Rapid Detection of Edible Oil Adulteration

Combining MALDI TOF Mass Spectrometry and Statistical Learning

Abstract

MALDI-TOF mass spectrometry has an important analytical advantage due to its simple and fast operation. The MALDI-TOF spectra of edible oils contain complex and subtle features that can be analyzed with multivariate statistical tools for differentiation and subsequent authentication. In this study, we combine MALDI-TOF mass spectrometry and statistical learning tools for the rapid classification of olive oils adulterated with soybean oils.

From a group of 46 olive oil spectra and 101 spectra of olive oil adulterated with 10% soybean oils, the cross validation of the classification model developed had an accuracy of >99% correct.

Keywords:
Olive Oil, Soybean Oil, MALDI, edible oil analysis, food adulteration, microflex LRF, flexAnalysis, Profile Analysis, MALDI-TOF, Applied Chemistry, Food adulteration, Authenticity Testing

Authors: Zhaoyang Liu, Kun Zhu, Kefei Wang · Bruker Daltonics, Shanghai, China
Jens Fuchser, Bob Galvin · Bruker Daltonik GmbH Bremen

Introduction

The adulteration of food stuffs for economic gain is an ancient fraud. High priced olive oils have long been targets for adulteration, often using significantly cheaper vegetable oils such as soybean oil. Mass spectrometry based techniques, particularly GC-MS and LC-MS, are widely used in developing analytical methods for identification of the adulterated olive oils. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has also been increasingly applied for such analyses [1-4]. Compared to chromatography based mass spectrometry, MALDI-TOF MS has the notable advantages of being simple and fast, and not requiring time-consuming chromatographic separation or derivatization.

One common approach of these studies involves the measurement of the profiles of the triacylglycerols (TAGs), the major constituents in all edible oils. Multivariate statistical tools such as principle component analysis (PCA) are often used to visualize the classification based on the difference in the TAG profiles in different oils [1-2]. Such an approach suffers two major drawbacks: first, the tedious identification and quantification of targeted TAGs, which may be unreliable due to their tendency to fragment during sample preparation and within ion sources [5]; second, the use of PCA as an unsupervised multivariate statistical visualization tool provides limited capabilities of prediction and result interpretation. Additionally, by targeting only the TAG's (m/z 850-1000), potential distinguishing features in other mass ranges may be missed.

In this study, we present a new method using a supervised statistical learning model to perform the non-targeted analysis of MALDI-TOF spectra collected from edible oil samples. The sparse least squares analysis discriminant analysis (sPLS-DA), a variant of least squares analysis discriminant analysis (PLS-DA), was used for building a supervised learning model to predict the classification. Cross validation result demonstrates that olive oils adulterated with 10% soybean oils can be predicated with 99.2% accuracy.

Experimental

Materials and Sample Preparation

Samples of five olive oil and eight soybean oil brands from local supermarkets were provided by a local FDA laboratory. The 10% soybean oil adulterated olive oil samples were prepared by mixing 9:1 (v/v) of the two oils accordingly. All oil samples were diluted (200x) with methanol/chloroform (2:1 v/v) before analysis.

The MALDI matrix solution was prepared by dissolving of CHCAE (α-cyano 4-hydroxycinnamic acid methyl ester, Bruker p/n 203071) in methanol/chloroform (2:1 v/v) for a working solution of 10 mg/ml.

Instrumentation

MALDI-TOF MS analysis was carried out on a benchtop microflex LRF model MALDI-TOF mass spectrometer (Bruker Daltonik GmbH, Bremen) equipped with a nitrogen laser at 337 nm and using a repetition rate of 60 Hz. The laser power was adjusted across a narrow (relative power) range to avoid detector saturation. The instrument was operated in reflectron mode with positive ion acquisition, and spectra in the range of m/z 100-3000 were recorded. For each spectrum, between 300 and 500 shots were summed.

For sample analysis, 1 µL of diluted oil sample was applied by pipette on a stainless steel MALDI plate, followed by pipetting 1 µL matrix solution onto the oil spot. Each spot was analyzed in triplicate. For the adulteration classification, a total of 147 spectra were collected: 46 on olive oil (OL) and 101 on 10% soybean oil adulterated olive oil (OS). Spectra were also collected for soybean oils alone (SB) and matrix alone under the same conditions.

