure because it produces high yields in fast and inexpensive way. In a previous study, Jomtib et al. (2011) used hexane, benzene, and toluene as co-solvents to determine the effect of adding co-solvents to the oil in various concentrations (10 - 50% v/v) on the formation of methyl esters. Benzene was used during the oil extraction procedure because it can extract a higher quantity of oil than other solvents. Because of the carcinogenic property of benzene, n-hexane is globally preferred solvent because of its extraction efficiency, easy availability, high stability, low evaporation loss, low corrosion, low greasy residue, and better odour and flavour (Saxena et al., 2011). The use of hexane as a solvent during oil extraction may also contribute to the occurrence of benzene in food (Masohan et al., 2000). The low boiling point of benzene compared to edible oil suggested that some residues remain in the oil after extraction. Recently, benzene residues were found in cottonseed oil and these remnants provide the motivation to identify and quantify benzene in edible oil because its presence is directly related to consumer health.

Benzene, a volatile organic compound has been classified as human carcinogen by Environmental Protection Agency (EPA) that can form when benzoate and ascorbic acids were present under the influence of heat, UV light, and metal ions as catalysts (Styarini et al., 2011). For general studies, level of benzene in drinking water (10 and 5 μg/L) is usually adopted by WHO and United States Environmental Protection Agency (USEPA) for references because there is no specific limit of benzene in food and beverages (Aprea et al., 2008; Vinci et al., 2012). Also the maximum limit for benzene concentration based on toxicity has been set in Europe at 5 mg/L (Atkinson et al., 1995). Most cases of benzene toxicity have been reported in Italy (Vigliani and Saita, 1964; Vigliani, 1976), and Turkey (Aksoy et al., 1972; Aksoy et al., 1974), which together had the highest rate of chronic myelogenous and myelogenous leukaemia.

Several analytical techniques have been applied to the detection of benzene in edible oil samples. Masohan et al. (2000) estimated the residue of benzene in crude and refined samples of rice bran and soy oil, and the oil-extracted cakes using gas chromatography and UV spectroscopy. In another study, Styarini et al. (2011) detected benzene in beverages using headspace gas chromatography. Furthermore, solid-phase micro-extraction and gas chromatography was used for the detection of benzene in beverages i.e., soft-drink, juice and tea samples (Sanchez et al., 2012). Within this context, it is evidently necessary to develop analytical techniques to make it possible to identify the solvent residue in edible oils as these methods required sample preparation and the use of chemicals which is destructive, leading to the end use of samples.

Fourier-transform infrared (FTIR) spectroscopy is a method used to determine the structures of molecules by their characteristic absorption of infrared radiation and the resulting molecular vibrational spectra. Spectroscopy is regularly used for both the qualitative and quantitative analysis of agricultural and food products and presents an alternative to time-consuming, wet-chemical, and destructive techniques (Lim et al., 2017; Mo et al., 2017; Qin et al., 2017; Ning et al., 2018). A key advantage of this technique is its high-speed operation; a sample can be analysed in seconds, and the spectrometer simultaneously collects all light frequencies that are transmitted or reflected from the sample. FTIR spectroscopy measurements are also non-destructive, making them successful in evaluating the quality of agriculture products and beverages.

FTIR spectroscopy has previously been combined with discriminant analysis and partial least-squares (PLS) analysis, and has been successfully used to quantify adulterants, such as refined oils and different types of vegetable and nut oils, in extra virgin olive oil (Lai 1995; Marigheto et al., 1998; Kupper et al., 2001). IR studies of edible oils generally use specific absorbance bands to evaluate traditional indices and other parameters of interest in relation to the composition of edible oils (van de Voort et al., 1992; Che Man and Setiowaty, 1999; Che Man and Mirghani, 2000; Setiowaty et al., 2000). For example, a common fraud is the adulteration of Moroccan olive oil mixed with other edible oils of lower commercial value (Flores et al., 2006). IR spectroscopy and PLS analysis were used to quantify the percentage of adulterants such as soybean oil, pure tea seed oil and sunflower oil in virgin walnut oil (Liang et al., 2013).

Currently, a wide range of vibrational spectroscopic techniques in combination with chemometrics have shown potential as sensitive and rapid techniques for the authentication and quality analysis of a variety of food products. Our study devoted to achieve quantitative detection of benzene residues in edible oil using FTIR spectroscopy. We specifically attempted predict the benzene concentration in edible oils using FTIR spectroscopy with an integrated PLS model. We conducted spectral analysis of five concentrations of benzene in edible oils, which were chosen as 0, 100, 200, 300, 400, and 500 mg/L. Based on the results, the different concentrations were identified and categorized using a multivariate analytical method.

Materials and Methods

Sample Preparation

Commercial samples of two different edible oils (cottonseed oil and sesame oil) were purchased from a market in the South Korea. Benzene was purchased from Sigma Aldrich (St. Louis, USA) which is essentially used to extract cooking oil from seeds. The edible oil samples were spiked with benzene at various concentrations (100, 200, 300, 400 and 500 mg/L). The spiked oil samples were placed in snap-cap vials and shaken with a high-speed shaker (Vortex-Genie® 2, Scientific Industries, Inc., Model G560, USA) for over 2 min. Ten samples of each of pure and benzene-spiked edible oil were used for FTIR data analysis; therefore, a total of 120 samples (60 samples for each edible oil) were used to measure their spectra by FTIR.

Spectral Measurements

The sample measurements were performed using a Nicolet 6700 (Thermo Scientific Co., Madison, USA) FTIR spectrophotometer was configured with an attenuated total reflectance (ATR) accessory, a deuterated triglycine sulphate (DTGS) detector, and a KBr beam splitter controlled by OMNIC software. The spectra were measured separately for each sample between 4,000 and 650 cm^{-1}, and a total of 32 successive scans were collected from each sample with 4 cm^{-1} intervals. For measurement, a drop of oil was deposited on the surface of the diamond crystal sampling plate using a 1 mL syringe. After measuring a sample, the oil was removed with a dry tissue and then the surface was rinsed firstly with alcohol and then with water before moving to another sample. Finally, the surface was dried with a clean tissue. As, benzene is a highly volatile compound, we believe that there are no traces remain in the ATR after cleaning. Before

recording the sample spectra, a background scan was obtained once for pure oil and once for adulterated oil samples with an empty sample plate.

