

A New Strategy for Quantitative Proportions in Complex Systems of Blended Oils by Triacylglycerols and Chemometrics Tools

Hong Yang^{1,2} · Pei-Jin Tong² · Yuan-Rong Jiang² · Wen-Ming Cao² · Xi-Chang Wang¹ · Ting-Ting Wei² · Tong Li² · Cheng Zhong² · Yi-Ming Zhang² · Chuan Zhou² · Zhuo-Ping Ding¹ · Yu-Quan Zhang²

Received: 1 June 2016 / Revised: 6 January 2017 / Accepted: 6 March 2017 / Published online: 28 March 2017
© AOCS 2017

Abstract Many oils sold in China and India are a blend of various oils to improve performance, stability, and nutritional characteristics, which are required in their respective markets. Quantitative analysis of the proportions of constitutive components is fundamental to the conformity and adulteration checking of edible blended oil products. A multi linear regression model with constrained linear least squares and exhaustion calculation was applied in this study. The source of the varieties in the model is a database (614 pure oils) of triacylglycerols (TAGs) collected by GC–FID and HPLC–RID. There were 20 groups of binary and ternary blended oils consisting of two or three oils out of five kinds, namely soybean, corn, peanut, rapeseed, and sunflower, which were analyzed and processed separately. Results showed that the method was able to predict the proportions of constitutive components in the edible blended oils, given that relative errors required less than 20%, the accuracy was 98.2% for the binary system if the proportion of each oil in blended oils was more than 20%, while the accuracy was 84.7% for the ternary system if the proportion of each oil in blended oils was more than 10%. The quantitative method is based on a simple analysis to determine the TAGs composition and thus it is useful for quick segregation and quality control of blended oils in routine analysis.

Keywords Edible blended oil · Proportion · Exhaustion calculation · Triacylglycerols (TAGs) · Multiple linear regression (MLR) · Constrained linear least squares (CLLS)

Abbreviations

CLLS	Constrained linear least squares
GA-PLS	Genetic algorithm-partial least squares
GC–FID	Gas chromatography–flame ionization detector
GNO	Peanut oil
HPLC–RID	High-performance liquid chromatography–refractive index detector
IR	Infrared spectroscopy
IS	Internal standard
LLL	Trilinolein
MIR	Mid-infrared spectroscopy
MLR	Multiple linear regression
MZO	Maize oil
NIR	Near infrared spectroscopy
NMR	Nuclear magnetic resonance
OLL	Dilinoleoyl oleoyl glycerol
OLO + SLL	Dioleoyl linoleoyl glycerol + dilinoleoyl stearoyl glycerol
OOO + SLS + SLO	Triolein + distearoyl linoleoyl glycerol + stearoyl linoleoyl oleoyl glycerol
PCA	Principal components analysis
PLL	Dilinoleoyl palmitoyl glycerol
PLO	Palmitoyl linoleoyl oleoyl glycerol
PLP	Dipalmitoyl linoleoyl glycerol
PLS-DA	Partial least squares–discriminant analysis
POP	Dipalmitoyl oleoyl glycerol

✉ Yuan-Rong Jiang
jiangyuanrong@cn.wilmar-intl.com

✉ Xi-Chang Wang
xcwang@shou.edu.cn

¹ College of Food Science and Technology, Shanghai Ocean University, Shanghai 201306, China

² Wilmar Biotechnology Research & Development Center (Shanghai) Co., Ltd., Shanghai 200137, China

POO + PLS	Palmitoyl linoleoyl stearoyl glycerol + dioleoyl palmitoyl glycerol
POS	Palmitoyl oleoyl stearoyl glycerol
RP-HPLC	Reversed-phase high-performance liquid chromatography
RSO	Rapeseed oil
SBO	Soybean oil
SFO	Sunflower oil
SIMCA	Soft independent modeling of class analogy
SOO	Dioleoyl stearoyl glycerol
TAGs	Triacylglycerols
UV	Ultraviolet spectroscopy

Symbols

β_j	Contribution of component j in the considered blended oil
β	Total mass of blended oil combining q components j according to different mass β_j'
β_j	Proportions of constitutive components
$\beta_{j,\text{actual}}$	Actual adding proportions of constitutive components
$\beta_{j,\text{pred}}$	Predicted adding proportions of constitutive components
N_{e_β}	Number of e_β within a given range
N_n	Number of samples in the sample collection n
P_{N_n}	Percentage of the relative errors within a given range in a sample collection
q	Additive (dilutive) components

Introduction

Most vegetable oils have limited applications in their original forms as a result of their specific chemical compositions [1]. For instance, frying oils rich in saturated fatty acids are good in terms of crisping and stability, but they have high melting points, poor operating performances, and are unhealthy. On the contrary, oils rich in polyunsaturated acids are healthy but rather prone to becoming oxidized and generating harmful substances during deep frying [2]. So, blended oils give manufactures greater flexibility to tailor the products to accomplish specific functional properties or to satisfy nutritional requirements. Considering the production cost and the availability of plant sources, blended oils have high market share in developing countries like China and India [3, 4]. Among the oils used for producing blended oils, five kinds of pure oils [soybean (SBO), corn (MZO), peanut (GNO), rapeseed (RSO), sunflower (SFO)]

are the main and important materials in China [5]. Edible blended oils, composed of different vegetable oils, have different nutritional qualities and prices. Adulteration of oil products, involving the replacement of expensive ingredients with cheap substitutes, could be very lucrative for a vendor or a raw material supplier. Therefore, the determination of the quality and authenticity of blended oils is needed to prevent economic losses, mislabeling problems, and unfair trading practices [6–9]. Moreover, the most important index to evaluate the quality of blended oils is to determine the proportions of the constitutive components in it.

To this day, many studies have been focusing on the detection of blended or adulterated vegetable oils; many analytical methods have been proposed, such as spectroscopy [infrared spectroscopy (IR), ultraviolet spectroscopy (UV), nuclear magnetic resonance (NMR), *etc.*] [10–13], chromatography (GC, LC, *etc.*) [14, 15], biotechnology (DNA, enzyme, *etc.*) [16], often followed by chemometric tools. Chemometrics is applied in both qualitative analysis and quantitative analysis. There are two major categories of qualitative analysis techniques—unsupervised (PCA, cluster analysis, *etc.*) and supervised [partial least squares–discriminant analysis (PLS-DA), *etc.*] [17]. Multivariate calibration methods have been applied to construct models for quantification analysis, such as PLS [18–21], principal component regression (PCR) [22], and artificial neural networks (ANN) [23]. With the launch and implementation of Commission Regulation EU No. 29/2012, an analytical method for verifying if the percentage of olive oil in a blended oil is lower or higher than 50% [24] is demanded; therefore, the detection and quantification of the co-occurring components in a blended oil are required in this new view of quality control for blended oils.