Statistical Analysis

All MALDI-TOF MS data files were imported into Bruker's Profile Analysis software for peak alignment and normalization, followed by export to a .csv file of m/z peaks vs intensities. All inherent CHCAE matrix peaks were removed and excluded from statistical analysis. The final .csv file used for statistical analysis contained normalized intensities of 190 m/z peaks within each of the 147 spectra.

The statistical computing was performed with the open source R language (https://www.r-project.org/) with an RStudio code editor (https://www.rstudio.com/) on a laptop with an i7 processor/12GB RAM running Windows 7. The mixOmics library (http://mixomics.org/) was installed for PLS-DA and sPLS-DA analysis. The spectral files were randomly divided into a training set containing approxmately 2/3 of the data, and a test set containing the remaining 1/3 of the data. The training set was used to build the statistical learning model, which was then used to predict the classification of the test set for cross validation. This computing procedure was repeated 10 times, each time with new randomzied training and test data sets, and the total classification accuracy was determined.

Results

Given that the profiles of MALDI-TOF spectra are dependent upon the choice of matrix, an initial study was carried out to evaluate various MALDI matrices for edible oil analysis. The results indicate that with CHCAE the MALDI-TOF spectra reveal the most abundant features, particularly in the low mass range of m/z 250-600, thus CHCAE was used for all subsequent studies. It was noted that few peaks (S/N >10) were observed beyond m/z 1200.

The MALDI-TOF spectra for pure olive oil (OL), pure soybean oil (SB) and 10% adulterated olive oils (OS) are shown in Figure 1. The oil spectra can be divided into four regions corresponding to four classes of compounds reported in the literature [4-7]: ~m/z 900 for TAGs; ~m/z 600 for diacylglycerols (DAGs); ~m/z 400 for monoacylglycerols (MAGs); and m/z <~350 for the ion fragments of fatty acid esters. Under the conditions tested, and as has been reported by other research groups [5-7], the peaks of DAG's, MAG's, and fatty acid esters are more intense than the TAGs, although the latter are the known primary components of edible oils. Interference from matrix peaks is a concern with the use of MALDI-TOF for analysis of small molecules (< m/z 600~700), as illustrated by several intense matrix peaks in Figure 1, thus these be must be removed before applying any statistical analyses.

Statistical learning (i.e., machine learning) essentially is to "teach" a computer to follow a certain algorithm to learn the pattern from the known data (training set), then to "validate" the learned knowledge by predicting the classification of "unknown" data (test set) [8]. In present study, 2/3 of 146 total data points were used as the training set, with the remaining 1/3 as a test set to validate the accuracy of the model (cross validation). In order to reduce the bias in determining the validation accuracy, this "learn-and-predict" cycle was performed 10 times, each time with a new random selection of training and test data sets.

R was selected for statistical learning by virtue of its versatility and flexibility as an open source (free) platform for statistical data analysis and visualization. Two algorithms were chosen for the current classification study, the partial least squares discriminant analysis (PLS-DA) and the (relatively new) sparse least squares discriminant analysis (sPLS-DA).

The results of cross validation from the two algorithms are given in Table 1, and representative score plots of first two components from each algorithm are shown in Figure 2. As shown, both PLS-DA and sPLS-DA can classify OL and OS groups with high accuracy rates of 95.9% and 99.2%, respectively. Nonetheless, the sPLS-DA provides a superior differentiation and classification of the two oil groups as shown in Figure 2.