Data Analysis

Data Pre-processing

Spectral data pre-processing is one of the most critical steps in a data mining process that deals with the preparation and transformation of the initial dataset. Several pre-processing methods have been proposed to model the effect of multiplicative light scattering (Chen et al., 2006). Multiplicative scatter correction (MSC) is a widely used technique (Geladi et al., 1985). In our study, MSC was used to remove the non-linearity in the data caused by scattering from the samples. The MSC operation undergoes into two steps: estimation of the correction coefficients,

$$\mathbf{x_{org}} = \mathbf{b_0} + \mathbf{b_{ref,1}} \mathbf{x_{ref}} + \mathbf{e} \tag{1}$$

and correction of the spectra

$$\mathbf{x_{corr}} = \frac{\mathbf{x_{org}} - \mathbf{b_0}}{\mathbf{b_{ref,1}}} \tag{2}$$

where the b variables are the correction coefficients, e is the modelled part and x_{org} is the original spectra measured by IR instrument, x_{ref} is the reference spectra which is the average over a set of samples and x_{corr} is the corrected spectra, respectively (Rinnan et al., 2009).

Multivariate Analysis

Chemometrics and multivariate data analysis provide the solution to many problems in qualitative and quantitative analysis and are especially useful in adulteration and quality assessment of food products (Muick et al., 1998). The more frequently used method is multivariate analysis, which is a collection of methods that can be used when several measurements are made on objects. Multivariate linear regression is an extension of multiple linear regression to model multiple responses (Jung and Park, 2015). This method is concerned with data sets that have more than one response variable for each observational or experimental unit. We can perform a certain measurement and store the value for a given phenomenon in a univariate or multivariate variable called $y = (y_1, y_2, \ldots, y_m)^T$ where m is number of independent variables (Ami et al., 2000).

Principal component analysis (PCA) and PLS are useful multivariate tools for spectral data analysis because of the quality of their calibration model and their ease of implementation (Goodarcre et al., 2003; Tapp et al., 2003; Wang et al., 2003). These methods are reliable by generating components as a new input variable to linearly compose original input variables for multivariate data analysis and modeling (Yang et al., 2015). Generally, the first few transformed variables are sufficient to account for most of the variations (e.g., PCA) or to maximize separability (e.g., PLS) of the whole data. In our data analysis, PCA was carried out on the MSC-processed FTIR spectra because it can be readily applied to spectroscopic data to perceive the nature and scattering level of the data. PCA is a well-known method in multivariate analysis that is frequently used to maximize the variance of a linear combination of the variables. This method uses sophisticated underlying mathematical principles to transform several possibly correlated variables into smaller number of variables (principal components) (Richardson, 2009). The principal components (PCs) are orthogonal and the first few principal

components (i.e., PC1, PC2, etc.) provide most of the information about the material. The linear combination created by principal components can be expressed in the form:

$$y = a_1x_1 + a_2x_2 + \ldots + a_nx_n \tag{3}$$

where PCA treats the peak positions as vectors $(x_1, x_2, \ldots, x_n)$ and forms a linear combinations of the vectors by assigning a weight $(a_1, a_2, \ldots, a_n)$ to each vector in the spectra (Rusak et al., 2003). When predictors are reduced to a smaller set of uncorrelated components partial least-squares regression (PLSR) can be used on these components rather than on the original data. PLSR is especially useful when predictors are highly collinear, or when there are more predictors than observations. PLSR provides information about the correlation structures of the variables and about their structural similarities or dissimilarities. In this study, PLSR was developed for preprocessed spectral data to predict the content of benzene residues in the edible oil samples. The PLSR equation is given as:

$$X = TP^T + E \tag{4}$$

$$Y = UQ^T + F \tag{5}$$

where a spectral data matrix X is decomposed into the score matrix T, loading matrix P, and error matrix E, and the reference values matrix Y is decomposed into the score matrix U, loading matrix Q, and error matrix F. The basis of present study mainly focus on constructing and selecting the subsets of features that are useful to build a good predictor. This approach contrasts with the problem of finding or ranking all potentially relevant variables. Selecting the most relevant variables is usually suboptimal for building a predictor, particularly if the variables are redundant. Conversely, a subset of useful variables may exclude many redundant variables. The quality of the calibration model is described by the squared coefficient of determination (R^2) and the standard error of prediction (SEP). The best calibration model for prediction was the one with the highest value of R^2 and the lowest SEP value.

Variable Selection

Selecting the most relevant variables is usually suboptimal for building a predictor, particularly if the variables are redundant. The variable importance in projection (VIP) includes a measurement of the variable dependency, which is considered as a benefit of this multivariate filter method. VIP calculates how much a variable contributes to the description of the dependent or reference data sets (Y) and the independent or spectral variables (X) (Lohumi et al., 2015). The VIP score of variable j is calculated by the following equation:

$$VIP_j = \sqrt{\frac{\sum_{F=1}^{F} W_{jf}^2 \cdot SSY_f \cdot J}{SSY_{total} \cdot F}} \tag{6}$$

where W_{jf} is the weight value for component f of variable j, SSY_f is the sum of squares of explained variance for the fth component, J is the number of variables, SSY_{total} is the total sum of the squares for the dependent variable, and F is the total number of components. A variable with higher VIP score is more relevant to the prediction of the response variable. Normally, the average of the squared values of the VIPs is equal to 1 (Cho et al., 2002). The criterion of VIP value with greater than 1 is then often used as a cut-off point for variable selection (Lazraq et al., 2003; Chong and Jun, 2005).

