Acta Chimica Slovaca, Vol.1, No 1, 2008, 165 – 174

Chemometric classification of edible oils

Dáša Kružlicová[a], Viera Mrázová[b], Katja Šnuderl[c], Ján Mocák[a,b,*]

Ernst Lankmayr[d]

[a] *Institute of Analytical Chemistry, Slovak University of Technology, SK-81237 Bratislava, Slovakia*
[b] *Department of Chemistry, University of Ss. Cyril & Methodius, SK-91701 Trnava, Slovakia*
[c] *Faculty of Chemistry and Chemical Engineering, University of Maribor, Maribor, SI-2000 Slovenia*
[d] *Institute for Analytical Chemistry, University of Technology Graz, A-8010 Graz, Austria*

**jan.mocak@stuba.sk*

Abstract

This paper deals with spectral characterization of different sorts of olive oil and pumpkin seed oil, supplemented by sensory evaluation of the oil quality. UV-Vis, NIR and FTIR absorption spectroscopy and fluorescence spectroscopy were used as the measurement techniques. Chemometrical data processing, mainly multidimensional data analysis, enabled detection of the most informative properties concerning the oil quality. Classification of both kinds of oil samples by sensorial quality was successfully performed by several techniques of discriminant analysis and artificial neural networks. Comparison of olive oil classification based on different spectral techniques enables to find most useful way of the spectral measurement.

Keywords: classification, sensorial quality, multidimensional data analysis, olive oil, pumpkin seed oil

Introduction

Olive oil and pumpkin seed oil

Vegetable oils are essential in global nutrition. Depending on the regional conditions, a variety of oils are produced in different qualities. Worldwide known and widely used is olive oil, which enjoys special and increasing popularity in many countries not only due to its delicate taste but also because of its nutrition value. The beneficial health effects of olive oil are caused mainly by high content of monounsaturated fatty acids as well as high content of

antioxidative substances. Olive oil has been linked to lowering the risk of heart disease, reduction in the level of the LDL (bad) cholesterol, lowering of blood pressure, decrease in blood sugar levels, increase in the absorption of vitamins A, D, E, and K, stimulation of the gall bladder to secrete bile, improvement of the metabolism and favourable effects to the skin and hair.

Especially in the Central European area pumpkin seed oil is very popular mainly due to its characteristic taste. The sorts of this oil are produced from the seeds of pumpkins (*Cucurbita pepo* L.) and consist of approximately 70 % unsaturated fatty acids. They contain also a number of hydrocarbons, tri-terpenoides, carotenoides, tocopheroles and phytosteroles. Besides its main use for the production of edible oil, the pumpkin seeds have been utilised in pharmaceutical industry to relieving from disorders of the prostate gland and hyperplasia of urinary bladder.

In general, chemical analysis of edible oils is difficult not only since it consists of a very complex mixture of chemical compounds but also due to a huge matrix effect. Modern instrumental analytical measurements combined with computer control and the use of advanced software creates new possibilities for solution of problems in various practical application fields, e.g. in assessing quality of raw and processed food (Lankmayr et al. 2004). In this work, a complete spectral characterization of commercial olive oil and pumpkin oil samples was performed. For this purpose, electronic spectra in UV and visible region, near infrared spectra, Fourier transform infrared spectra and fluorescence spectra were recorded.

Characterization, classification and authentication of edible oils

Edible oils belong to the commodities, which are very frequent objects of falsification (Petka et al. 2001, Lankmayr et al. 2004). It is therefore necessary to develop such procedures which make possible their characterization in detail, classification by the chosen classification criteria, and finally, their authentication (Penza and Cassano 2004, Ferreira et al. 2007). To perform the last task it is necessary to investigate and verify the selected samples with regard to the vegetable variety, origin location or confirming the producer, and sometimes also the year of production.

Various analytical methods and chemometrical techniques were used and described for characterization and classification of edible oils. A very important problem of their authentication is concerned in lit. (Dennis 1998, Mannina et al. 2001, Mannina et al. 2003, Brescia et al. 2003, Brodnjak-Vončina et al 2005); some of the published works are focused to the spectral properties (Vigli et al. 2003, Downey et al. 2003, Christy et al 2004, Marini et

al. 2004, Rezzi et al. 2005, D'Imperio et al. 2007), which were proved, in general, most advantageous when investigating vegetable oils.

Tied up to the measured spectral data, several chemometric techniques are used in this work, namely the techniques of multidimensional data analysis including various kinds of discriminant analysis as well as artificial neural networks.

