Chemometrics and Intelligent Laboratory Systems 169 (2017) 94–99

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Targeted multivariate adulteration detection based on fatty acid profiles and Monte Carlo one-class partial least squares

Liangxiao Zhang [a,b,d,f,*], Zhe Yuan [a,b], Peiwu Li [a,c,d,e,**], Xuefang Wang [a,e], Jin Mao [a,d,e], Qi Zhang [a,c,e], Chundi Hu [g]

[a] Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China
[b] Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430062, China
[c] Key Laboratory of Detection for Mycotoxins, Ministry of Agriculture, Wuhan 430062, China
[d] Laboratory of Risk Assessment for Oilseed Products (Wuhan), Ministry of Agriculture, Wuhan 430062, China
[e] Quality Inspection and Test Center for Oilseed Products, Ministry of Agriculture, Wuhan 430062, China
[f] Hubei Collaborative Innovation Center for Green Transformation of Bio-Resources, Wuhan 430062, China
[g] Pharmaceutical College of Hubei University of Science and Technology, Xianning 437100, China

ARTICLE INFO

Keywords:
Monte Carlo one-class partial least squares (MCOCPLS)
Targeted multivariate adulteration detection
Virgin olive oil (VOO)
Chemometrics
Fatty acid profiles

ABSTRACT

To develop effective adulteration detection methods is essential as food quality and safety draw particular concern all over the world. In this study, Monte Carlo one-class partial least squares (MCOCPLS) was proposed and employed as a novel one class classification model for authentication identification by using virgin olive oil (VOO) as an example. Monte Carlo sampling was proposed for selecting variable subspace to improve the performance of one-class partial least squares (OCPLS) classifier. MCOCPLS was used to establish a one-class model, the performance of which was validated by an independent test set consisting of 5000 adulterated oils simulated by the Monte Carlo method. The prediction for the best model of MCOCPLS reaches a correct rate of 99.10%. Moreover, authentic VOOs were analyzed and assessed for the adulteration risk. In conclusion, the proposed MCOCPLS method could be used to effectively detect olive oils adulterated with other vegetable oils at a concentration of as low as 3%. Therefore, MCOCPLS provides an effective tool and new insights in adulteration detection for edible oils and other foods.

Virgin olive oil (VOO) is an important daily dietary supplement with beneficial effects on human health because it is a source of healthy unsaturated fatty acids and hundreds of micronutrients, including phenol compounds, vitamin E and carotenes [1]. However, according to the Database of Food Ingredient Fraud and Economically Motivated Adulteration (EMA), olive oil is reported as the most common target for adulteration in scholarly journals [2], making its adulteration detection an important issue. Recently, the following four strategies have been proposed to develop an adulteration detector for VOO: (a) detection of DNA-based genetic markers of adulterants [3,4]; (b) detection of the marker metabolite of adulterants [5]; (c) fast identification based on spectroscopy or sensor, including nuclear magnetic resonance (NMR) [6], infrared spectroscopy [7–9], Raman spectroscopy [10]; and (d) detection based on metabolomics including triacylglycerol (TAG), fatty acid [11,12] and pytosterol [13] profiles. With the explicit chemical properties and advantage of multivariate analysis, the authentication identification methods based on metabolomics were the most promising.

Chemometrics plays an important role in classifying pure and adulterated edible oils based on multivariate analysis [14]. Recently, chemometric methods such as self-organizing maps based on chaotic parameters, cluster discriminant analysis (CDA), support vector machine (SVM) and random forests (RF) were used to distinguish edible oils from refined recycled cooking oils [15–17], identify edible oils from different regions, and detect adulteration of extra VOO with inferior edible oils [18], respectively. Generally, existing adulteration detection methods were taken as a two- or multi-class classification model to discriminate pure edible oils from adulterated oils with one or more known oils. Since the adulterants in edible oils are usually unknown, the authentication identification generally requires a one-class classification technique in chemometrics [19]. Therefore, an effective authentication identification

* Corresponding author. Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China.
** Corresponding author. Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China.
E-mail addresses: liangxiao_zhang@hotmail.com (L. Zhang), peiwuli@oilcrops.cn (P. Li).

http://dx.doi.org/10.1016/j.chemolab.2017.09.002
Received 31 May 2017; Received in revised form 3 September 2017; Accepted 5 September 2017
Available online 8 September 2017

method involves the following key points: (a) searching for stable components in all samples of one kind of edible oil; (b) an effective one-class classification method. In this study, Monte Carlo one-class partial least squares (MCOCPLS) was proposed and employed to establish a better one-class classification model for identifying the authenticity of VOO. In the MCOCPLS method, Monte Carlo sampling for variables was used to select better variables, and simulated adulterated VOOs were used to illustrate the use of this method.

1. Methods

1.1. Materials and reagents

To ensure that oil samples could represent the actual status of edible oils, 21 VOO samples were purchased from the supermarkets. 20 multivariate adulterated olive oils were designed with soybean oil, rapeseed oil and maize oil added to different olive oils by 3% (w/w) at the ratios of 1:1:0, 1:0:1, 0:1:1 and 1:1:1 (w/w/w), respectively. Supelco 37 component FAME mix (No. 47885-U) was purchased from Sigma (St. Louis, MO, USA), and 11-octadecenoic acid (C18:1n-7, >97.0 purity) and 7-hexadecenoic acid methyl ester were also obtained from Sigma (St. Louis, MO, USA).

1.2. Experimental derivatization procedure

As described in the previous study [20–22], about 0.06 g of vegetable oil sample was diluted with 2 mL solvent of diethyl ether, and petroleum ether (v/v 1:1) and 1 mL 0.4 M KOH-CH_3OH was added, vortexed for 30s and placed at room temperature for 2.5 h. Then, 2 mL redistilled water was added, vortexed and centrifuged at 4500 rpm for 2 min. A volume of 200 μL organic phase was collected and diluted by 800 μL petroleum ether prior to GC-MS analysis.