The selectivity ratio (SR) denotes the ratio between the explained and the residual (unexplained) variance for each variable in the target projection (Farres et al., 2015). A high value denotes a variable with good predictive performance (Anderssen et al., 2006). This method essentially visualizes the important variables of a multivariate data set in the prediction of a property (Rajalahti et al., 2009a). The target projection model that calculates the explained and residual variance for each variable can be written as:

$$X = t_{TP}\, P_{TP}T + E_{TP} = X_{TP} + E_{TP} \tag{7}$$

where t_{TP} is the target-projection score, P_{TP} is target projection loading and E_{TP} is target-projection residual (Lohumi et al., 2015). All multivariate analysis was performed using MATLAB software version 7.0.4 (The Mathworks, Natick, USA).

Results and Discussion

Spectral exploration

The raw FTIR spectra of both sesame and cottonseed oils are shown in Fig. 1. The raw spectra revealed some peaks in both the fingerprint and functional group regions for sesame and cottonseed oils. However, some parts of the functional group region such as the region 4000 - 3156 cm^{-1}, which is assigned to the hydroxyl stretching band, no notable difference were observed in intensity of spectra of oil with different concentration of benzene, therefore, we discarded this region. The absorption in the 2700 - 3000 cm^{-1} region is associated with methylene stretching (McMullin et al., 2015), and we obtained a small peak in the region from 2435 - 2246 cm^{-1} because of the effect of background CO_2. The variations in these spectral
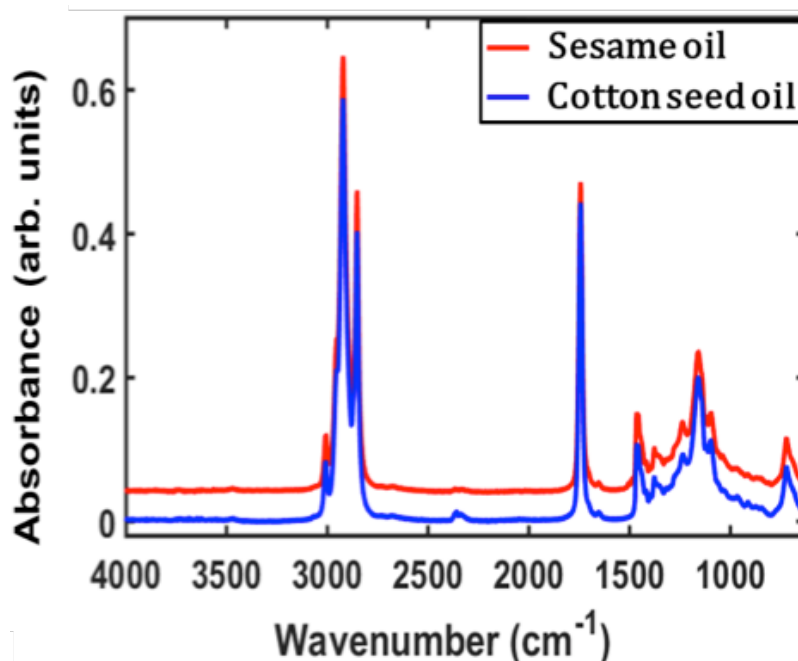

Fig. 1. Raw Fourier transform infrared (FTIR) spectra for pure sesame and cottonseed oil.

regions were not attributed to changes in the sample composition. Therefore, only the remaining spectral region was used for further data analysis to minimize the influence of these regions on model development.

The peak at ~ 3009 cm^{-1} was assigned to the C=CH (cis) stretching vibration mode, and the band at ~ 1742 cm^{-1} was related to the stretching mode of –C=O bonds in ester groups, which are found in samples with a high content of saturated fatty acids (Lerma-Garcia et al., 2010; Rohman and Man, 2010). The peak at ~ 1461 cm^{-1} and ~ 1378 cm^{-1} represented the –C–H (CH$_2$) bending (scissoring) mode of vibration and the –C–H (CH$_3$) bending symmetric vibrational mode, respectively. In the fingerprint region, the peaks at 1235 and 1161 cm^{-1}, and 1118 and 1098 cm^{-1} were related to the C–O (ester) and C–O stretching mode of vibration. Trans –CH=CH- out of plane bending peaks were observed at ~ 964, 914, 871 and 844 cm^{-1}, while the peak at ~ 721 cm^{-1} is related to the –C=O stretching mode. The functional groups and vibrational modes in the FTIR spectra of edible oil were similar those reported previously (Lerma-Garcia et al., 2010; Rohman and Man, 2010).

Principal component analysis interpretation

The selected FTIR data were preprocessed using an MSC method before conducting the multivariate data analysis and then PCA was applied on the edible oils data to check for both possible outliers and natural data groupings. The purpose of the PCA method is to concentrate the source of variability in the data into the first few PCs by decomposing the data matrix (Hori and Sugiyama, 2002). The score plot is a projection of data onto a subspace that is used to interpret the relations between observations. The resulting scatter plot of the PC scores showed one outlier (marked with a blue box in Fig. 2) from each oil group. In Fig. 2, the representative points of the sesame oil (Fig. 2a) and cottonseed oil (Fig. 2b) samples are mapped in the space spanned by the first two principal components. These score plots showed that a reasonable clustering was present for the different concentrations of benzene added to both edible oils.
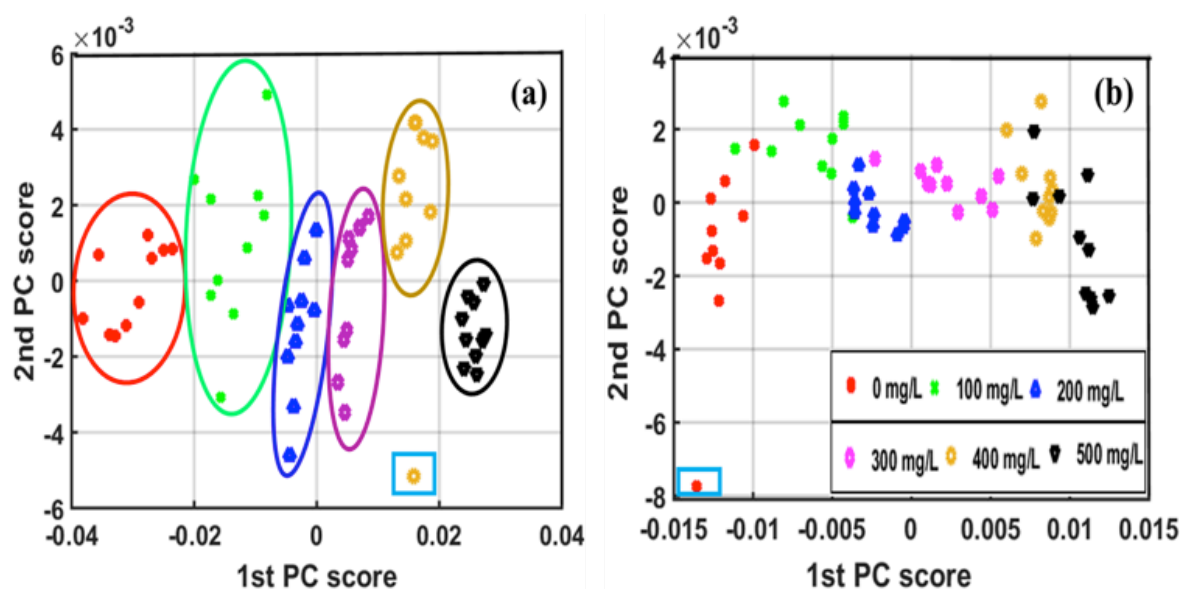

