

FTIR Spectroscopy and Chemometric Class Modeling Techniques for Authentication of Chinese Sesame Oil

De-Hua Deng · Lu Xu · Zi-Hong Ye · Hai-Feng Cui ·
Chen-Bo Cai · Xiao-Ping Yu

Received: 9 October 2011 / Revised: 18 November 2011 / Accepted: 22 December 2011 / Published online: 13 January 2012
© AOCS 2012

Abstract This investigation was aimed at developing a rapid analysis method for authentication of Chinese sesame oils by FTIR spectrometry and chemometrics. Ninety-five sesame oil samples were collected from the six main producing areas of China to include most if not all of the significant spectral variations likely to be encountered in future authentic materials. Two class modeling techniques, the soft independent modeling of class analogy (SIMCA) and the partial least squares class model (PLSCM) were investigated and the data preprocessing techniques including smoothing, derivative and standard normal variate (SNV) tests were performed to improve the classification performance. It was demonstrated that SIMCA and PLSCM can detect various adulterated sesame oils doped with 3% or more (w/w) of other cheaper oils, including rapeseed, soybean, palm and peanut oils. First derivative, second derivative and SNV tests significantly enhanced the class models by reducing baseline and background shifts. Smoothing of raw spectra led to inferior identification performance and proved itself to be unsuitable because some of the detailed frequency details

were lost during smoothing. The best model performance was obtained with second derivative spectra by SIMCA (sensitivity 0.905 and specificity 0.944) and PLSCM (sensitivity 0.952 and specificity 0.937). Although it is difficult to perform an exhaustive sampling of all types of pure sesame oils and potential adulterations, PLS and SIMCA combined with FTIR spectrometry can detect most of current adulterations of sesame oils on the Chinese market.

Keywords Sesame oil · FTIR · Class modeling techniques · Soft independent modeling of class analogy · Partial least squares class model

Introduction

Sesame (*Sesamum indicum* L.) is one of the oldest domesticated oilseed crops [1]. It is grown in tropical and sub-tropical areas over 7.54 million hectares worldwide, yielding 3.3 million tons of seed [2]. China especially is one of the most important producers of sesame, accounting for approximately a 25% share in the world's total production [3]. Pure and genuine Chinese sesame oil is dark brown with high contents of oleic, linoleic, palmitic and stearic acids, as well as low concentrations of palmitoleic, linolenic and eicosenoic acids [4, 5]. Sesame oil is highly valued for its unique and pleasant flavor and high nutrition and health functions. In China, pure and authentic sesame oil is traditionally regarded as a high-quality vegetable oil and serves as an important source of edible and cooking oils. Unfortunately, in the domestic market, pure sesame oil is sometimes adulterated with some cheaper vegetable oils, such as soybean oil, corn oil, rapeseed oil and palm oil. Therefore, a continuing demand exists for rapid, inexpensive and effective techniques for routine authenticity analysis of sesame oil products.

D.-H. Deng · L. Xu (✉) · Z.-H. Ye · H.-F. Cui · X.-P. Yu (✉)
College of Life Sciences, China Jiliang University,
Hangzhou 310018, People's Republic of China
e-mail: lxchemo@163.com

X.-P. Yu
e-mail: xyp@cjlu.edu.cn

D.-H. Deng
College of Chemistry and Chemical Engineering,
Anyang Normal University, Anyang 455002,
People's Republic of China

C.-B. Cai
Department of Chemistry and Life Science,
Chuxiong Normal University, Chuxiong 675000,
People's Republic of China

The quality and chemical composition of sesame oil depend on various factors, including varieties, fields, processing and storage, producing areas, etc. [6–13]. Numerous researches have been devoted to the investigation of the chemical compositions and properties of sesame oils influenced by such factors. Such investigations are crucial for understanding the quality of sesame oils. However, it is still difficult to find a few marker components that can unambiguously indicate the origins and species of raw seeds. In traditional sensory analysis [14, 15], the quality of oils can be evaluated by professional tasters on a sensory board. The tasters are experienced specialists who evaluate product quality based on the flavor, color, and taste of oil. Because it may take years and is very expensive to train a skilled oil taster, it would be attractive to evaluate oil quality by some nonhuman techniques.

In recent years, it has seen increasing interest in applying spectroscopic and chemometric tools to food authentication [16–18]. The principle of such techniques is that chemical compositions of complicated samples are characterized by multi-channel analytical signals; then useful information concerning food quality can be extracted by multivariate analysis methods. Compared with the most widely used near infrared (NIR) spectroscopy techniques, mid-infrared possesses some advantages, including similar speed of measurement, moderate instrument cost and relative ease of sample presentation, especially for liquid and paste samples [18–21].