1.3. Fatty acid analysis

According to the previous study [20–22], the analyses were performed by Agilent GC-7890 gas chromatograph interfaced to an Agilent 5973 mass spectrometer. In the gas chromatography system, a fused silica capillary column DB-23 (30 m × 0.25 mm; i. d. 0.15 μm film) (Agilent Technologies) was used. Helium (99.999% purity) was used as carrier gas at a flow-rate of 1.2 mL min^{-1}. The column was first set at 100 °C and held for 0.2 min, and the temperature was subsequently increased to 215 °C at the rate of 10 °C/min and held for 0.1 min, which was finally increased to 224 °C at the rate of 2 °C/min and held for additional 0.2 min (total program time, 16.5 min). This is the optimum temperature programming for both the separation effect and run time. Mass spectrometric conditions were as follows: ionization mode: electron ionization (EI); electron energy: 70 eV; temperatures of the injector, ion-source and detector at 220, 250 and 150 °C, respectively. The solvent cut time was 3 min; splitting ratio: 20:1; selected ion monitoring (SIM) mode: *m/z* 55, 67, 74 and 79.

The identification of fatty acids in SIM mode was conducted according to the protocol in our previous study [22]. The percentage composition (percentage of peak areas) of fatty acids was employed as the quantitative results for edible oils.

1.4. Multivariate analysis

The one-class partial least squares (OCPLS) classifier was proposed on the basis of partial least squares (PLS) using a distance-based sample density measurement as the response variable [14,23,24]. Initially, a PLS model was developed using analytical data and response vector with all elements being 1. In addition, the PLS residuals were used to compute the critical value of Q statistics. The predicted residual sum of squares (PRESS) obtained by Monte Carlo cross validation (MCCV) or leave-one-out cross validation (LOOCV) can be used to estimate the number of significant LVs. Two parameters can be calculated for the OCPLS model including the score distance (SD) and upper confidence limit (UCL). The SD of an object in the space spanned by the primary OCPLS components and prediction residuals of response variable 1. The UCL of SD based on the first K LVs is calculated using the F-distribution. Finally, the SD and centered model residual (ACR) were plotted in OCLPS outlier diagnosis to screen the outliers.

In this study, Monte Carlo sampling for variable subspace was combined with OCPLS to select a better one-class classification model for VOOs. MCOCPLS mainly works in three steps: (1) randomly drawing the subsets of a fixed number of variables by sampling without replacement; (2) building an OCPLS model for each sub-dataset; (3) selecting the model with the highest correct classification rate as the final model.

In the previous study [20], chemical databases of the fatty acid compositions of 25 soybean oils, 81 peanut oils, 62 sunflower seed oils, 63 rapeseed oils, 80 sesame oils, 42 flaxseed oils, 16 grape seed oils, 11 almond oils, 7 cottonseed oils and 14 corn oils were established. According to the previous studies [20,25,26], 10,000 adulterated oils were simulated by adding these 10 kinds of vegetable oils into authentic olive oils at different proportions. For each adulterated oil, (1) random mixing proportions for 10 kinds of edible oils are created and then the sum of these proportions is normalized to the adulteration level; (2) one kind of authentic edible oil (one olive oil in this study) is randomly picked out; (3) the fatty acid composition is calculated by weighted sum of fatty acid compositions of the selected 10 other oils and 1 olive oil; (4) steps (1)–(3) are repeated for 10,000 times for the adulteration level from 1% to 5%. For example, we could obtain an adulterated olive oil by adding 3% of 10 kinds of edible oils at random mixing proportions to 97% olive oil. 10,000 adulterated oils were randomly divided into training set (5000) and test set (5000).

The data matrix includes the relative contents of fatty acids in edible oils. Since the chemical properties of fatty acids are relatively stable, the fatty acid composition of the blended oil is equal to the weight summary of the fatty acid compositions of individual oils. To establish a robust adulteration detection model for VOOs, adulterated oils were simulated by adulterating with low proportions of one or more vegetable oils using the Monte Carlo method.

The programs of MCOCPLS and Monte Carlo simulation of the adulterated oils were coded in Matlab 2011a for windows (The Mathworks, Natick, MA). Matlab codes of OCPLS were kindly provided by Dr. Xu [27].

2. Results and discussion

2.1. MCOCPLS for authentication identification

In this study, the fatty acid profiles of VOOs were obtained by GC/MS in SIM mode. By combining the mass spectrometric characteristics and equivalent chain length (ECL), 28 fatty acids were identified. The fatty acid profiles of vegetable oils are described by their percentage contents for subsequent multivariate analysis (see Table 1). Since the chemical properties of the fatty acids are relatively stable, the fatty acid composition of the blended oils is equal to the weight summary of the fatty acid compositions of individual oils. To establish a robust adulteration detection model for VOOs, 5000 adulterated oils in training set were simulated by the Monte Carlo method as the pure oils were adulterated with 10 vegetable oils at different proportions and the adulterated oils were employed to train the MCOCPLS model for 21 VOOs, while the other 5000 adulterated oils were used to test this model. The fatty acid compositions of VOOs were shown in Table 1, and the score plots of Principal Component Analysis (PCA) for pure and simulated adulterated VOOs were illustrated in Fig. 1. From Fig. 1, the adulterated VOOs overlapped with the pure VOOs.

In OCPLS, the variables and their combinations influence the prediction performance of the models. However, the number of variable combinations is tremendous. Therefore, Monte Carlo sampling was