Currently spectroscopy techniques (particularly NIR, Raman, MIR spectroscopy) and chemometrics have been used to detect for authenticity in blended oils as rapid and low-cost methods [25]. These works report MIR spectroscopy as an effective technique to detect various kinds of adulteration [26–32]. Oussama *et al.* [12] focused on the detection and quantification of extra virgin olive oil adulterated with soybean oil or sunflower oil using MIR spectroscopy and PLS [33]. A PLS regression model was able to predict the proportion of sunflower oil or soybean oil adulterants in a range from 1% to 24% (w/w) with relative prediction errors lower than 3% for external validation samples. Özdemir and Öztürk [34] measured NIR transmittance of pure olive oils and olive oils adulterated with sunflower oil and corn oil. The proposed method was able to predict olive oil adulteration in binary blends in a range of 5–95% (w/w). However, in the case of ternary blends of olive oil, sunflower oil, and corn oil, a good prediction was observed for the olive oil content but a limited

prediction was observed for the other two vegetable oils. Compared with the spectral method, the biggest advantage of chromatography is that it is a powerful separation technique [17]. Measurement of TAGs by a high temperature gas chromatography (HTGC)–MS system was applied for the quantification of olive oil in binary blends with vegetable oils (sunflower, corn, seeds, sesame, and soya) with several percentages (10–90%) using multivariate statistical analyses, such as soft independent modeling of class analogy (SIMCA), PLS, and genetic algorithm–partial least squares (GA-PLS) [15]. Qualitative classification and compositional quantification of multivariate blends represent a complex multi-scale question which has been studied much less to this point in time. Xu *et al.* [35] reported an accurate and effective method for the identification and quantification of blends of sunflower oil, soybean oil, and peanut oil with three-dimensional fluorescence spectra (3DFS). 3DFS contains more information including greater numbers of samples and more excitation wavelengths (or emission wavelengths) to distinguish overlapping analytcs. However, it has not been widely used because of its high price.

All the vegetable oils including blended oils are generally mixtures of TAGs which make up to 95–98% (w/w) of composition and show characteristic distributions with different concentration levels. The characteristics and the composition of the oils extracted from various vegetable seeds depend on several factors, such as soil, climate, processing technology, harvesting, and chemical degradation process occurring during storage [36, 37]. Although TAGs are more stable than other compositions, like sterols and vitamin E, pure oils do not have exact similarity in their whole TAGs profiles, let alone blended oils. In order to apply this model extensively, large samples are needed to be collected and analyzed. It is only through this database that the composition of blended oil can be routinely quantified [38].

TAG analysis methods have attracted significant attention over the last decades. Papers dealing with the chromatographic analysis of TAGs in edible oils can be found elsewhere [17]. These include gas chromatography (GC) [39, 40], liquid chromatography (LC) [41–43], and supercritical fluid chromatography (SFC) [44]. Depending on different separation principles, these techniques have unique strengths, which can complement each other. However, there have been few studies on the combination of TAG values from GC and HPLC to build mathematical models.

Proportions of co-occurring components in a blended oil are governed by additive–dilutive processes which obey a simple rule [45], i.e., the simulation method complies with the law of conservation of mass. Multiple linear regression (MLR) is commonly used in chromatography to determine relevant proportions of different miscible solvents for target separations of co-occurring chemical components in complex matrices [46]. In this study, the

independent variables in the complex matrix chromatogram are the TAG compositions of pure oils and a blended oil, with which the proportions of constitutive components, i.e., the dependent variables, can be obtained.

The purpose of this study was to verify whether it was possible to quantify the co-occurring components in multi-blended oils with varied proportions of soybean oil, corn oil, peanut oil, rapeseed oil, and sunflower oil by TAGs and chemometric tools, which has not previously been described. There were 20 groups of binary and ternary blended oils consisting of two or three oils out of those five kinds of oils, which were analyzed and processed separately. Chemometrics tools, such as PCA and MLR, were applied in this study.

Materials and Methods

Pure Oils

Most of the oils used to construct the blended oil are refined oils. The refined vegetable oils were provided by manufacturers. The unrefined oil (peanut oil, GNO) was extracted by mechanical processing in order to maintain more flavor substances. The vegetable oils comprised soybean oil (SBO) (147 pure oils), peanut oil (GNO) (179 pure oils), maize oil (MZO) (107 pure oils), sunflower oil (SFO) (65 pure oils), and rapeseed oil (RSO) (116 pure oils). SBO, MZO, and GNO were from China; SFO was from Australia, Argentina, and China; and RSO was from the Netherlands, Australia, and China.

Lab-made Blends Oils

An original simulation approach based on complex networks was carried out to evaluate the effect on the MLR method [45]. In simplex networks, geometrical locations of blends are governed by the within-blend constraint under which the coordinates (mass) of co-occurring oil varieties obey the unit-sum rule:

$$\sum_{j=1}^q \beta'_j = \beta \quad (1)$$

where q is the number of the total components in the considered blended oil, β'_j is the mass of component j , and β is the total mass of the blended oil. Then the mass β'_j is directly converted into proportions β_j by β'_j/β . The blended oil is made by q (≥ 2) additive (dilutive) components j ($j = 1$ to q) which obey a unit (constant) sum constraint:

$$\sum_{j=1}^q \beta_j = 1 \quad (2)$$

As a result of the unit sum of q proportions, the relative variations of q components can be geometrically represented within a $(q - 1)$ -dimensions simplex space (Fig. 1). In other words, simplex is a variation space with $(q - 1)$ dimensions in which one can represent the relative variations of q components.

Figure 1 shows a simplex space corresponding to a segment or an equilateral triangle, representing blended oils with $q = 2$ or 3 components, respectively. The different points in a simplex space have q coordinates corresponding to the proportions β_j of q components.

The q apexes of the simplex space correspond to an extreme “blend” exclusively made by one component (pure oil state). For example, for $q = 3$, there are three extreme states corresponding to the three apexes of an equilateral triangle with the coordinates $(\beta_1 = 1, 0, 0)$, $(0, \beta_2 = 1, 0)$, $(0, 0, \beta_3 = 1)$ (Fig. 1b). Points located inside segments represent blends of two components (bi-varietal blends) excluding a third one, e.g., $(\beta_1, \beta_2, 0)$, $(\beta_1, 0, \beta_3)$, and $(0, \beta_2, \beta_3)$. Points located within simplex space correspond to blends of three components without any null mass: $(\beta_1, \beta_2, \beta_3)$ with all $\beta_j > 0$.

Random sampling is a method to simulate this application in a real-life situation, since it is unknown which pure oils would be used in the blended oil. There were 20 groups of binary and ternary systems which consisted two or three oils out of those five kinds of oils (SBO, MZO, GNO, RSO, SFO) by means of random sampling. They were prepared step by step with fixed intervals (interval of binary system was 5%, interval of ternary system was 10%). The proportions of constitutive components in blended oils varied from 5% to 95% in the binary system and from 10% to 80% in the ternary system.

Binary system (10 groups, 150 blended oils): SBO–MZO, SBO–GNO, SBO–RSO, SBO–SFO, MZO–RSO, MZO–SFO, GNO–RSO, RSO–SFO, GNO–SFO, MZO–GNO.

Ternary system (10 groups, 360 blended oils): SBO–MZO–GNO, SBO–MZO–RSO, SBO–MZO–SFO, SBO–GNO–RSO, SBO–GNO–SFO, SBO–RSO–SFO, MZO–GNO–RSO, MZO–GNO–SFO, MZO–RSO–SFO, GNO–RSO–SFO.

All the vegetable oil samples were stored in dark bottles at -4 °C until analysis.