This work was aimed at developing a reliable and quick analytical method for authentication of Chinese sesame oils by Fourier-transform infrared (FTIR) spectroscopy and chemometric methods, focusing mainly on the following aspects. First, the sampling procedure should be representative and comprehensive to include most if not all of the significant variations likely to be encountered in the authentic materials [18, 19]. In this paper, real and pure sesame oils from the six main producing areas in China were collected to make up a sufficient training set. Second, it was pointed out that food authentication can be reduced to the question of whether an unknown object should be accepted or rejected by a class of interest. The objective of a class model is to describe the representative samples belonging to the class and answer the above question based on predictions of the unknown objects. In these cases, the commonly used discriminant analysis (DA) methods devoted to differentiating two or more predefined classes are of limited use because food authentication will be confronted with various known and unknown adulterations [22]. Therefore, instead of DA, class modeling techniques (CMT) [22, 23] were used in this paper. Finally, based on the above two considerations, the objective of data analysis was to enhance class models to detect even very subtle signal difference caused by low doping concentrations.

Soft independent modeling of class analogy (SIMCA) [24] and a recently proposed partial least squares class model (PLSCM) [25] were used to develop class models. To remove undesirable factors in the raw data, smoothing [26], derivatives [26] and standard normal variate (SNV) [27] were investigated to optimize data preprocessing. For data analysis, given the performances of different models are similar or have no significant differences, the models with least complexity and least preprocessing procedures were sought to ensure the generalization of models.

Materials and Methods

Sample Preparation

A set of 95 pure and authentic sesame oil samples from the six most important sesame producing areas of China were collected from domestic markets, including Henan (18), Hebei (16), Anhui (18), Hubei (16), Shandong (15) and Jiangsu (12). All the pure sesame oil samples were obtained from the cold pressing of sesame seeds. Samples were stored in a cool, dark area before spectrometry analysis.

Eighteen cheaper and common vegetable oils, including rapeseed oil (5), soybean oil (6), palm oil (5) and peanut oil (2) were purchased from domestic markets for adulteration. Then adulterated sesame oil samples were prepared by mixing the pure sesame oils with different ratios of the cheaper oils ranging from 2 to 20%. For each doping level, a supersaturated design [28] was used to select 18 pure sesame oil samples, each of which was adulterated with one of the 18 cheaper oils. Therefore, with doping levels at 2, 3, 4, 5, 8, 10 and 20 percent (w/w), 126 adulterated sesame oil samples were prepared and used as negative test samples. The doped samples are stored under the same conditions as the pure sesame oil samples before spectrometric analysis.

FTIR Spectrometric Analysis

The FTIR transmission spectra were measured within the spectral range from 4,000 to 400 cm^{-1} on a Nicolet Avatar 360 infrared spectrophotometer (Thermo Scientific, Waltham, MA) with a DTGS KBr detector. No further preprocessing of the oil samples was performed, and the spectra were measured in a KBr demountable absorption cell (150 μm) without any solvents. The resolution was 4 cm^{-1} , and the scanning interval was 1.929 cm^{-1} . Therefore, each spectrum had 1,868 individual data points for chemometric analysis. The scanning time was set to be 64, because an increase in scanning time did not significantly improve the quality of signal. Some of the raw spectra are demonstrated in Fig. 1.