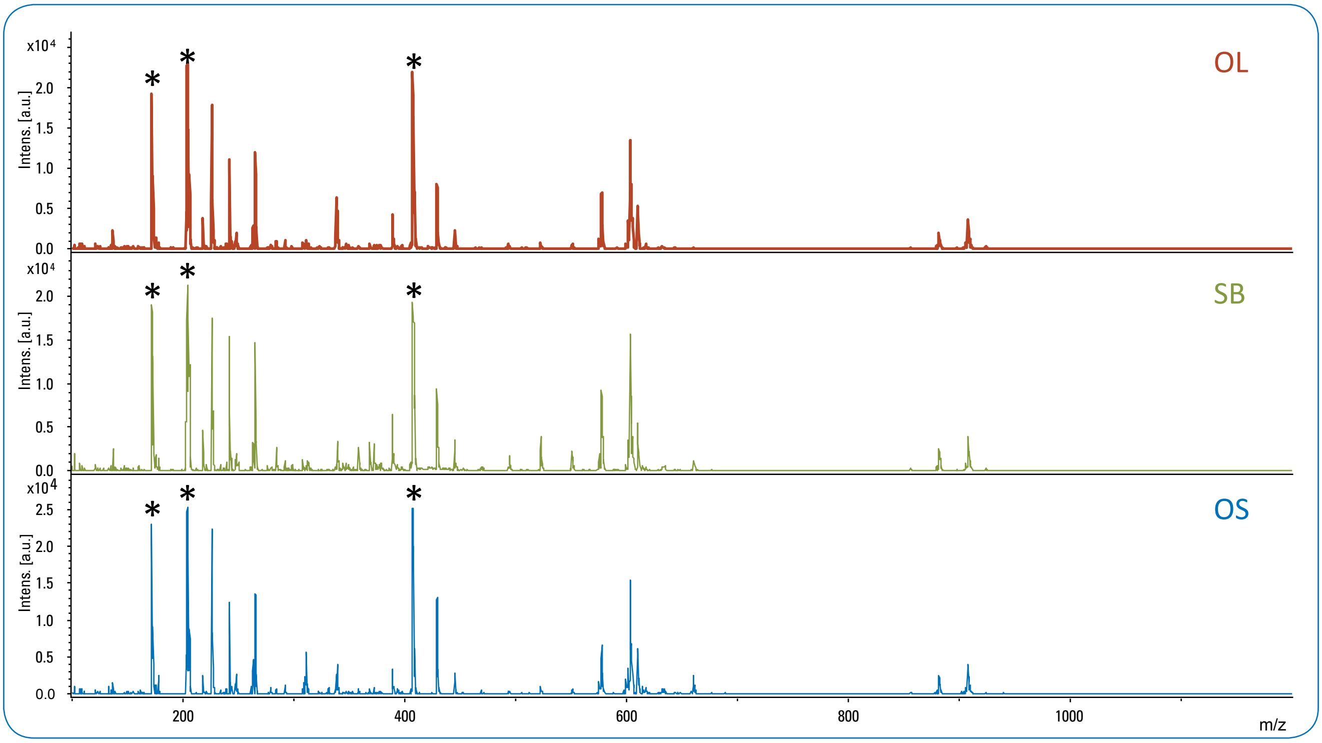

Fig. 1: MALDI-TOF spectra of pure olive oil (OL), soybean oil (SB) and olive oil adulterated with 10% soybean oil (OS). The matrix peaks are labeled with.*