Fig. 2. Principal component analysis (PCA) score plot for sesame oil (a) and cottonseed oil (b) after applying multiplicative scatter correction (MSC) preprocessing.

Further, we attempted to interpret the first three PC loadings (PC1, PC2, and PC3), explaining about 98% of total variance in terms of chemical composition. These loadings give a correlation between a component and a variable that estimates the information they share. Using these plots can extract information about which variable have the largest effect on each component. In addition, these loading plots are helpful for characterizing each component in terms of variables. The loading of PC1 (Fig. 3a) shows a small peak at around 3000 cm^{-1}, a distinct peak around 1500 cm^{-1}, and an upwards trend at the end. In addition, PC2 (Fig. 3b) shows two strong negative peaks at ~ 2000 cm^{-1} and a small positive peak at ~ 1600 cm^{-1}, while PC3 (Fig. 3c) shows a negative peak in the same region (1600 cm^{-1}) caused by the variation in benzene concentration among the samples shown in Fig. 3d.
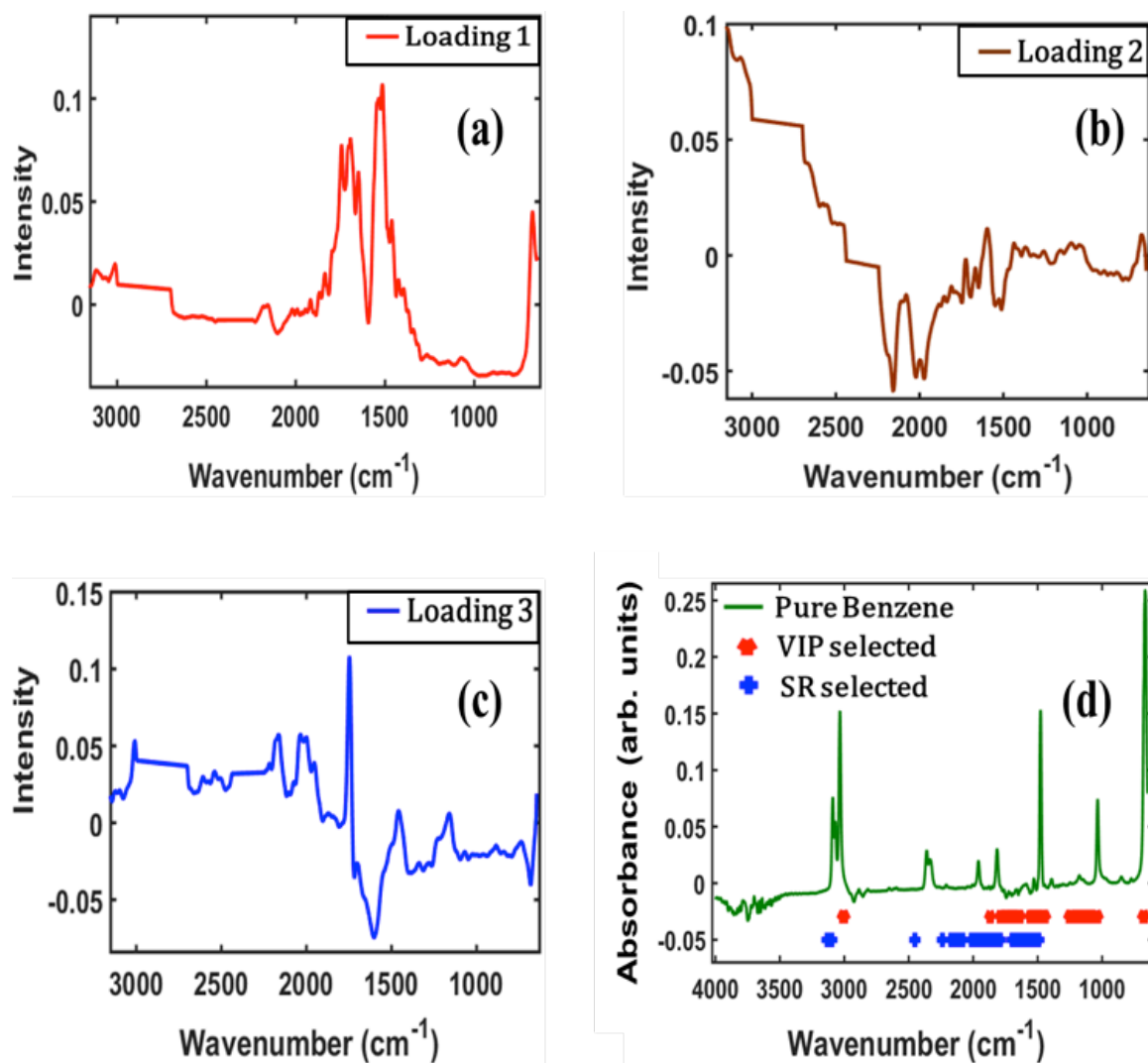

Fig. 3. First three principal components loading plots from principal component analysis (PCA) for benzene-adulterated sesame oil (a), (b), (c) and the spectrum of pure benzene with variable importance in projection (VIP) and selectivity ratio (SR) for selected variables (d).