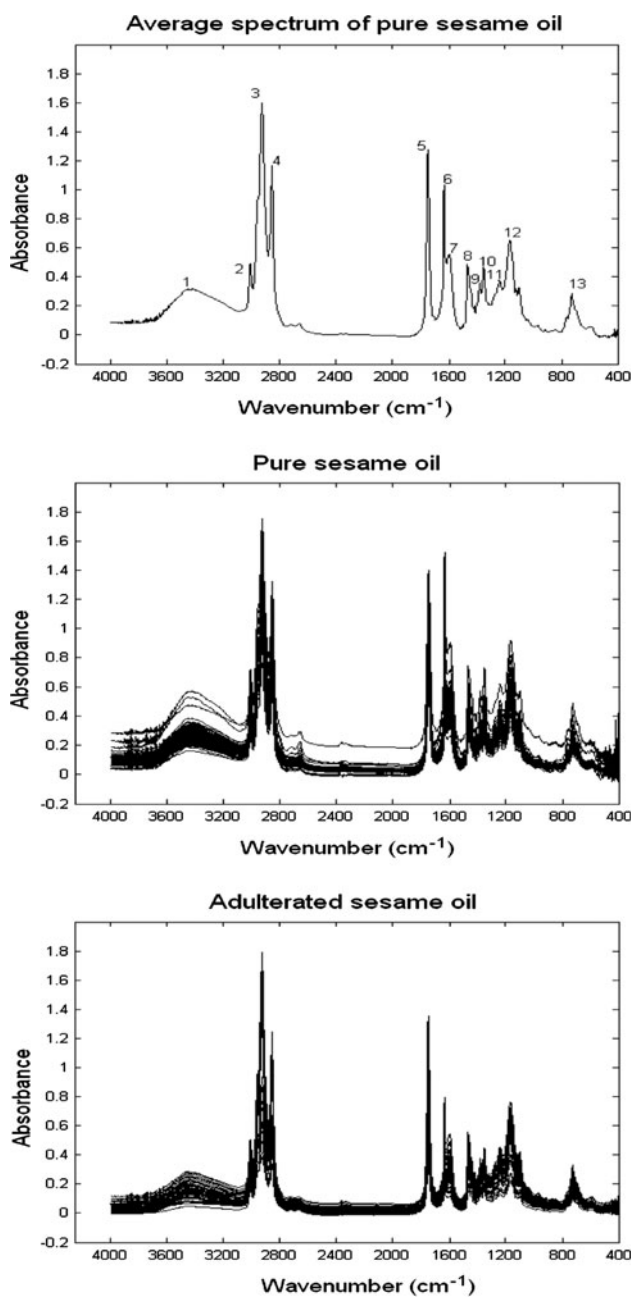


Fig. 1 FTIR spectra of pure and adulterated sesame oil; the doping levels ranged from 2 to 20%

Chemometric Data Analysis

The reliability of class models depends largely on the representativity of training samples. Therefore, robust principal component analysis (rPCA) [29, 30] was performed to detect and exclude the outliers. A robust PCA model can avoid the masking effects caused by the coexistence of multiple outliers. For pure sesame oils, the DUPLEX method [31] was used to split the measured spectra data into a representative training set and a test set.

DUPLEX selects two samples with largest distance and puts them in the training set, then selects two other samples with largest distance among the remaining samples and puts them in the test set, and so on. By alternatively selecting the spectral data for the calibration set and the test set, DUPLEX gives data in the test set with a distribution almost equal to that of the training set.

Different data preprocessing techniques were investigated to optimize the training and predicting performance of class models. Smoothing was often used to reduce random noise in spectral data and enhance the signal-to-noise ratio (SNR). The polynomial fitting algorithm by Savitzky and Golay [26] was adopted for this purpose considering its popularity and simplicity. Moreover, taking derivatives can enhance spectral resolution and remove baseline and background, so first and second order derivatives were also investigated. Because derivatives tend to degrade the SNR by enhancing noise, the derivative spectra were also computed by polynomial fitting algorithms [26]. Standard normal variate (SNV) [27] was originally proposed to reduce scattering effects and was also proved to be effective in correcting the interference caused by variations of optical path. Therefore, SNV tests were performed to standardize the spectral variations caused by different liquid membrane thickness.

The classical SIMCA [24] and a recently proposed PLSCM method [25] were used to developing class models for authentication of pure sesame oils. For one-class problems, SIMCA describes the structure of the training samples by the principal components (PC) space spanned by a few significant PCs. The residual error is then subject to an F test procedure for estimating the decision region. The residual error tends to be underestimated when it is computed directly from PCA of the training samples. It was recognized that SIMCA could lead to a large number of objects that are wrongly rejected (a large α -error), therefore, the strategy suggested in Ref. [24] were adopted. By using scores predicted by leave-one-out cross-validation (LOOCV) instead of the original scores obtained after PCA on the class objects, this procedure was shown to lead to a reduction of the number of false outliers. PLSCM was demonstrated to have a comparable performance to SIMCA. It builds a PLS regression model relating the spectral variables to a response vector of $\mathbf{1}$ with all the elements being ones. The vector $\mathbf{1}$ used as a response vector means all the objects in the same class should be distributed as close to each other as possible. While SIMCA projects the data onto a few PCs explaining most of the data variances, PLSCM considers both the explained variances and compactness of a class. The predicting error of the response variable is assumed to have a normal distribution and used to deduce the decision region. Because PLSCM can be performed in the framework of multivariate calibration, the determination of the model