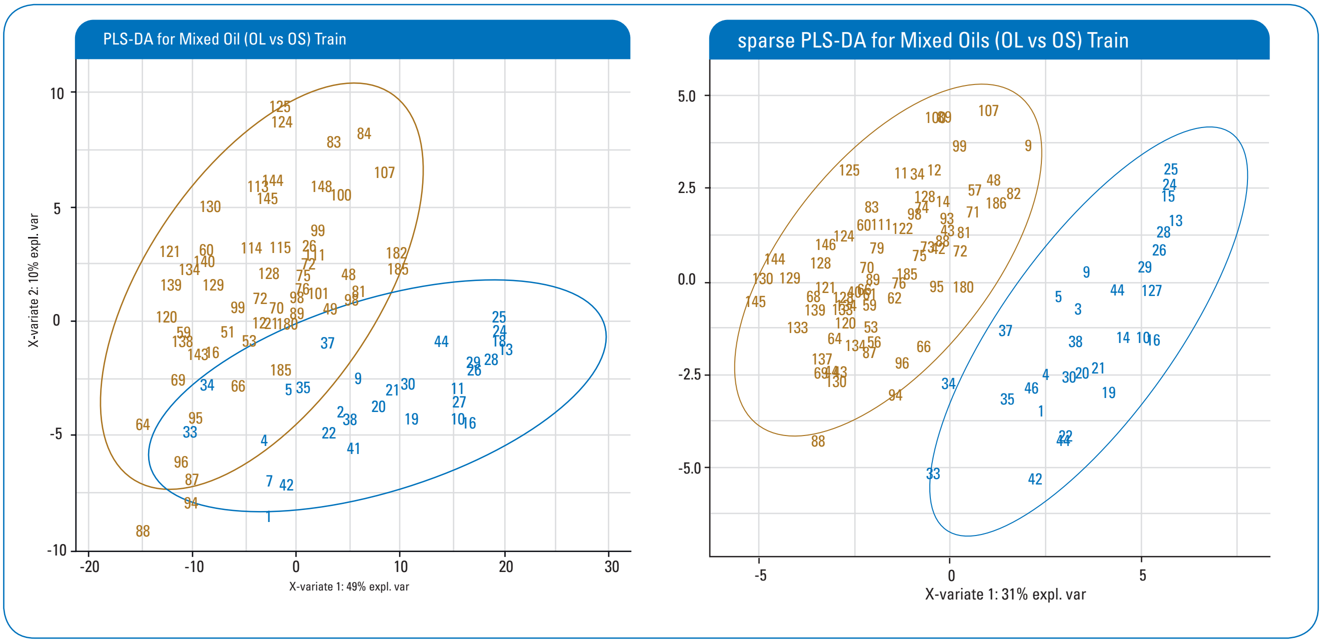

Fig. 2: Representative score plot of first vs second components of a training set from the PLS-DA and sparse PLS-DA models (red: OL; blue: OS).

Model	**PLS-DA**		**sPLS-DA**	
Actual / **Predict**	OL	OS	OL	OS
OL	146	6	159	1
OS	15	342	3	350
% Correct	95.9%		99.2%	

Table 1: Cross Validation of Statistical Learning Models.

An additional advantage of the sPLS-DA over PLS-DA is the result interpretability, as sPLS-DA performs both the variable selection and classification simultaneously. In current study, it was determined that a minimum of three components, with each limited to the use of 10 variables (predictors), i.e. a total of 30 m/z peaks (= 3 x 10), were sufficient to achieve the indicated classification results; increasing the components and/or predictors does not improve the cross validation accuracy. A subset of these major predictor m/z peaks, which could be considered as potential "biomarkers" for further investigation of the subtle spectra differences of the OL and OS groups, are listed in Table 2. Overlaid mass spectra of classification biomarker peaks (likely DAG, MAG, and fatty acid esters fragment ions) for olive oil, soybean oil and 10% adulterated olive oil are shown in Figure 3.