Chemicals

Glyceryl triundecanoate and tripentadecanoin (purity $\geq 95\%$) were purchased from Sigma. HPLC-grade hexane, acetonitrile, and ethyl acetate were purchased from Fisher Scientific. Other chemicals and reagents of analytical grade were obtained from reliable commercial sources.

Internal Standard Solution Preparation

Tripentadecanoin was diluted with toluene to a concentration of 1.0 mg/mL. This internal standard solution (IS-a) was used for sample preparation of GC–FID analysis.

Glyceryl triundecanoate was diluted with ethyl acetate to a concentration of 20 mg/mL. This internal standard solution (IS-b) was used for sample preparation of HPLC–RID analysis.

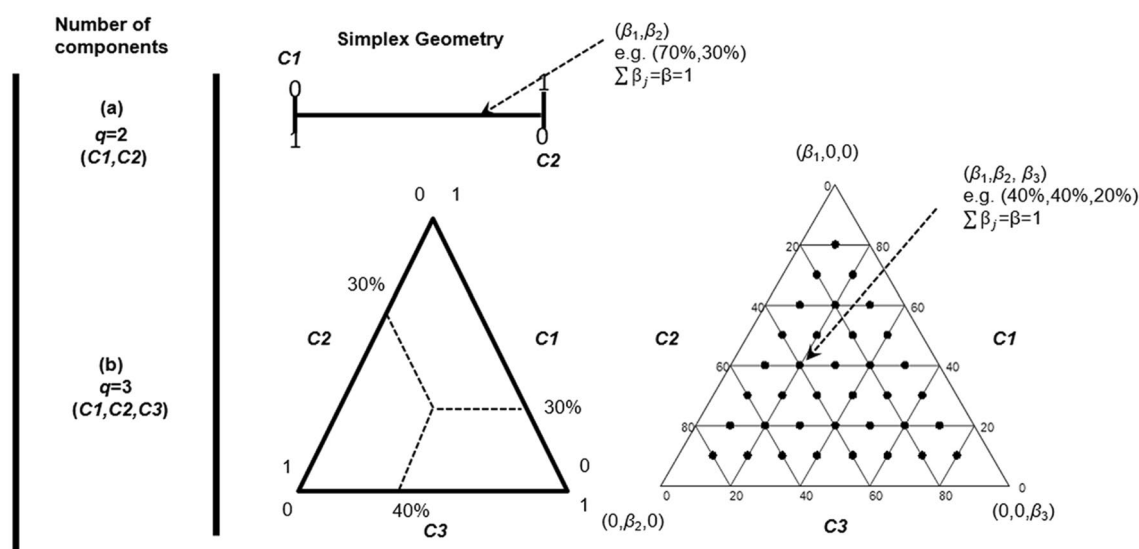


Fig. 1 Geometrical representations for different simplex systems showing different blended parameters

Sample Preparation

For GC–FID analysis, the controlled oil was dissolved to a final concentration of 0.4% (w/w) with 2 mL internal standard solution (IS-a) and 8 mL hexane. The well-mixed diluted oil sample was directly injected into the GC system, without any chemical derivation or purification step prior to chromatographic analysis.

For HPLC–RID analysis, the controlled oil was dissolved to a final concentration of 20 mg/mL with 100 μ L internal standard solution (IS-b) and 900 μ L ethyl acetate. The well-mixed diluted oil sample was injected directly into the HPLC system.

GC–FID and HPLC–RID Analysis of TAG Species in Oil Samples

GC–FID analyses of TAG species in oil samples were evaluated following the AOCS official method Ce 5-86 (2009). All separations were performed with an Agilent 7890 gas chromatograph (USA) equipped with a split/splitless injector coupled to a flame ionization detector. A split injection was used with a ratio of 25:1. The samples were introduced using a robotized autosampler module and then a sample volume of 1 μ L was injected into the GC system.

A capillary column coated with 65% diphenyl–35% dimethylpolysiloxane stationary phase (Restek Rtx-65TG; 30 m \times 0.32 mm \times 0.25 μ m; maximum temperature 370 $^{\circ}$ C; Restek Corp., Bellefonte, PA, USA) was used. The GC oven temperature was programmed from 250 to 350 $^{\circ}$ C at 10 $^{\circ}$ C/min, and held for 20 min. The injection port was held at 370 $^{\circ}$ C isothermally. Helium (99.995%) was used as the carrier gas and its flow rate was 1.0 mL/min; FID, 360 $^{\circ}$ C.

HPLC–RID analyses of TAG species in oil samples were evaluated following the AOCS official method Ce 5b-89 (2009). Chromatographic analysis was carried out using a reversed-phase HPLC system (Agilent 1260, USA) equipped with a differential refraction detector coupled to an Hibar RP-18 analytical column (Merck LiChrosorb[®] RP-18, 3 \times 150 mm, 5 μ m). The column temperature was kept at 40 $^{\circ}$ C. The injection volume was kept at 10 μ L. The mobile phase composed of acetonitrile and ethyl acetate (8:2) was used for isocratic separation. The flow rate was 1.8 mL/min. RID conditions were polarity, positive; peak width >0.2 min.

Data Analysis

The TAG profiles from GC–FID and HPLC were handled using Open LAB Data Station and results were calculated using Microsoft Excel 2007. The internal normalization

method was used to complete the quantitative analysis, assuming that the detector response was the same for all compounds. All experiments were performed in triplicate.

The statistical analyses including PCA and MLR were developed by Matlab (Ver. 7.10 (2010a): The MATHWORKS, USA).

PCA is frequently used for the purpose of generating a reduced set of variables which account for the most of the variability in the original data. In this study, PCA created the three-dimensional score plot, which was based on the 11 TAG composition data sets collected by GC–FID, to visualize the differences of those five vegetable oils.

The main steps of the MLR with constrained linear least squares method and exhaustion method for proportions prediction of constitutive components are shown as follows.

Step 1. The database for TAG composition of pure oils was prepared.

Step 2 (Input). Matrix representations of simplex mixture systems were given.

For instance, $q = 3$,

$$T_{i\text{Blendedoil}} = \beta_1 T_{i1} + \beta_2 T_{i2} + \beta_3 T_{i3} \quad (3-1)$$

Some restrained conditions for solving linear program can be written as

$$0 \leq \beta_1, \beta_2, \dots, \beta_j \quad (3-2)$$

where $T_{i\text{Blendedoil}}$, T_{i1} , T_{i2} , T_{i3} are the contents of TAG i in the blended oil, type 1, type 2, and type 3, respectively; and β_1 , β_2 , and β_3 are the proportions of type 1, type 2, and type 3 in blended oil.

Matrices representing mixture systems were analyzed under an exhaustive method. Each of the matrices was carried out by exhaustive sampling of TAG compositions which were shared in the three pure oil groups corresponding to their proportions β_j . The ternary system SBO–MZO–GNO (30%, 50%, 20%) consists of $\beta_{\text{SBO}} = 30\%$, $\beta_{\text{MZO}} = 50\%$, and $\beta_{\text{GNO}} = 20\%$ TAG compositions exhaustively taken from the first (SBO), second (MZO), and third (GNO) oil groups, respectively. In this paper, the database includes five groups, i.e., SBO (147 pure oils), MZO (107 pure oils), GNO (179 pure oils), SFO (65 pure oils), and RSO (116 pure oils). Therefore, N corresponding to the total number of the TAG composition combinations from its constitutive pure oils is $N = C_{147}^1 \times C_{107}^1 \times C_{179}^1 = 2815491$.