Table 1 Peak assignments for pure sesame oil

No.	Peak (cm ⁻¹)	Assignment	Remarks
1	3,423	–OH	Stretching of phenolic O–H
2	3,008	–CH	Stretching of unsaturated –CH
3	2,925	–CH	Stretching of unsaturated –CH
4	2,854	–CH	Stretching of saturated –CH
5	1,747	–C=O	Stretching of C=O
6	1,631	Phenyl	Phenyl skeletal frequency
7	1,593	Phenyl	Phenyl skeletal frequency
8	1,466	–CH ₂	CH ₂ bending (~1,480 cm ⁻¹)
9	1,379	–CH ₃	Methyl symmetric bending
10	1,350	–CN	Stretching of –CN in aromatic amines
11	1,240	C–O–C	Asymmetrical stretching vibration
12	1,163	C–H	In-plane bending of aromatic C–H
13	723	–(CH ₂) _n –	Out-of-plane rocking

complexity and the decision region for PLSCM seems more straightforward than for SIMCA. More details concerning the optimization of parameters in SIMCA and PLSCM will be presented later.

Results and Discussion

As shown in Fig. 1, the spectra of adulterated and pure sesame oils [32] have very similar absorbance bands in the range of 400–4,000 cm⁻¹. Peak assignments for spectra of pure sesame oil as shown in Fig. 1 are listed in Table 1. Smoothing, first and second S-G derivatives and SNV were used to preprocess the raw spectra. Some of the raw and preprocessed spectra of pure and adulterated sesame oil are demonstrated in Figs. 2, 3. A significant difference between pure and adulterated sesame oil is the absorbance

at 3,423 cm⁻¹, however, the difference becomes very subtle with low doping concentrations. This can be further seen from the spectra of adulteration by peanut oils ranging from 2 to 20 percent in Figs. 2, 3. Therefore, class models are necessary to extract the detailed information from spectral data for characterizing pure sesame oils.

Because the FTIR spectral data were high-dimensional (for the raw spectra, $p = 1,868$), the improved rPCA [30] was used, which was shown to be numerically more stable for high-dimensional data and had a much lower computational cost. According to the computed score distance (SD) and orthogonal distance (OD), an rPCA diagnosis plot classifies the samples into four groups: regular data (with small SD and small OD), good PCA-leverage points (with large SD and small OD), orthogonal outliers (with small SD and large OD), and bad PCA-leverage points (with large SD and large OD). Outlier detection was performed on the raw spectra by rPCA with a significance level of 0.05. The number of PCs was determined by robust pooled predicted residual sum of squares (PRESS) values. Because the first seven PCs accounted for 95.35% of the total data variance, rPCA used seven PCs for outlier detection. The rPCA diagnosis plot of the 95 pure sesame oil samples is shown in Fig. 4. OD is a measure of the distance from the sample to the model space spanned by selected PCs and SD describes the sample dispersion in the class projected onto the model space. Therefore, both orthogonal outliers and bad PCA-leverage points should be excluded from the training set. Because the pure sesame oil samples came from different producing areas, there might be considerable difference in the contents of different chemical components. Therefore, good PCA-leverage points should be reserved to represent the spectral variations among the pure sesame oil samples from different producing areas. In Fig. 4, four orthogonal outliers were detected and

Table 2 Results by SIMCA and PLSCM with different data preprocessing methods

Model	Preprocessing	TP	FN	TN	FP	Sensitivity	Specificity
SMICA	Raw data (9)	17	4	103	23	0.810	0.817
	Smoothing (10)	16	5	106	20	0.762	0.841
	First derivative (7)	18	3	112	14	0.857	0.889
	Second derivative (7)	19	2	119	7	0.905	0.944
	SNV (7)	18	3	117	9	0.857	0.929
PLSCM	Raw data (10)	15	6	102	24	0.714	0.810
	Smoothing (9)	17	4	109	17	0.810	0.865
	First derivative (7)	18	3	114	12	0.857	0.905
	Second derivative (6)	20	1	118	8	0.952	0.937
	SNV (8)	18	3	113	13	0.857	0.897

Here “positive” and “negative” denote pure and adulterated sesame oil, respectively. Numbers in parentheses indicate numbers of PLSCM or SIMCA components

TP number of true positives, *FN* number of false negatives, *TN* number of true negatives, *FP* number of false positives

Sesame oil adulterated with different levels of peanut oil

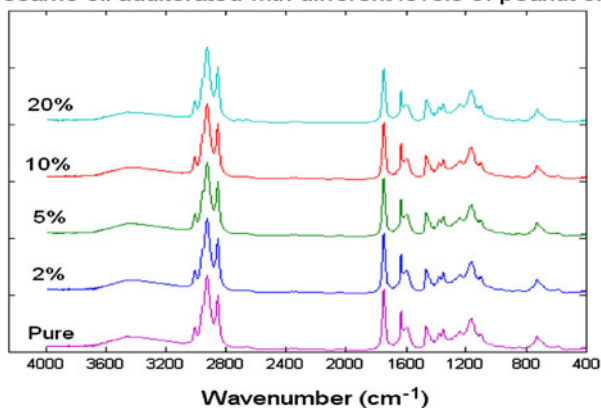


Fig. 2 FTIR spectra of sesame oil adulterated with different levels of peanut oil; the difference in absorbance at $3,423\text{ cm}^{-1}$ becomes very difficult to distinguish with the naked eye with low doping concentrations

excluded. The DUPLEX algorithm was then used to split the remaining 91 pure sesame oil samples into a training set with 70 samples and a test set of 21 samples. Therefore, the test set had 21 positive samples and 126 negative samples.