PLSR Model

A PLSR model was employed to develop a predictive model for detecting the added benzene concentration in edible oils. After discarding the two outlier samples (sesame oil: 400 mg/L; cottonseed oil: 0 mg/L), the samples were categorized based on the adulterant concentration. Totally 118 samples (after the removal of two outliers) from both edible oils was divided into calibration (70 samples) and validation (48 samples) set in a ratio of 6 : 4 to evaluate the accuracy of the model. The PLSR model was developed using the MSC-preprocessed spectra of pure and adulterated oil samples. In the multivariate analysis, two data sets were used for calibration: X (independent variables, i.e., spectral data) and Y (dependent variable, i.e., adulteration percentage), and regressed to develop the model for prediction. The validation set, which was not used in model development, was used to test the predictive accuracy of the developed model.

The calibration model gave a very high correlation value (R^2) of 0.99 with a standard error of calibration (SEC) of 15.1 mg/L. However, the R^2 and SEP for prediction were 0.95 and 38.5 mg/L, respectively (Table.1). Fig. 4a shows the actual
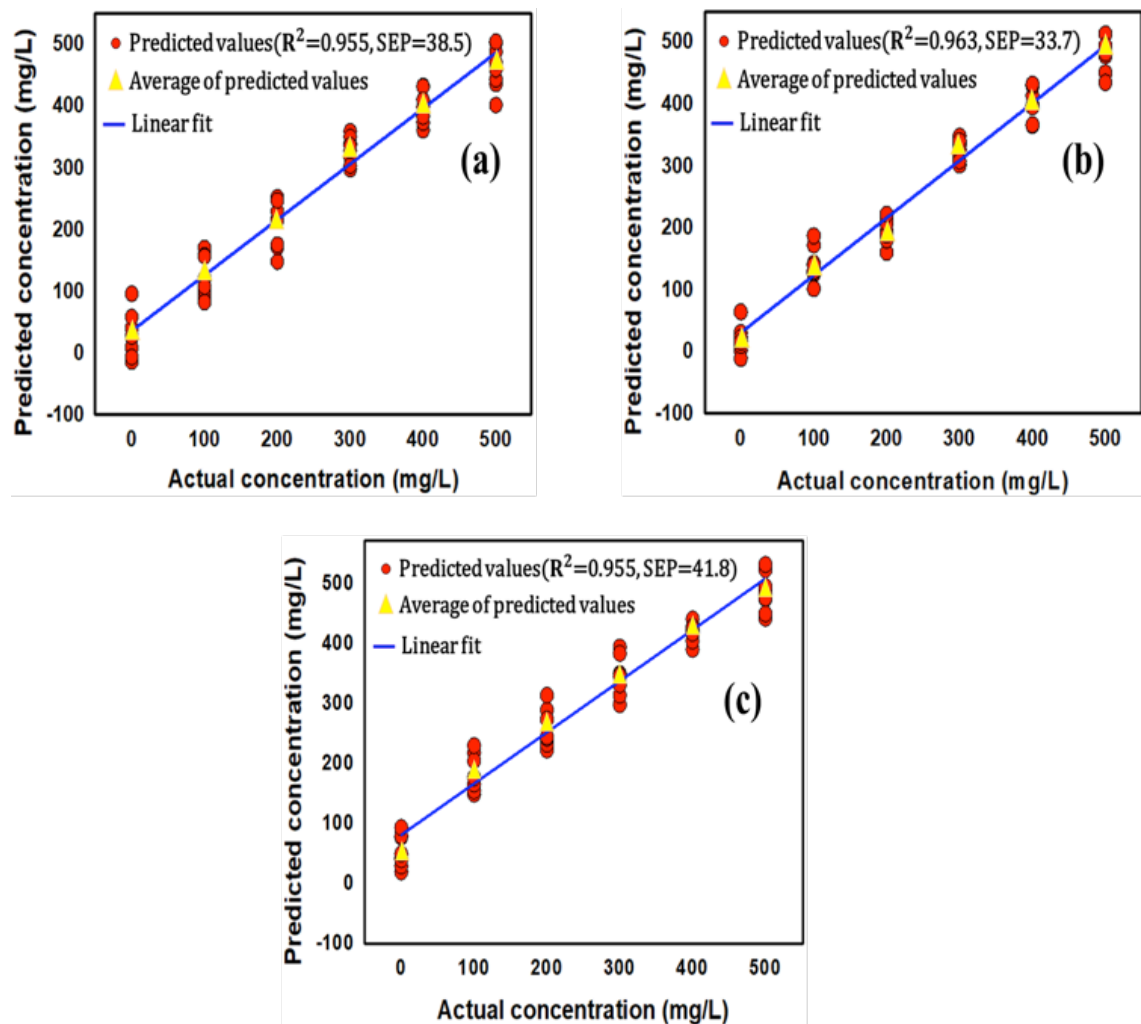

Fig. 4. Regression plot of the actual versus calculated percentages of benzene in the validation set of the whole spectral region (a), VIP (b) and SR (c) variable selection methods. VIP, variable importance in projection; SR, selectivity ratio.

and predicted concentrations of benzene in edible oils by the PLSR model for the validation set. We also determined the optimal number of factors based on the lowest root mean square error (RMSE) in the validation process, and seven factors were selected.

Two model-based variable selection methods, i.e., VIP and SR were executed on PLS-based results to select the optimal variables. The VIP measurement includes the variable dependency which is a key benefit of the multivariate filter method. However, the SR is used to avoid model overfitting and improves the predictive competence. SR is usually applied to filter out irrelevant variables (Kvalheim and Karstang, 1989; Rajalahti et al., 2009b). By assigning a threshold value 1.2 for VIP and 0.03 for SR, we selected a total of 166 and 141 variables for pure and adulterated oil samples, respectively. Then, the PLSR model were developed for both variable selection methods. A summary of the results is shown in Table 1. All the values found for the parameters in Table 1 suggested that the model developed with selected variables afforded either higher or similar prediction accuracy compared to the PLS model developed with whole corrected variables. However, the highest coefficient of determination ($R_p^2 = 0.96$) with standard error of prediction (SEP = 33.7 mg/L) was obtained using the VIP variable selection method. Fig. 4b and 4c show the excellent prediction ability of the PLSR model developed with selected variables.