Step 3 (Output). Calculation of constitutive component proportions.

The composition of the TAGs (11 TAGs–GC or six TAGs–HPLC) from pure oils and blends was considered as independent variable. According to the example in step 3 (Input), the whole calculation process needs to be executed 2,815,491 times. Each of the matrices/mixtures will be calculated to obtain a set of predicted results/proportions. The standard technique for performing

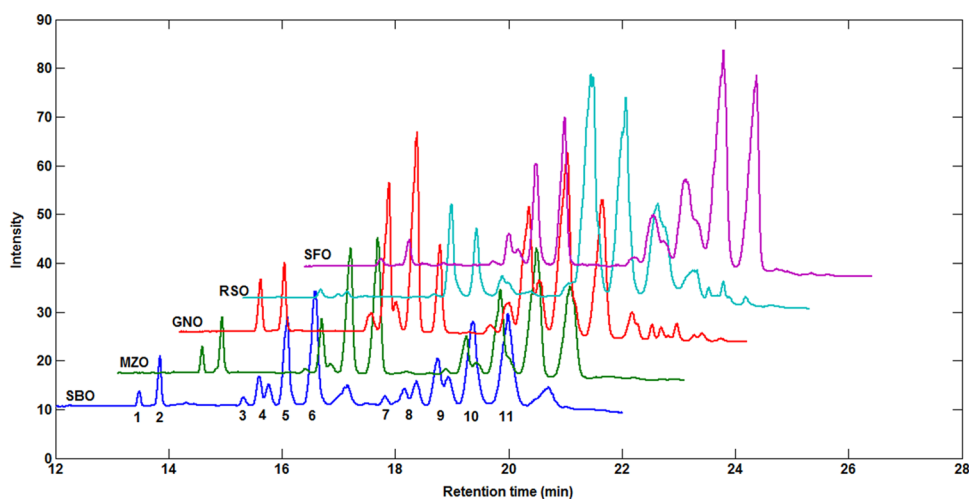
the MLR is least squares regression. However, the results could be unreasonable. In some cases, each composition in the objective function is not absolutely independent, but is restrained by some physical and mathematical conditions. As shown in Eq. 2, Eq. 3-1, and Eq. 3-2, there are three conditions in this model: all proportions should fit in the linear regression; the proportions should be always positive; the sum of the proportions in the blended oil should equal 1. To overcome the drawback, constrained linear least squares (CLLS) is used to resolve the unreasonable fitting results when proportions of constitutive components in blended oil were predicted. The program will provide all the statistical parameters (e.g., minimum, maximum, mean, standard deviation (SD), etc.) as output. Data processing by MLR with constrained linear least squares method and exhaustion method in this study was carried out by a homemade Matlab program.

Results and Discussion

TAG Analysis by GC–FID

Compared to the other TAG analysis methods, GC has features of simple operation, stable running, and does not require a long time to start analysis. Therefore, GC has been applied as a measurement of the quality and the purity of vegetable oils in this paper. TAG analysis of these oils reveals a quite homogeneous composition dominated by C50, C52, and C54 groups which are mainly composed of palmitic (P-16:0), stearic (S-18:0), oleic (O-18:1), linoleic (L-18:2), and linolenic (Ln-18:3) acid residues, and these groups account for 95–98% of the investigated vegetable oils. TAG chromatograms of GC–FID are shown in Fig. 2: C50 group includes POP and PLP, C52 group includes POS, POO + PLS, PLO, and PLL, and C54 group includes SOO, OOO + SLS + SLO, OLO + SLL, OLL, and LLL.

Fig. 2 GC–FID chromatograms of five kinds of vegetable oils (all kinds of oils analyzed in the study). Numbers (1–11) represent 11 typical TAGs: 1 POP, 2 PLP, 3 POS, 4 POO + PLS, 5 PLO, 6 PLL, 7 SOO, 8 OOO + SLS + SLO, 9 OLO + SLL, 10 OLL, 11 LLL



Statistical Analysis

Table 1 shows the relative TAG compositions and the minimum, maximum, and mean values for each TAG of 614 pure vegetable oils by GC–FID. The one-way ANOVA for each TAG of different kinds of oils was applied by using the SPSS software. The results of the ANOVA (the significance levels are less than 10^{-4} , $P = 10^{-4}$) indicated there was a statistically significant difference in the content of TAGs within the different kinds of oils used in this experiment.

Afterwards, the data set of five kinds of pure vegetable oils (614) having different origins and refining conditions was classified by PCA (Fig. 3). The PCA graphic shows the similarities and differences among the pure oils, and similar oils tend to form clusters and dissimilar oils are separated away. As shown in Table 1 and Fig. 3, except for SBO and MZO, the other kinds of pure vegetable oils, which form three clusters, are noticeably separated. The TAG compositions of SBO and MZO are very similar; hence, it can be inaccurate when predicting the proportions of SBO and MZO in their co-existence groups via the developed MLR simulation.

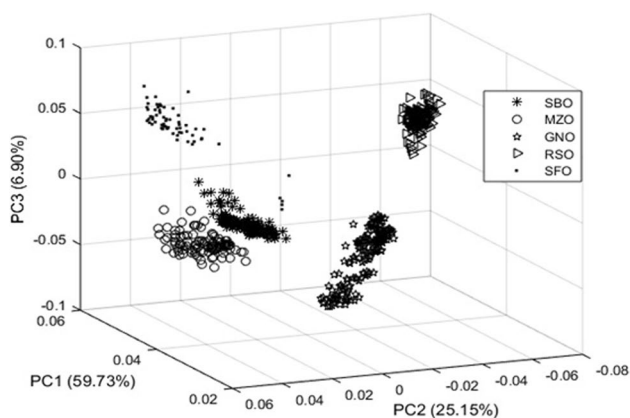
Furthermore, the current analysis condition of GC–FID leads to poor peak shapes when samples contain longer fatty acids and TAGs that have the same carbon number of long chains which cannot be separated individually. Therefore, reversed-phase high-performance liquid chromatography (RP-HPLC) is also employed for the separation of TAGs, because it operates on the principle of both chain length and degree of unsaturation of fatty acids, thus providing a better separation of individual TAG molecules, especially TAGs containing polyunsaturated fatty acids.