PLSCM and SIMCA were applied to model the pure sesame oil samples. For SIMCA, the decision region proposed in [24] was adopted to reduce the risk of having a large number of objects wrongly rejected. Cross validation was performed to evaluate the number of significant PCs, the criterion of 95% total explained variances was also considered. For PLSCM, Monte Carlo cross validation (MCCV) [33] with 10% samples left out was used to determine the number of PLS components and the sampling time was 100. The PRESS values by MCCV were subject to the F test proposed in [34, 35]. As suggested, a significance of 0.25 was adopted to select the least number of latent variables with a PRESS value not significantly larger than the minimum value according to the F test. Sensitivity and specificity were used to evaluate the performance of different models and preprocessing options. The prediction results of test samples by SIMCA and PLSCM were summarized in Table 2. Seen from Table 2, preprocessing generally improved the classification performance in terms of sensitivity and specificity. However, the models based on smoothed spectra had inferior performance, which might be attributed to the possible loss of detailed frequency information [36]. First derivative, second derivative and SNV significantly enhanced the class models by reducing the baseline and backgrounds. The model complexity of SIMCA and PLSCM models based on such preprocessing was less than those based on the smoothed and the raw spectra. For both PLSCM and SIMCA, the best class models were obtained by taking second derivative spectra and the prediction results were demonstrated in Figs. 5, 6. In Figs. 5b, 6b, the adulterated sesame oil samples were arranged according to an ascending doping

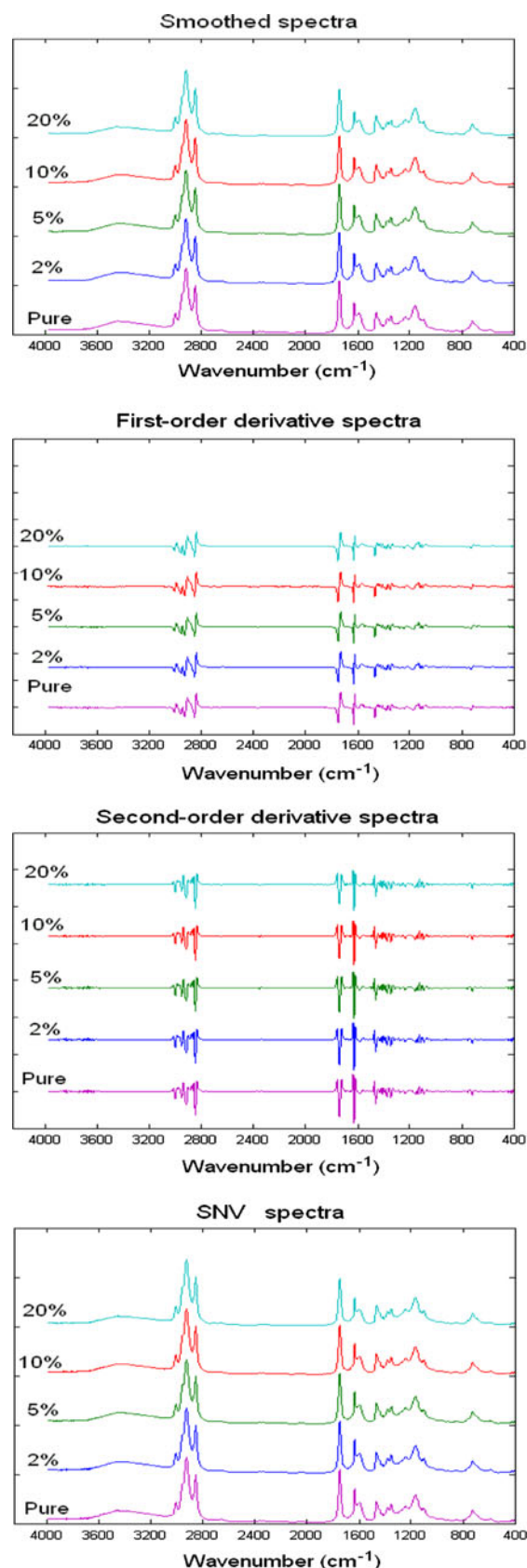


Fig. 3 Preprocessed spectra of sesame oil adulterated with peanut oil of different levels by smoothing, taking derivatives and SNV