The VIP and SR selected variables are represented against the spectra of pure benzene in Fig. 3d. The spectra showed that most of the selected variables by VIP method were related to the benzene-sensitive bands while SR selected variables are dissimilar to those selected by VIP. A simple visual comparison of the variables selected using these two different methods suggested that the VIP selected variables were more genuine than SR selected variables when compared with the pure benzene spectrum. This improved performance in VIP could be because SR was limited by the selection of a reliable threshold for assessing the significance of a selected discriminating variable. The selected variables for VIP and SR were 15.9% and 13.5% of the total variables, respectively. Visual inspection of the beta coefficient from PLSR model (Fig. 5a) showed that certain peaks within certain spectral regions were important for differentiating between pure and adulterated oil samples. These distinct peaks are influenced by the different benzene concentration of the other group of samples. However, minor peaks can be caused by the spectral variations between the two different kinds of edible oils.

Residual plots which illustrate the residual against the corresponding fitted values or the explanatory variables have been widely used to diagnose the regression model in terms of model structure such as numbers and types of variables,

Table 1. Prediction results by partial least squares regression (PLSR), variable importance in projection (VIP), and selectivity ratio (SR) variable-selection methods for detecting pure and benzene-adulterated edible oils.

Methods	Calibration		Validation		Prediction		Factors/
	$R_c^{2\ u}$	SEC (mg/L)[v]	$R_v^{2\ w}$	SEV (mg/L)[x]	$R_p^{2\ y}$	SEP (mg/L)[z]	Variables
Whole spectra	0.99	15.1	0.95	36.6	0.95	38.5	7/1046
VIP	0.99	18.9	0.95	39.5	0.96	33.7	8/166
SR	0.98	25.9	0.95	37.6	0.95	41.7	5/141

R^2 values for [u]calibration, [w]validation, and [y]prediction. [v]standard error of calibration, [x]standard error of validation, and [z]standard error of prediction are the standard errors of calibration, validation, and prediction, respectively.

inclusion or exclusion of interaction terms, and necessity of higher order terms or non-linear terms (Kutner et al., 2008). An increasing trend in the residuals plot suggests that the error variance increases with the independent variable; while a decreasing trend indicates that the error variance decreases with the independent variable. Also, one of residual plots showing the standardized residuals vs. the predicted values is useful in detecting violations in linearity (Stevens, 2009). Fig. 5b shows a residual plot against the sample number to study the relationship between the different concentrations of benzene and the values predicted by the whole variable, PLS-VIP, and PLS-SR models for edible oil samples. The obtained residual figure shows an identical pattern for both whole variables, and PLS-VIP as these methods get negative values for low concentration. On comparing with these two methods, PLS-SR extract more negative values for all concentration which gives a small change in the random pattern of the residuals. Thus, it shows an agreement between actual and predicted values for benzene concentration and provides a decent fit for a linear model.
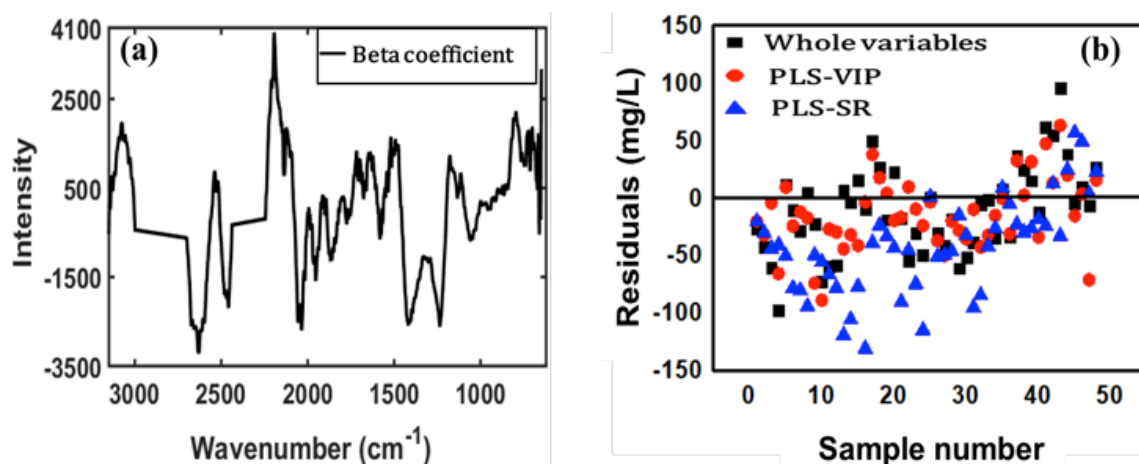

Fig. 5. Beta coefficient plot from the PLSR (a), Residual plot for whole variables, PLS-VIP and PLS-SR method (b). PLSR, partial least squares regression; VIP, variable importance in projection; SR, selectivity ratio.

Conclusion

In this study, FTIR spectroscopy combined with PLS multivariate analysis was demonstrated to be capable of detecting trace amounts of benzene in edible oils. Variable selection methods were additionally adopted to select the informative variables and avoid model over fitting and they improved the model accuracy developed by PLS. Also, the selected variables were authentic by showing distinct peaks in the same spectral regions when compared to the pure benzene spectrum. The results of this study indicate that specific FTIR spectral regions are effective for the determination of benzene traces in edible oils. Our approach highlights that FTIR spectroscopy is a rapid technique that can be performed with no sample preparation, and thus has a potential to be an effective analytical tool for the detection of benzene trace in a variety of edible oils.