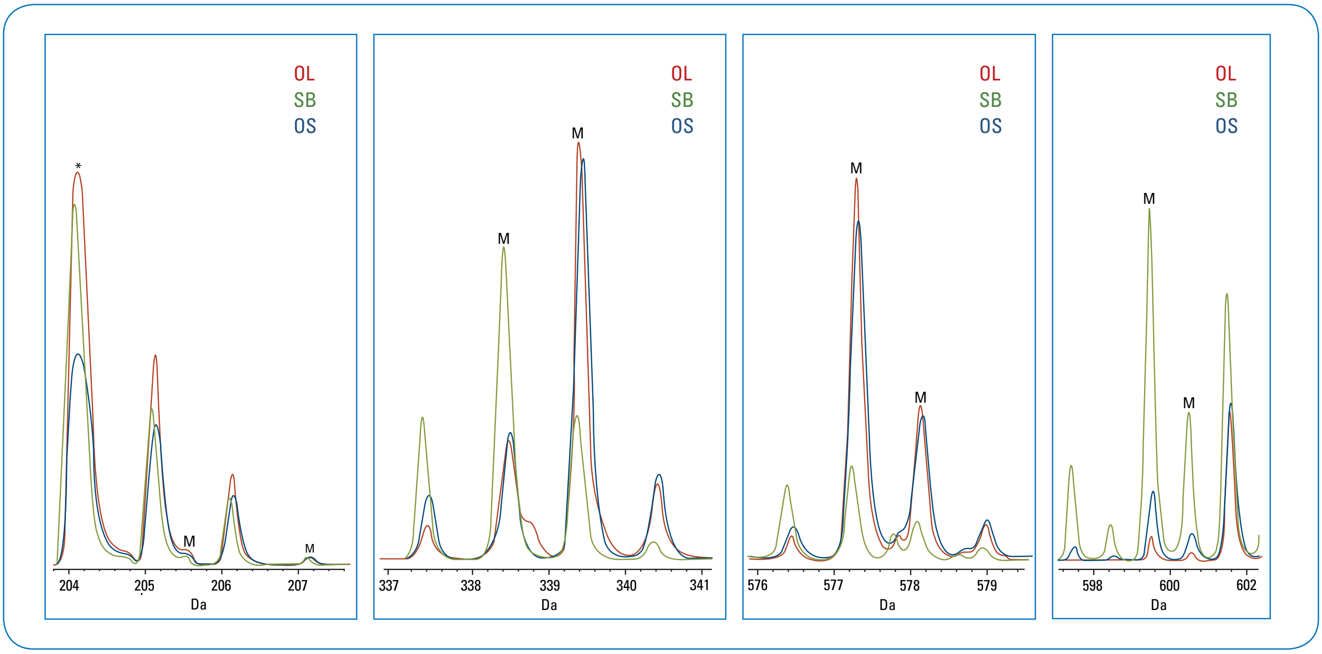

*Fig. 3: Major biomarker peaks ("M") identified from sPLS-DA analysis as predicators for differentation of olive oil from soybean oil adulterated olive oil (*indicates a matrix peak) The scales for the spectra within each mass range are arbitrary for illustration of the differences of the three oils.*

	Component 1	Component 2	Component 3
Predictors (m/z)	394.5	577.6	338.5
	599.5	578.6	339.4
	600.5	205.5	
		207.2	

Table 2: Major Predictors Identified from the sPLS-DA Model.

Conclusion

The combination of the MALDI-TOF mass spectrometry and statistical learning can be applied for robust and accurate classification of pure olive and olive oils adulterated with 10% soybean oils. The analysis method is fast and simple, with no need for characterization or quantification of any particular chemical component(s) within the oils. Furthermore, supervised statistical learning with sPLS-DA has been shown to be useful in identifying potential biomarkers from complex spectral features, as in non-targeted "omics" studies.

Learn More

You are looking for further Information? Check out the Link or scan the QR Code.

www.bruker.com/microflex-lrf

References

[1] A. Jakab, K. Nagy, K. Heberger, K.Vekey, E. Forgacs. Differentiation of vegetable oils by mass spectrometry with statistical analysis, Rapid Commun. Mass Spectrom. 2002, 15, 2291-2297.

[2] B. Chapagain, Z. Wiseman, MALDI-TOF/MS fingerprinting of triacylglycerols (TAGs) in olive oils produced in the Israeli Negev Desert, J. Agric. Food Chem, 2009, 57, 1135-1142.

[3] G Asbury, K. Al-Saad, W. F. Siems, R. Hannan, H. Hill, Analysis of triacylglycerols and whole oils by matrix-assisted laser desorption/ionization time of flight mass spectrometry, J. Am. Soc. Mass Spectrom. 1999, 10, 983-991.

[4] J. Lay, R. Liyangage, B. Durham, J. Brooks, Rapid characterization of edible oils by direct matrix assisted laser desorption/ionization time-of-flight mass spectrometry analysis using triacylglycerols, Rapid Commun. Mass Spectrom. 2006, 20, 952-958.

[5] J. Gidden, R Liyangage, B. Durham, J. Lay, Reducing fragmentation observed in the matrixassisted laser desorption/ionization time-of-flight mass spectrometric analysis of triacylglycerols in vegetable oils, Rapid Commun. Mass Spectrom. 2007, 21, 1951-1957.

[6] C. Calvano, F. Palmisano, C. Zambonin, Laser desorption/ionization time-of-flight mass spectrometry of triacylglycerols in oils, Rapid Commun. Mass Spectrom. 2005, 19, 1315-1320.

[7] A. Kubo, T. Satoh, Y. Itoh, M. Hashimoto, J. Tamura, R. Cody, Structural analysis of triacylglycerols by using a MALDI-TOF/TOF system with monoisotopic precursor selection, J. Am. Soc. Mass. Spectrom. 2013, 24, 684-689.

[8] G. James, D. Witten, T. Hastie, R Tibshirani, An Introduction to Statistical Learning, , Springer Science+Business Media New York, 2013 (6h corrected printing 2015), Chapters 2-6.

Bruker Daltonics GmbH & Co. KG

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com - www.bruker.com