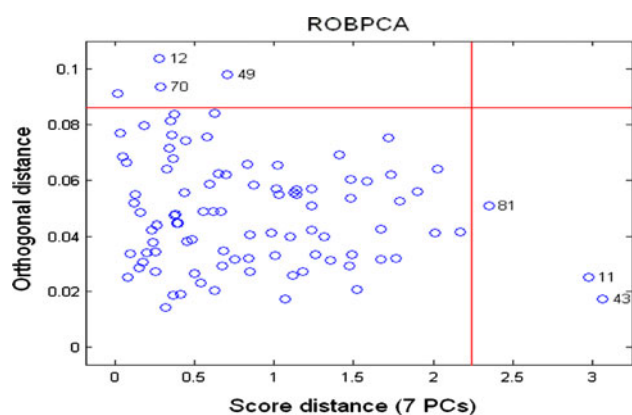


Fig. 4 Robust PCA outlier diagnosis plots for 95 pure sesame oil samples from Henan (18), Hebei (16), Anhui (18), Hubei (16), Shandong (15) and Jiangsu (12) of China

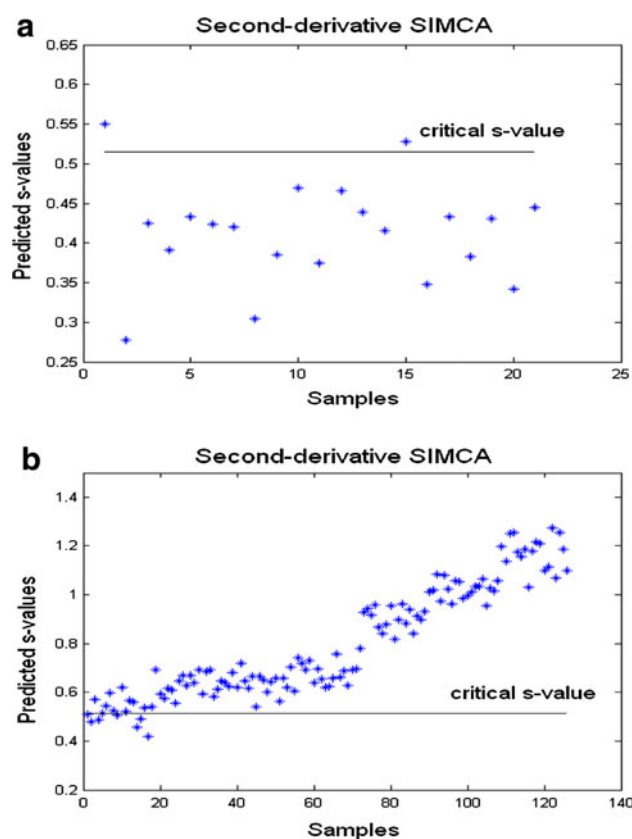


Fig. 5 Predictions of test samples by second derivative SIMCA with 6 PCs for **a** 21 pure sesame oils and **b** 126 adulterated sesame oils in an ascending order of doping levels

concentration, namely 2, 3, 4, 5, 8, 10, and 20%. The wrongly accepted samples (false positives) had a doping concentration of 2% and all the adulterated samples with a doping concentration equal to or higher than 3% were correctly detected. Moreover, as expected, the distance of an adulterated sample to the critical value increases with doping concentration, indicating the identification of a doping level over 3% was straightforward. The comparison of different