TAG Analysis by HPLC–RID

Chromatograms of TAGs in SBO, MZO, GNO, RSO, and SFO using HPLC are shown in Fig. 4. Peak 1 eluted at

Table 1 Eleven TAG compositions (% mass) in five kinds of vegetable oils by GC–FID and ANOVA results

	POP	PLP	POS	POO + PLS	PLO	PLL	SOO	OOO + SLS + SLO	OLO + SLL	OLL	LLL
SBO											
MIN	0.42	1.65	0.47	3.66	7.91	11.41	0.87	5.40	11.74	15.89	17.64
MAX	0.96	2.48	0.98	6.23	10.98	15.73	2.11	11.41	18.81	23.45	30.52
AVG	0.70	2.06	0.74	5.12	9.59	13.69	1.47	8.16	15.41	20.27	21.85
RSD	10.91	7.78	11.23	8.54	7.04	7.67	14.87	11.72	8.39	7.31	10.99
MZO											
MIN	0.81	2.08	0.27	4.38	11.75	12.88	0.53	6.18	15.05	23.86	17.29
MAX	1.70	2.96	0.51	7.05	16.26	18.05	1.03	10.93	22.65	34.45	26.67
AVG	1.10	2.55	0.38	5.51	13.59	15.53	0.77	8.24	18.61	28.97	21.61
RSD	13.18	7.94	13.84	10.92	8.86	6.79	12.71	9.50	9.66	8.54	8.33
GNO											
MIN	1.11	1.45	0.80	8.69	10.16	4.64	2.78	17.59	20.14	13.34	4.61
MAX	1.77	2.82	1.40	12.48	16.11	8.24	6.27	39.82	31.05	22.95	8.73
AVG	1.31	1.82	1.03	10.65	12.63	5.73	4.57	29.21	24.51	18.08	6.23
RSD	7.32	13.93	12.57	7.72	10.83	10.28	20.28	19.40	8.67	10.47	13.43
RSO											
MIN	0.20	0.20	0.17	5.57	3.88	1.82	2.17	35.56	23.90	14.09	3.88
MAX	0.42	0.48	0.31	7.05	5.18	3.33	4.11	51.24	32.89	19.15	6.62
AVG	0.26	0.31	0.23	6.25	4.48	2.37	2.91	43.99	28.80	16.57	5.03
RSD	11.39	16.11	9.79	4.43	5.54	11.58	12.76	6.47	5.13	5.66	10.56
SFO											
MIN	0.15	0.68	0.16	2.71	5.18	7.87	0.82	7.26	13.45	20.66	15.99
MAX	0.31	1.44	0.36	3.81	8.40	13.84	1.93	14.68	24.50	38.99	37.02
AVG	0.25	0.95	0.28	3.33	7.20	10.87	1.16	10.32	20.40	33.75	29.18
RSD	15.22	13.67	13.84	8.40	9.47	10.63	13.76	12.53	11.18	12.35	14.75
<i>F</i>	1930.25	1577.17	1553.12	1804.15	1101.73	3717.86	992.91	2061.05	635.07	735.96	2257.36
<i>P</i>	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴	<10 ⁻⁴

**Fig. 3** Principal components analysis (PCA) plot for TAGs of the overall vegetable oils

12.13 min and only appears in SBO. Furthermore, there is a great difference in the content of peak 2 between SBO and MZO. Peak 1 and peak 2 were identified to correspond to 3-linolenoyl-1,2-dilinoleoyl-*sn*-glycerol (LLL_n) and trilinolein (LLL) by LC–MS/MS, respectively. These

MS results (protonated molecule $[M + H]^+$ and “diacylglycerol” ions, $[DG]^+$) were in accordance with the report published by Neff and Byrdwell in 1995 [47]. Table 2 shows the relative TAG compositions and the minimum, maximum, and mean values for each TAG of 614 pure vegetable oils by HPLC–RID. These TAGs could be new noticeable markers to determine the proportions of the constitutive components in blended oil which cannot be distinguished by the GC. Thus, all the SBO and MZO were re-analyzed by HPLC–RID to find better independent variables and to offer more accurate proportion results.

Development of a MLR-based Method for Determination of the Proportions of the Constitutive Components in Blended Oils

Mass balance is a universal principle. TAG compositions of pure oils and a blended oil were considered as the independent variables and proportions of constitutive components were defined as the dependent variables. MLR was used to determine the relationship between multiple

Fig. 4 HPLC-RID chromatograms of five kinds of vegetable oils (all kinds of oils analyzed in the study). Numbers (1–6) represent six typical TAGs: 1' LLLn, 2' LLL, 3' LLO, 4' LLP, 5' OOL, 6' POL

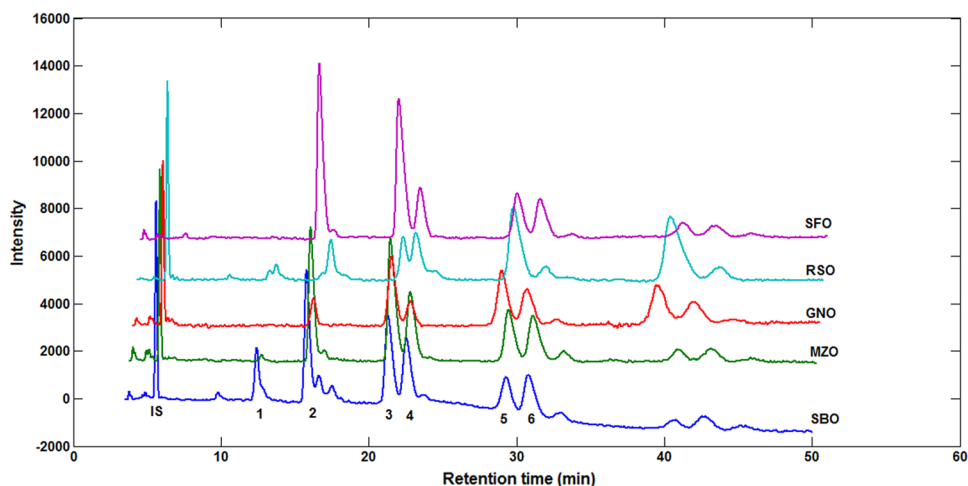


Table 2 Six TAG compositions (% mass) in five kinds of vegetable oils by HPLC-RID and ANOVA results

	LLLn	LLL	LLO	LLP	OOL	POL
SBO						
MIN	6.01	21.68	13.01	11.62	6.49	8.51
MAX	8.05	24.92	17.42	14.51	9.23	12.59
AVG	6.64	23.81	15.35	13.29	8.18	10.43
RSD	7.64	7.03	7.22	7.25	8.00	7.87
MZO						
MIN	0.54	17.80	21.43	13.32	9.65	10.71
MAX	1.08	21.05	23.17	16.12	13.20	13.02
AVG	0.73	18.94	22.32	14.73	12.01	12.19
RSD	10.02	7.87	7.85	8.08	7.45	7.98
GNO						
MIN	0	3.98	14.54	4.79	17.74	13.47
MAX	0	2.93	17.96	6.15	20.77	15.39
AVG	0	3.37	16.78	5.71	18.51	14.04
RSD	0	8.11	7.04	7.25	6.99	7.15
RSO						
MIN	3.50	7.91	6.58	7.78	17.97	0
MAX	3.90	5.34	7.54	13.89	19.64	0
AVG	3.40	6.34	7.03	8.48	18.65	0
RSD	8.09	8.08	7.98	8.35	7.00	0
SFO						
MIN	0	21.46	21.83	9.41	7.23	10.41
MAX	0	30.09	29.82	12.49	12.35	12.46
AVG	0	26.71	27.21	11.03	10.90	11.43
RSD	0	7.23	6.05	6.01	8.22	8.12
<i>F</i>	540.97	347.80	155.53	81.52	130.28	249.00
<i>P</i> value	<10 ⁻³	<10 ⁻³	<10 ⁻³	<10 ⁻³	<10 ⁻³	<10 ⁻³

independent predictor variables (the TAGs composition of pure oils and a blended oil) and dependent variables (the proportions of constitutive components). The details of such simulation have been shown in Sect. “Data Analysis”.