Acknowledgment

Authors Information

Ritu Joshi, Chungnam National University Department of Biosystems Machinery Engineering College of Agricultural and Life Science, Ph.D. student

Byoung-Kwan Cho, https://orcid.org/0000-0002-8397-9853

Santosh Lohumi, Chungnam National University Department of Biosystems Machinery Engineering College of Agricultural and Life Science, Postdoctoral researcher

Rahul Joshi, Chungnam National University Department of Biosystems Machinery Engineering College of Agricultural and Life Science, Ph.D. student

Jayoung Lee, Chungnam National University Department of Biosystems Machinery Engineering College of Agricultural and Life Science, Master student

Hoonsoo Lee, https://orcid.org/0000-0001-8074-4234

Changyeun Mo, https://orcid.org/0000-0002-9088-5978

References

Aksoy M, Dincol K, Erdem S, Dincol G. 1972. Acute leukaemia due to chronic exposure to benzene. The American Journal of Medicine 52:160-166.

Aksoy M, Erdem S, Dincol G. 1974. Leukaemia in shoe workers exposed chronically to benzene. Blood 44:837-841.

Anderssen E, Dyrstad K, Westad F, Martens H. 2006. Reducing over-optimism in variable selection by cross-model validation. Chemometrics and Intelligent Laboratory System 84:69-74.

Ami D, Mereghetti P, Doglia SM. 2000. Multivariate analysis for Fourier transform infrared spectra of complex biological systems and processes. Multivariate Analysis in Management, Engineering and the Sciences, IntechOpen, 2013:189-220. http://dx.doi.org/10.5772/53850.

Aprea E, Biasioli F, Carlin S, Mark TD, Gasperi F. 2008. Monitoring benzene formation from benzoate in model systems by proton transfer reaction-mass spectrometry. International Journal of Mass Spectrometry 275:117-121.

Atkinson R, Boissard C, Cao X, Chandler J, Crump DR, Davies TJ, Delaney M, Derwent RG, Dollard GJ, Duckham SC, Dumitrean P, Field RA, Hewitt CN, Jones BMR, Midgley PM, Murlis J, Nason PD, Passant NR, Watkins D. 1995. Volatile organic compounds in the atmosphere. Environmental Science and Technology 4 Hester RE and Harrison RM eds. Royal Society of Chemistry, London, UK.

Budowski P, Markely KS. 1951. The chemical and physiological properties of sesame oil. Chemical Reviews 48:125-151.

Che Man YB, Setiowaty G. 1999. Determination of anisidine value in thermally oxidized palm olein by Fourier

transform infrared spectroscopy. Journal of the American Oil Chemist's Society 76:243-247.

Che Man YB, Mirghani MES. 2000. Rapid method for determining moisture content in crude palm oil by Fourier transform infrared spectroscopy. Journal of the American Oil Chemist's Society 77:631-637.

Chen ZP, Morris J, Martin E. 2006. Extracting chemical information from spectral data with multiplicative light scattering effects by optical path-length estimation and correction. Analytical Chemistry 78:7674-7681.

Cho JH, Lee D, Park JH, Kim K, Lee IB. 2002. Optimal approach for classification of acute leukemia subtypes based on gene expression data. Biotechnology Progress 18:847-854.

Chong IG, Jun CH. 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems 78:103-112.

Dowd MK, Boykin DL, Meredith WR, Campbell Jr BT, Bourland FM, Gannaway JR, Glass KM, Zhang J. 2010. Breeding and genetics: Fatty acid profiles of cottonseed genotypes from the national cotton variety trials. Journal of Cotton Science 14:64-73.

Farres M, Platikano S, Tsakovski S, Tauler R. 2015. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. Journal of Chemometrics 29:528-536.

Flores G, Ruiz Del Castillo ML, Herraiz M, Blanch GP. 2006. Study of the adulteration of olive oil with hazelnut oil by on-line coupled high performance liquid chromatographic and gas chromatographic analysis of filbertone. Food Chemistry 97:742-749.

Geladi P, McDougall D, Martens H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Applied Spectroscopy 39:491-500.

Goodarcre R, York EV, Heald JK, Scott IM. 2003. Chemometric discrimination of unfractionated plant extracts analyzed by electrospray mass spectrometry. Phytochemistry 62:859-863.

Hong SJ, Lee AY, Han YH, Park JM, So JD, Kim GS. 2018. Rancidity prediction of soybean oil by using near infrared spectroscopy techniques. Journal of Biosystems Engineering 43:219-228.

Hori R, Sugiyama J. 2002. A combined FT-IR microscopy and principal component analysis on softwood cell walls. Carbohydrate Polymers 52:449-453.

Jomtib N, Prommuak C, Goto M, Sasaki M, Shotipruk A. 2011. Effect of co-solvents on transesterification of refined palm oil in supercritical methanol. Engineering Journal 15:49-58.

Jung SY, Park CS. 2015. Variable selection with nonconcave penalty function on reduced-rank regression. Communications for Statistical Applications and Methods 22:41-54.

Kupper L, Heise HM, Lampen P, Davies AN, McIntyre P. 2001. Authentication and quantification of extra virgin olive oils by attenuated total reflectance infrared spectroscopy using silver halide fiber probes and partial least-squares calibration. Applied Spectroscopy 55:563-570.

Kutner M, Nachtsheim CJ, Neter J, Li W. 2008. Applied linear statistical models (5th ed.). Journal of American Statistical Association 103:880-880.

Kvalheim OM, Karstang TV. 1989. Interpretation of latent-variable regression models. Chemometrics and Intelligent Laboratory System 7:39-51.

Lai YW, Kemsley EK, Wilson RH. 1995. Quantitative analysis of potential adulterants of extra virgin olive oil using

infrared spectroscopy. Food Chemistry 53:95-98.

Lazraq A, Cleroux R, Gauchi JP. 2003. Selecting both latent and explanatory variables in the PLS1 regression model. Chemometrics and Intelligent Laboratory Systems 66:117-126.

Lerma-Garcia MJ, Ramis-Ramos G, Herrero-Martinez JM, Simo-Alfonso EF. 2010. Authentication of extra virgin olive oils by Fourier transform infrared spectroscopy. Food Chemistry 118:78-83.

Liang P, Wang H, Chen C, Ge F, Liu D, Li S, Han B, Xiong X, Zhao S. 2013. The use of Fourier transform infrared spectroscopy for quantification of adulteration in virgin walnut oil. Journal of Spectroscopy 2013.