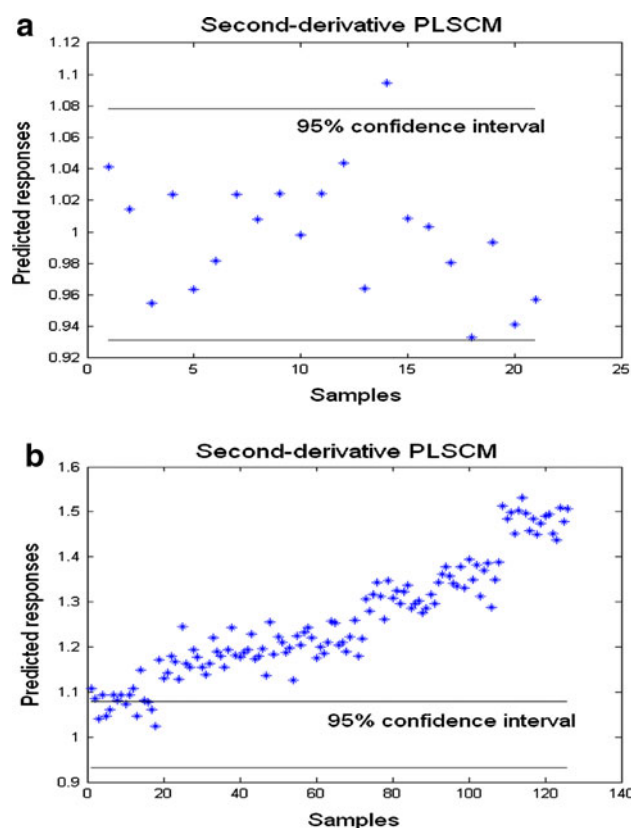


Fig. 6 Predictions of test samples by second derivative PLSCM with 7 components for **a** 21 pure sesame oils and **b** 126 adulterated sesame oils in an ascending order of doping levels

preprocessing methods demonstrated that the spectral variations caused by scattering effects and baseline shifts played a more important role than an inferior SNR.

Conclusions

This paper has demonstrated that FTIR combined with chemometric class models provide a reliable and practical method for detecting adulterated sesame oils. Although it is difficult to perform an exhaustive sampling of all types of pure sesame oils and potential adulterations, PLS and SIMCA proved themselves to be successful at authenticating pure sesame oil samples from the six main producing areas in China. Moreover, the data pretreatments of derivatives and SNV were effective in reducing the influence of spectral variations caused by baseline shifts and scattering effects. The combination of FTIR and class modeling methods were able to detect adulterations with doping concentrations over 3%, which is enough to identify most of the current adulterations in Chinese markets. Wavelength selection might be useful to further enhance the class models to detect even lower levels of adulteration, however, care should be taken when performing a

wavelength selection. Because too few spectral variables would fail in characterizing oils sufficiently, the use of a class model would be restricted. How to make a tradeoff between sufficient characterization of samples and proper variable selection/reduction will be the focus of our future investigations. Moreover, because the diverse origins of samples can increase model complexity and degrade model performance, we will attempt to combine clustering methods with class models to tackle this problem.

Acknowledgments This work was financially supported by the National Public Welfare Industry Projects of China (no.201210010), Hangzhou Programs for Agricultural Science and Technology Development (no.20101032B28) and the Key Scientific and Technological Innovation Team Program of Zhejiang Province (no.2010R50028). Dr. Chen-Bo Cai is grateful to the financial aid of the Applied and Basic Research Project of Yunan Provincial Science and Technology Department (no.2010CD087).