TAG compositions of GC as independent variables include 1, POP (dipalmitoyl oleoyl glycerol); 2, PLP (dipalmitoyl linoleoyl glycerol); 3, POS (palmitoyl oleoyl stearoyl glycerol); 4, POO + PLS (dioleoyl palmitoyl glycerol + palmitoyl linoleoyl stearoyl glycerol); 5, PLO (palmitoyl linoleoyl oleoyl glycerol); 6, PLL (dilinoleoyl palmitoyl glycerol); 7, SOO (dioleoyl stearoyl glycerol); 8, OOO + SLS + SLO (triolein + distearoyl linoleoyl glycerol + stearoyl linoleoyl oleoyl glycerol); 9, OLO + SLL (dioleoyl linoleoyl glycerol) + dilinoleoyl stearoyl glycerol); 10, OLL (dilinoleoyl oleoyl glycerol); and 11, LLL (trilinolein).

In order to improve the prediction accuracy, the binary and ternary blended oils with SBO and MZO co-occurring were analyzed and processed by MLR twice, and the independent variables of MLR were from GC–FID and HPLC–RID, respectively.

TAG compositions of HPLC as independent variables include 1', LLLn (3-linolenoyl-1,2-dilinoleoyl-*sn*-glycerol); 2', LLL (trilinolein); 3', LLO (1, 2-dilinoleoyl-3-oleoyl-*rac*-glycerol); 4', LLP (dilinoleoyl palmitoyl glycerol); 5', OOL (dioleoyl linoleoyl glycerol); and 6', POL (palmitoyl linoleoyl oleoyl glycerol).

The calculation steps to analyze SBO and MZO co-existence in blended oils are as follows: firstly, the sum of β_{SBO} and β_{MZO} and the proportions of other oils are determined using MLR with the TAG variables from GC–FID. Secondly, β_{SBO} and β_{MZO} are further predicted using MLR with the TAG variables from HPLC–RID.

Prediction and Evaluation of the MLR Method in the Real Blended Oils of Binary and Ternary Systems

Considering the extensive sources of collected oils, exhaustive algorithm was applied to calculate the proportions of all the possible combinations. The program provided all the statistical parameters (e.g., minimum, maximum, mean, standard deviation (SD), etc.) as output. The accuracy of these values is not constant for all blended oils as it depends on the type and the proportion of the constitutive components in blended oils. The mean value was used as the parameter to calculate the relative error in the statistical analysis.

The relative error e_{β_j} of β_j was calculated with the following formula:

$$e_{\beta_j} = \left| \beta_{j,actual} - \beta_{j,pred} \right| / \beta_{j,actual} \times 100\% \quad (4)$$

$$e_{\beta} = \max_j e_{\beta_j} \quad (5)$$

where e_{β} is the $\max_j e_{\beta_j}$ in a blended oil, also regarded as the relative error of each oil and illustrated in Fig. 5, $\beta_{j,actual}$ denotes the actual additive proportion of co-occurring components when preparing a lab-made blended oil, and $\beta_{j,pred}$ denotes the predicted additive proportion of co-occurring components in a blended oil.

To analyze the relative errors of each of the five oils (SBO, MZO, GNO, RSO, SFO) in related groups in the binary system and the ternary system, another approach named “pinwheel and chessboard” was applied.

For binary systems, the center of the pinwheel represents a kind of pure oil. For instance, A1 in Fig. 5 represents SBO, where the 4 axes represent the other oils (MZO, GNO, RSO, SFO), and the proportion of SBO decreases gradually while the proportions of other oils increase as the axis extends farther out from the center. The extent and tendency of the deviation can be observed from colors, lengths, and locations of the bars, i.e., the white representing e_{β_j} is less than 20%, the gray representing e_{β_j} is from 20% to 30%, and the black representing e_{β_j} is more than 30%. The formula to calculate e_{β_j} is shown as Eq. 4.

For ternary systems, the center of the chessboard represents a kind of pure oil. For instance, B1 in Fig. 5 represents SBO, where each triangle in the chessboard represents a group which contains SBO, and the other two triangle vertexes represent the other two oils. Similar to the binary system, the proportion of SBO decreases gradually while the proportions of the other oils increase as vertexes extend outward from the center. The extent and tendency of the deviation can be observed from the colors and locations of the points, same as in the binary system.

$$P_{N_n} = N_{e_{\beta}} / N_n \quad (6)$$

In Eq. 6, $N_{e_{\beta}}$ is the number of e_{β} (considering only the maximum of e_{β_j} in a blended oil sample) within a given range, N_n is the number of samples in the sample collection n , and P_{N_n} is defined as the percentage of the relative errors within a given range in a sample collection, and is regarded as an evaluation indicator for the samples. N_1 and N_5 are for the statistical results of binary system and ternary system based on GC–FID variables; N_2 – N_4 and N_6 – N_7 are for the statistical results of binary system and ternary system based on GC–FID variables combined with HPLC variables. The results are shown in Table 3. Regarding a given range, P_{N_n} of the method with GC–FID variables and HPLC variables ($P_{N_4} = 98.2\%$ and $P_{N_7} = 84.7\%$, $e_{\beta_j} < 20\%$) is always larger than P_{N_n} of the method with GC–FID variables ($P_{N_1} = 61.3\%$ and $P_{N_5} = 48.3\%$, $e_{\beta} < 20\%$). In addition, the proposed method