Lim JG, Kim GY, Mo CY, Oh KM, Kim GS, Yoo HC, Ham HH, Kim YT, Kim SM, Kim MS. 2017. Rapid and nondestructive discrimination of *Fusarium Asiaticum* and *Fusarium Graminearum* in hulled barley (*Hordeum vulgare* L.) using near-infrared spectroscopy. Journal of Biosystems Engineering 42:301-313.

Lohumi S, Lee S, Cho BK. 2015. Optimal variable selection for Fourier transform infrared spectroscopic analysis of starch-adulterated garlic powder. Sensors and Actuators B 216:622-658.

Marigheto NA, Kemsley EK, Defernez M, Wilson RH. 1998. A comparison of mid-infrared and Raman spectroscopies for the authentication of edible oils. Journal of the American Oil Chemist's Society 75:987-992.

Masohan A, Parsad G, Khanna MK, Chopra SK, Rawat BS, Garg MO. 2000. Estimation of trace amounts of benzene in solvent-extracted vegetable oils and oil seed cakes. Analyst 125:1687-1689.

McMullin D, Mizaikoff B, Krska R. 2015. Advancements in IR spectroscopic approaches for the determination of fungal derived contaminations in food crops. Analytical and Bioanalytical Chemistry 407:653-660.

Mo C, Lim J, Kwon SW, Lim DK, Kim MS, Kim G, Kang J, Kwon KD, Cho BK. 2017. Hyperspectral imaging and partial least square discriminant analysis for geographical origin discrimination of white rice. Journal of Biosystems Engineering 42:293-300.

Muick L, Norgaad L, Englesen S, Bro R, Andersson C. 1998. Chemometrics in food science-a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. Chemometrics and Intelligent Laboratory System 14:31-60.

Ning XF, Gong YJ, Chen YL, Li H. 2018. Construction of a ginsenoside content-predicting model based on hyperspectral imaging. Journal of Biosystems Engineering 43:369-378.

Park MK, Yoo JH, Lee JB, Im GJ, Kim DH, Kim W-II. 2013. Detection of heavy metal content in sesame oil samples grown in Korea using microwave-assisted acid digestion. Journal of Food Hygiene and Safety 28:45-49.

Qin J, Kim MS, Chao K, Cho BK. 2017. Raman chemical imaging technology for food and agricultural application. Journal of Biosystems Engineering 42:170-189.

Rajalahti T, Arneberg R, Berven F, Myhr KM, Ulvik RJ, Kvalheim OM. 2009a. Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. Chemometrics and Intelligent Laboratory System 95:35-48.

Rajalahti T, Arneberg R, Kroksveen AC, Berle M, Myhr KM, Kvalheim OM. 2009b. Discriminating variable test and selectivity ratio plot: Quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles. Analytical Chemistry 81:2581-2590.

Richardson M. 2009. Principal component analysis. pp. 1-23. Accessed in https://pdfs.semanticscholar.org/37de/3370 1dd73c89a7b410d69070085b583bfd38.pdf on 1 September 2018

Rinnan A, Nørgaard L, van den Berg F, Thygesen J, Bro R, Engelsen SB. 2009. Data pre-processing. Infrared Spectroscopy for Food Quality Analysis and Control Edited by Da-Wen Sun. Elsevier Inc., Amsterdam, Netherlands.

Rohman A, Man YBC. 2010. Fourier transform infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. Food Research International 43:886-892.

Rusak DA, Brown LM, Martin SD. 2003. Classification of vegetable oils by principal component analysis of FTIR spectra. Journal of Chemical Education 80:541-543.

Sanchez AB, Budziak D, Martendal E, Carasek E. 2012. Determination of benzene in beverages by solid-phase micro-extraction and gas chromatography. Scientia Chromatographic 4:209-216.

Saxena DK, Sharma S, Sambi S. 2011. Comparative extraction of cottonseed oil by n-hexane and ethanol. Journal of Engineering and Applied Sciences 6:84-89.

Setiowaty G, Che Man YB, Jinap S, Moh MH. 2000. Quantitative determination of peroxide value in thermally oxidized palm olein by Fourier transform infrared spectroscopy. Phytochemical Analysis 11:74-78.

Stevens JP. 2009. Applied multivariate statistics for the social sciences (5th ed.). Routledge, New York, USA.

Styarini DO, Zaus O, Hamim N. 2011. Validation and uncertainty estimation of analytical method for determination of benzene in beverages. Eurasian Journal of Analytical Chemistry 6:159-172.

Tapp HS, Defernez M, Kemsley EK. 2003. FTIR spectroscopy and multivariate analysis can distinguish geographical origin of extra virgin olive oil. Journal of Agricultural and Food Chemistry 51:6110-6115.

Van de Voort FR, Sedman J, Emo G, Ismail AA. 1992. Rapid and direct iodine value and saponification number determination of fats and oils by attenuated total reflectance/Fourier transform infrared spectroscopy. Journal of American Oil Chemist's Society 69:118-1123.

Vigliani EC, Saita G. 1964. Benzene and leukaemia. The New England Journal of Medicine 271:872-876.

Vigliani EC. 1976. Leukaemia associated with benzene exposure. Annals of the New York Academy of Sciences 271:143-151.

Vinci RM, Jacxsens L, Loco JV. 2012. Assessment of human exposure to benzene through foods from the Belgian market. Chemosphere 88:1001-1007.

Wang YL, Bollard ME, Keun H, Antti H, Beckonert O, Ebbels TM, Lindon JC, Holmes E, Tang HR, Nicholson JK. 2003. Spectral editing and pattern recognition methods applied to high-resolution magic-angle spinning H-1 nuclear magnetic resonance spectroscopy of liver tissues. Analytical Biochemistry 322:26-32.

Yang CC, Novell CG, Marin DP, Ginel JEG, Varo AG, Cho HJ, Kim MS. 2015. Differentiate of beef and fish meals in animal feeds using chemometrics and analytic models. Journal of Biosystems Engineering 40:153-158.