References

- Tashiro T, Fukuda Y, Osawa T, Namiki M (1990) Oil and minor components of sesame (*Sesamum indicum* L.) strains. *J Am Oil Chem Soc* 67:508–511
- Jyothi B, Ansari NA, Vijay Y, Anuradha G, Sarkar A, Sudhakar R, Siddiq EA (2011) Assessment of resistance to *Fusarium* wilt disease in sesame (*Sesamum indicum* L.) germplasm. *Australasian Plant Pathol* 40:471–475
- Salunkhe DK, Chavan JK, Adsule RN, Kadam SS (1992) World oilseeds: chemistry technology and utilization. Van Nostrand Reinhold, New York
- Jamieson GS, Baughman WF (1924) The chemical composition of sesame oil. *J Am Chem Soc* 46:775–778
- Budowski P, Markley KS (1951) The chemical and physiological properties of sesame oil. *Chem Rev* 48:125–151
- Abou-Gharbia HA, Shahidi F, Adel A, Shehata Y, Youssef MM (1997) Effects of processing on oxidative stability of sesame oil extracted from intact and dehulled seeds. *J Am Oil Chem Soc* 74:215–221
- Arslan Ç, Uzun B, Ülger S, İlhan Çağırman M (2007) Determination of oil content and fatty acid composition of sesame mutants suited for intensive management conditions. *J Am Oil Chem Soc* 84:917–920
- Shahidi F, Amarowicz R, Abou-Gharbia HA, Adel A, Shehata Y (1997) Endogenous antioxidants and stability of sesame oil as affected by processing and storage. *J Am Oil Chem Soc* 74:143–147
- Uzun B, Arslan Ç, Furat Ş (2008) Variation in fatty acid compositions oil content and oil yield in a germplasm collection of sesame (*Sesamum indicum* L.). *J Am Oil Chem Soc* 85:1135–1142
- Abou-Gharbia HA, Adel A, Shehata Y, Shahidi F (2000) Effect of processing on oxidative stability and lipid classes of sesame oil. *Food Res Int* 33:331–340
- Were BA, Onkware AO, Gudu S, Welander M, Carlsson AS (2006) Seed oil content and fatty acid composition in East African sesame (*Sesamum indicum* L.) accessions evaluated over 3 years. *Field Crop Res* 97:254–260
- Rangkadilok N, Pholphana N, Mahidol C, Wongyai W, Saengsooksree K, Nookabkaew S, Satayavivad J (2010) Variation of sesamin, sesamol and tocopherols in sesame (*Sesamum indicum* L.) seeds and oil products in Thailand. *Food Chem* 122:724–730
- Yen GC, Shyu SL (1989) Oxidative stability of sesame oil prepared from sesame seed with different roasting temperatures. *Food Chem* 31:215–224
- Shimoda M, Shiratsuchi H, Nakada Y, Wu Y, Osajima Y (1996) Identification and sensory characterization of volatile flavor compounds in sesame seed oil. *J Agric Food Chem* 44:3909–3912
- Vlachos A, Arvanitoyannis IS (2008) A review of rice authentication/adulteration methods and results. *Crit Rev Food Sci* 48:553–598
- Alishahi A, Farahmand H, Prieto N, Cozzolino D (2010) Identification of transgenic foods using NIR spectroscopy: a review. *Spectrochim Acta A Mol Biomol Spectrosc* 75:1–7
- Cen H, He Y (2007) Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci Tech* 18:72–83
- Tzouros NE, Arvanitoyannis IS (2001) Agricultural produces: synopsis of employed quality control methods for the authentication of foods and for the classification of foods according to their variety of geographical origin. *Crit Rev Food Sci* 41:287–319
- Karoui R, Downey G, Blecker C (2010) Mid-infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships—a review. *Chem Rev* 110:6144–6168
- Wilson RH, Tapp HS (1999) Mid-infrared spectroscopy for food analysis: recent new applications and relevant developments in sample presentation methods. *Trend Anal Chem* 18:85–93
- Arvanitoyannis IS, Vlachos A (2007) Implementation of physicochemical and sensory analysis in conjunction with multivariate analysis towards assessing olive oil authentication/adulteration. *Crit Rev Food Sci* 47:441–498
- Forina M, Oliveri P, Lanteri S, Casale M (2008) Class-modeling techniques, classic and new, for old and new problems. *Chemom Intell Lab Syst* 93:132–148
- Brereton RG (2011) One-class classifiers. *J Chemom* 25:225–246
- De Maesschalck R, Candolfi A, Massart DL, Heuerding S (1999) Decision criteria for soft independent modelling of class analogy applied to near infrared data. *Chemom Intell Lab Syst* 47:65–77
- Xu L, Fu HY, Jiang N, Yu XP (2010) A new class model based on partial least square regression and its applications for identifying authenticity of bezoar samples. *Chin J Anal Chem* 38:175–180
- Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least-squares procedures. *Anal Chem* 36:1627–1639
- Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard normal variate transformation and detrending of near infrared diffuse reflectance. *Appl Spectrosc* 43:772–777
- Tang B, Wu CFJ (1997) A method for constructing super-saturated designs and its Es2 optimality. *Can J Stat* 25:191–201
- Stanimirova I, Walczak B, Massart DL, Simeonov VA (2004) Comparison between two robust PCA algorithms. *Chemom Intell Lab Syst* 71:83–95
- Hubert M, Rousseeuw PJ, Verboven SA (2002) Fast method for robust principal components with applications to chemometrics. *Chemom Intell Lab Syst* 60:101–111
- Snee RD (1977) Validation of regression models, methods and examples. *Technometrics* 19:415–428
- Mirghani MES, Che Man YB, Jinap S, Baharin BS, Bakar J (2003) Application of FTIR spectroscopy in determining sesamol in sesame seed oil. *J Am Oil Chem Soc* 80:1–4
- Xu QS, Liang YZ (2001) Monte Carlo cross validation. *Chemom Intell Lab Syst* 56:1–11
- Haaland DM, Thomas EV (1988) Partial least-squares methods for spectral analyses. 1. relation to other quantitative calibration methods and the extraction of qualitative information. *Anal Chem* 60:1193–1202
- Haaland DM, Thomas EV (1988) Partial least-squares methods for spectral analyses. 2. application to simulated and glass spectral data. *Anal Chem* 60:1202–1208
- Kokalj M, Rihtarič M, Kreft S (2011) Commonly applied smoothing of IR spectra showed inappropriate for the identification of plant leaf samples. *Chemom Intell Lab Syst* 108:154–161