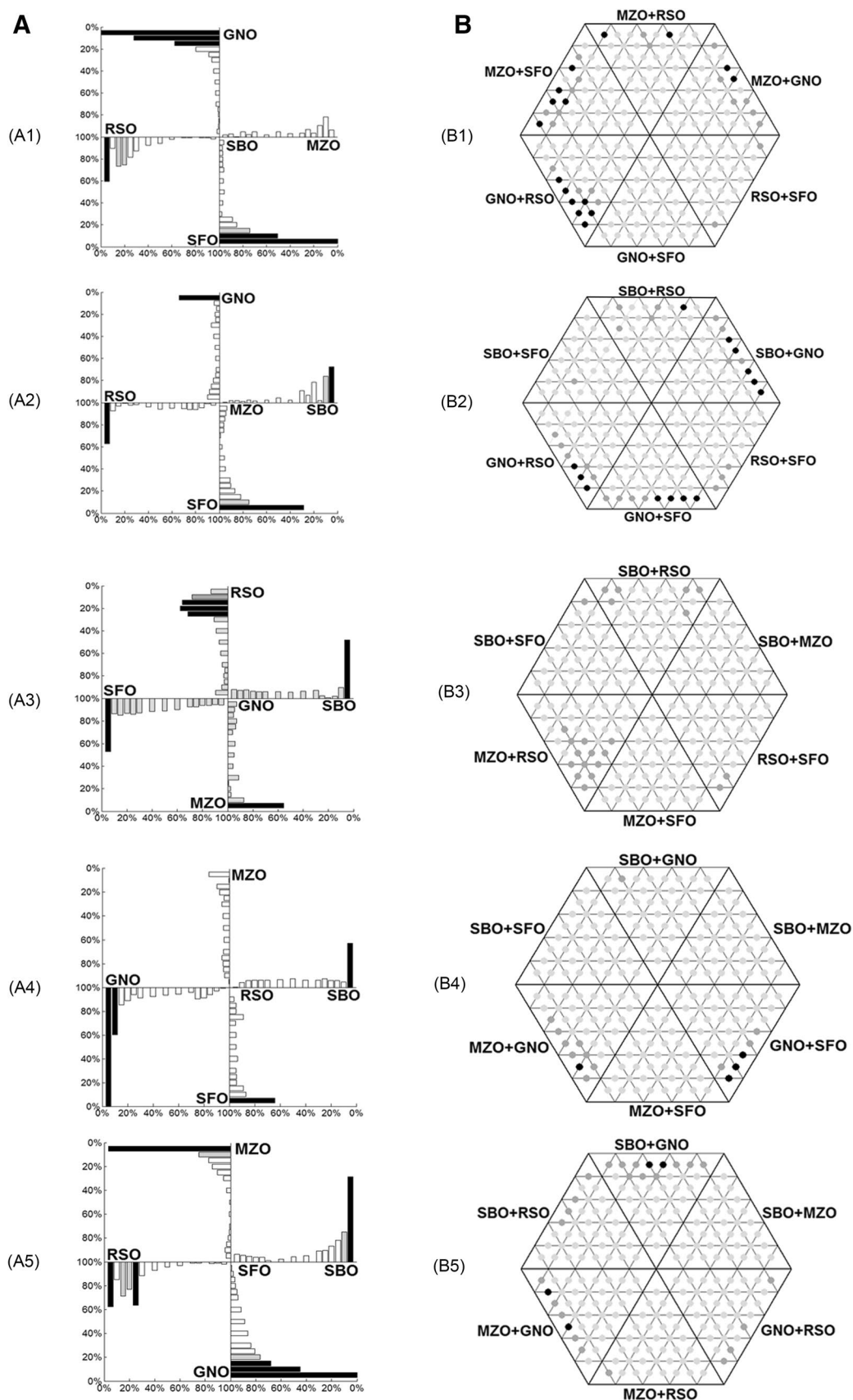


Fig. 5 Relative errors, e_{β_j} , of each kind of the five oils in related groups of the binary and the ternary system

Table 3 Statistical results of the percentage of the relative errors in different ways for the binary system and the ternary system

	Bin-sys ($n = 1$)	Bin-sys' ($n = 2$)	Bin-sys' ($n = 3$)	Bin-sys ($n = 4$)	Ter-sys ($n = 5$)	Ter-sys' ($n = 6$)	Ter-sys' ($n = 7$)
	$N_1 = 150$	$N_2 = 150$	$N_3 = 130,$ $\beta_j > 10\%$	$N_4 = 110,$ $\beta_j > 20\%$	$N_5 = 360$	$N_6 = 360$	$N_7 = 150$ $\beta_j > 10\%$
$N_n (e_{\beta_j} < 20\%)$	92	112	118	108	174	242	127
$P_{N_n} (e_{\beta_j} < 20\%)$	61.3%	74.7%	90.8%	98.2%	48.3%	67.2%	84.7%
$N_n (20\% < e_{\beta_j} < 30\%)$	15	11	6	0	58	53	18
$P_{N_n} (20\% < e_{\beta_j} < 30\%)$	10.0%	7.3%	4.6%	0%	16.2%	14.7%	12%
$N_n (e_{\beta_j} > 30\%)$	43	27	6	2	128	65	5
$P_{N_n} (e_{\beta_j} > 30\%)$	28.7%	18%	4.6%	2.8%	35.5%	18.1%	3.3%

P_{N_n} is defined as the percentage of the relative errors falling in a given range in a sample collection. N_1 and N_5 are for the statistical results of binary system and ternary system based on GC–FID variables, N_2 – N_4 and N_6 – N_7 are for the statistical results of binary system and ternary system based on GC–FID variables combined with HPLC variables. β_j stands for adding proportion, while e_{β_j} stands for relative error

requires a minimum proportion of each oil, i.e., at least 10% for MZO, 15% for RSO, 20% for SBO, and 30% for GNO and SFO in a binary blend, and at least 10% for GNO, 20% for MZO, RSO, and SFO, and 30% for SBO in a ternary blend, respectively ($e_{\beta_j} < 30\%$). It is interesting to note that the larger proportion β_j (Eq. 2) is added in blended oils, the lower e_{β_j} (Eq. 4) is obtained in the binary/ternary system, i.e., the predicted results become more accurate in the binary/ternary system. As shown in the Fig. 5, there are almost all white dots/bars near the centers of the pinwheels and the chessboards. The proportions β_j are getting smaller as they move away from the centers, and the further from the centers, the more possibilities of getting high e_{β_j} since β_j is the denominator in Eq. 4. In other words, there is a negative correlation between β_j and e_{β_j} (e.g., given that prediction errors are 10%, e_{β_j} is 50% when β_j is 20%, while e_{β_j} is 12.5% when β_j is 80%); therefore it is better to take β_j into consideration when the standards were set up for e_{β_j} in the process of a practical application. Furthermore, this mathematical model can be extended to determine the proportions of the constitutive components in other kinds of blended oils.

Conclusions

In view of the results attained within this study, it can be concluded that the proposed method of MLR with constrained linear least squares and exhaustion calculation based on the combination of TAG values from GC and HPLC analysis is a promising and reasonable way to verify the proportion of constitutive components in some blended vegetable oils. To verify the effectiveness of the experimental protocols and the models, it needs to be validated in more blended oils and in a larger database with a

greater diversity and quantity of pure oils. Once approved, it can be applied for quick segregation and quality control in routine analysis to help regulate the blended oil market. Additionally, more specific and stable markers will be further studied to enhance the accuracy of the predicted proportions of constitutive components in the edible blended oils.

Acknowledgements We would like to thank the Wilmar (Shanghai) Biotechnology Research & Development Center Co., Ltd. for their support of this research work.

References

- Chen CW, Chong CL, Ghazali HM, Lai OM (2007) Interpretation of triacylglycerol profiles of palm oil, palm kernel oil and their binary blends. *Food Chem* 100:178–191
- Choe E, Min DB (2007) Chemistry of deep-fat frying oils. *J Food Sci* 72(5):77–86
- Nasirullah, Ankaiah KN, Krishnamurthy MN, Nagaraja KV (1991) Quality characteristics of edible vegetable oil blends. *J Am Oil Chem Soc* 68(6):446–447
- Park YW, Chang PS, Lee JH (2010) Application of triacylglycerol and fatty acid analyses to discriminate blended sesame oil with soybean oil. *Food Chem* 123:377–383
- Zhou ZY (2012) Study on the strategy of the vegetable oil industry in China. Doctoral dissertation, Chinese Academy of Agricultural Sciences
- Hai Z, Wang J (2006) Electronic nose and data analysis for detection of maize oil adulteration in sesame oil. *Sens Actuators B Chem* 119(2):449–455
- Lee DS, Lee ES, Kim HJ, Kim SO, Kim K (2001) Reversed phase liquid chromatographic determination of triacylglycerol composition in sesame oils and the chemometric detection of adulteration. *Anal Chim Acta* 429:321–330
- Ramón A, Ramón AR (2000) Authentication of vegetable oils by chromatographic techniques. *J Chromatogr A* 881:93–104
- Zabaras D, Gordon MH (2004) Detection of pressed hazelnut oil in virgin olive oil by analysis of polar components: improvement and validation of the method. *Food Chem* 84(3):475–483

10. Nunes CA (2014) Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. *Food Res Int* 60:255–261
11. Rohman A, Che Man YB (2014) The use of FTIR spectroscopy and chemometrics for rapid authentication of extra virgin olive oil. *J Am Oil Chem Soc* 91(2):207–213
12. Oussama A, Elabadi F, Platikanov S, Kzaiber F, Tauler R (2012) Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores. *J Am Oil Chem Soc* 89:1807–1812
13. Zhang XF, Zou MQ, Qi XH, Liu F, Zhang C, Yin F (2011) Quantitative detection of adulterated olive oil by Raman spectroscopy and chemometrics. *J Raman Spectrosc* 42:1784–1788
14. de la Mata-Espinosa P, Bosque-Sendra JM, Bro R, Cuadros-Rodríguez L (2011) Olive oil quantification of edible vegetable oil blends using triacylglycerols chromatographic fingerprints and chemometric tools. *Talanta* 85:177–182
15. Ruiz-Samblás C, Federico Marini F, Cuadros-Rodríguez L, González-Casado A (2012) Quantification of blending of olive oils and edible vegetable oils by triacylglycerol fingerprint gas chromatography and chemometric tools. *J Chromatogr B* 910:71–77
16. Yajun W, Ying C (2008) Detection of olive oil using the Evagreen real-time PCR method. *Eur Food Res Technol* 227:1117–1124
17. Bosque-Sendra JM, Cuadros-Rodríguez L, Ruiz-Samblás C, de la Mata AP (2012) Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data—a review. *Anal Chim Acta* 724:1–11
18. Szostak R, Mazurek S (2004) FT-Raman quantitative determination of ambroxol in tablets. *J Mol Struct* 704:229–233
19. Johansson J, Sparen A, Svensson O, Folestad S, Claybourn M (2007) Quantitative transmission Raman spectroscopy of pharmaceutical tablets and capsule. *Appl Spectrosc* 61(11):1211–1218
20. Szostak R, Mazurek S (2002) Quantitative determination of acetylsalicylic acid and acetaminophen in tablets by FT-Raman spectroscopy. *Analyst* 127(1):144–148
21. Mazurek S, Szostak R (2009) Quantification of atorvas-tatin calcium in tablets by FT-Raman spectroscopy. *J Pharm Biomed Anal* 49(1):168–172
22. Mazurek S, Szostak R (2006) Quantitative determination of captopril and prednisolone in tablets by FT-Raman spectroscopy. *J Pharm Biomed Anal* 40(5):1225–1230
23. Mazurek S, Szostak R (2008) Quantitative determination of diclofenac sodium in solid dosage forms by FT-Raman spectroscopy. *J Pharm Biomed Anal* 48(3):814–821
24. EC (2012) Commission Regulation (EC) No. 29/2012, on marketing standards for olive oil. *Off J Eur Commun* 55:L12
25. Lohumi S, Lee S, Lee H, Cho BK (2015) A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends Food Sci Tech* 46:85–98
26. Nurrulhidayah AF, Rohmanb A, Amina I, Shuhaimia M, Khatibd A (2013) Analysis of chicken fat as adulterant in butter using Fourier transform infrared spectroscopy and chemometrics. *Grasas Aceites* 64:349–355
27. Deng DH, Xu L, Ye ZH, Cui HF, Cai CB, Yu XP (2012) FTIR spectroscopy and chemometric class modeling techniques for authentication of Chinese sesame oil. *J Am Oil Chem Soc* 89:1003–1009
28. Javidnia K, Parish M, Karimi S, Hemmateenejad B (2013) Discrimination of edible oils and fats by combination of multivariate pattern recognition and FT-IR spectroscopy: a comparative study between different modeling methods. *Spectrochim Acta A* 104:175–181
29. Marigheto NA, Kemsley EK, Defernez M, Wilson RH (1998) A comparison of mid-infrared and Raman spectroscopies for the authentication of edible oils. *J Am Oil Chem Soc* 75:987–992
30. Rohman A, Man YB (2010) FTIR spectroscopy combined with chemometrics for analysis of lard in the mixtures with body fats of lamb, cow, and chicken. *Food Res Int* 17:519–526
31. Rohman A, Sunarminingsih R, Man YB (2012) The employment of FTIR spectroscopy and chemometrics for classification and quantification of mutton fat in cod liver oil. *Am J Food Technol* 7:151–159
32. Rohman A, Man YB (2009) Analysis of cod-liver oil adulteration using Fourier transform infrared (FTIR) spectroscopy. *J Am Oil Chem Soc* 86:1149–1153
33. Oussama A, Elabadi F, Platikanov S, Kzaiber F, Tauler R (2012) Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores. *J Am Oil Chem Soc* 89:1807–1812
34. Özdemir D, Öztürk B (2007) Near infrared spectroscopic determination of olive oil adulteration with sunflower and corn oil. *J Food Drug Anal* 15:40–47
35. Xu J, Liu XF, Wang YT (2016) A detection method of vegetable oils in edible blended oil based on three-dimensional fluorescence spectroscopy technique. *Food Chem* 212:72–77
36. Choudhary M, Grover K, Kaur G (2015) Development of rice bran oil blends for quality improvement. *Food Chem* 173(15):770–777
37. De Leonardis M (2012) Heat-oxidation stability of palm oil blended with extra virgin olive oil. *Food Chem* 135(3):1769–1776
38. Cubero-Leon E, Peñalver R, Maquet A (2014) Review on metabolomics for food authentication. *Food Res Int* 60:95–107
39. Park JR, Lee DS (2003) Detection of adulteration in olive oils using triacylglycerols compositions by high temperature gas chromatography. *Bull Korean Chem Soc* 24(4):527–530
40. Christopoulou E, Lazarakia M, Komaitis M, Kaselimis B K (2004) Effectiveness of determinations of fatty acids and triglycerides for the detection of adulteration of olive oils with vegetable oils. *Food Chem* 84:463–474
41. Salghi R, Armbruster W, Schwack W (2014) Detection of argan oil adulteration with vegetable oils by high-performance liquid chromatography–evaporative light scattering detection. *Food Chem* 153:387–392
42. Salivaras E, McCurdy AR (1992) Detection of olive oil adulteration with canola oil from triacylglycerol analysis by reversed-phase high-performance liquid chromatography. *J Am Oil Chem Soc* 69(9):935–938
43. Cunha SC, Oliveira MBPP (2006) Discrimination of vegetable oils by triacylglycerols evaluation of profile using HPLC/ELSD. *Food Chem* 95:518–524
44. Funada Y, Hirata Y (1999) Development of a simulation program for the analysis of oils and fats by subcritical fluid chromatography. *Anal Chim Acta* 401:73–82
45. Semmar N, Artaud J (2015) A new simplex-based approach predicting olive oil blend compositions from fatty acid data. *J Food Compos Anal* 43:149–159
46. Rebecchi SR, Vélez MA, Vaira S, Perotti MC (2016) Adulteration of Argentinean milk fats with animal fats: detection by fatty acids analysis and multivariate regression techniques. *Food Chem* 192:1025–1032
47. Neff WC, Byrdwell WC (1995) Soybean oil triacylglycerol analysis by reversed-phase high-performance liquid chromatography coupled with atmospheric pressure chemical ionization mass spectrometry. *J Am Oil Chem Soc* 72:1185–1191