Experimental

Spectral measurements

UV-Vis spectra of the edible oil samples were recorded in 1 cm cell in the region 200–700 nm using computer controlled spectrophotometer Varian, Cary 50 Conc (Varian, Victoria, Australia), complemented by software package Cary Win UV, which was used for the data acquisition and processing. Near infrared spectra were recorded by means of an Analect Diamond-20 FT-NIR spectrometer (Hamilton Sundstrand, Pomona, CA, USA) with Analect FX90 software for data processing. A 0.1 cm pathlength quartz cell (160.001-QS, Hellma, Germany) was thermostatted at 25°C with a modified temperature control system (DMA58, Anton Paar, Austria). A Fourier transform infrared spectrometer MATTSON 3000 (Mattson Instruments, Bucks, England), interfaced to a personal computer, was used for recording the FTIR spectra. Fluorescent spectra were recorded by means of the Cary spectrophotometer equipped by Cary Total Fluorescence Accessory. The corresponding samples were diluted 1:300 (v/v) by isooctane (spectroscopy grade, Merck, Darmstadt, Germany).

The recorded spectra of all kinds were digitized using the selected step in wavelength or wave number and saved to the PC hard disc; the absorbances of the digitized spectra were used as the spectral variables suitable for further chemometric processing (Kružlicová et al. 2007).

Oil samples

The olive oil samples originated from three main Greek peripheries (Central Greece, Peloponnese and Crete). Altogether 193 samples of 5 different olive oils of the type *M* (31 samples), *K* (37), *E* (13), *N* (94), and *T* (18) were investigated. The sample designation by the above mentioned codes was made and demanded by the sample donators.

The absorbance of 70 commercially available pumpkin oils were measured in the UV-Vis spectral region. Absorbance of 80 and 82 pumpkin oils was obtained by the NIR and FT

IR spectral measurements, respectively. The dominant part of the samples of Styrian origin was the same in measurements by all above mentioned techniques. For fluorescence measurements, which were performed later and the originally taken samples were too old, further sets of samples was used (36 samples categorized in 2 classes and 56 samples for 3 classes), in which only 11 samples were consistent with those measured by the UV-Vis, NIR a FT IR spectra.

The sensorial properties of both kinds of oil samples were evaluated in a ten point scale by the panel of experts who rated smell, taste and visual character of the sample. The presence of water was thoroughly checked since a colloidal mixture of water and residual proteins considerably diminishes the oil quality. According to sensorial quality the collected oil samples were finally divided into three categories: excellent (6.5–9.0 points), satisfactory (3.5–6.4) and bad (1.0–3.4) quality samples; for the sake of simplicity they will be denoted as best, good and worst. In this way a target categorical variable was formed, which is needed for chemometrical oil classification by discriminant techniques. In case of olive oils another classification was made by the oil type (5 varieties) as well as geographic origin (3 locations).

Chemometric methods and software

Four methods of discriminant analysis (DA) - linear and quadratic discriminant analysis (LDA, QDA), logistic regression (LR) and K-th nearest neighbour method (KNN) were used for classification of the oil samples together with artificial neural network (ANN), operating in three layer perceptron mode.

Success in the oil sample classification was evaluated using two ways of cross validation: by a leave-one-out (jackknife) method as well as using a special test set of the samples not included into training process, by which the classification model was calculated. Classification success was given as the ratio of the number of correctly classified samples to the total number of samples and expressed also in %. The calculations were performed mainly by the commercial software packages SPSS 15.0, SAS 9.1.3 and Statgraphics Plus, ver. 5.1.

Results and Discussion

Classification of two sets of oil samples (olive oil, pumpkin seed oil) was performed in three basic tasks: (1) investigation and optimisation of the conditions leading to best discrimination among the oil categories created either by sensorial oil properties, or by fruit variety determining the type of oil (only in case of olive oil samples), (2) investigation, which

spectral method leads to best discrimination of the oil samples, (3) comparison of the applied techniques of multidimensional data analysis with the aim to find the best way of classification. The first task is in fact given by the best combination of the tasks two and three.

Comparison of spectral methods used for characterizing the samples was made for both kinds of oil. While the olive oil study was completely new, the study of pumpkin oil represents continuation of our previous work (Lankmayr et al. 2004). Unlike before, in order to obtain a legitimate comparison of the used chemometrical techniques a very thorough cross validation of their results was applied in the present work. This strategy was justified mainly due to utilization of nonlinear classification algorithms, typical for artificial neural networks. A real danger in the ANN application (but existing also in discriminant techniques) is overtraining when the classification model is calculated (Kruzlicova et al 2009). Consequently, it often happens that despite an excellent classification of the samples belonging to the training data set, a poor prediction of the samples out of the training set is observed.

Further problem, which is typical for spectral data is connected to an excessive amount of descriptors (absorbances) compared to the number of studied objects (e.g. oil samples). Due to usually very high correlation between the spectral variables, their massive decrease is therefore needed. In this work a stepwise feature reduction (Khattree and Naik 2000) was applied, accessible by all software packages, which we have used. As the result only the absorbances at most informative wavelengths or wave numbers were chosen. Figure 1 illustrates a typical result of linear discrimination analysis performed with the pumpkin oil samples where 57 wavelengths were selected from the total number of originally measured 226 wavelengths. Except one sample, all others are well discriminated into the preselected categories made by sensorial quality.

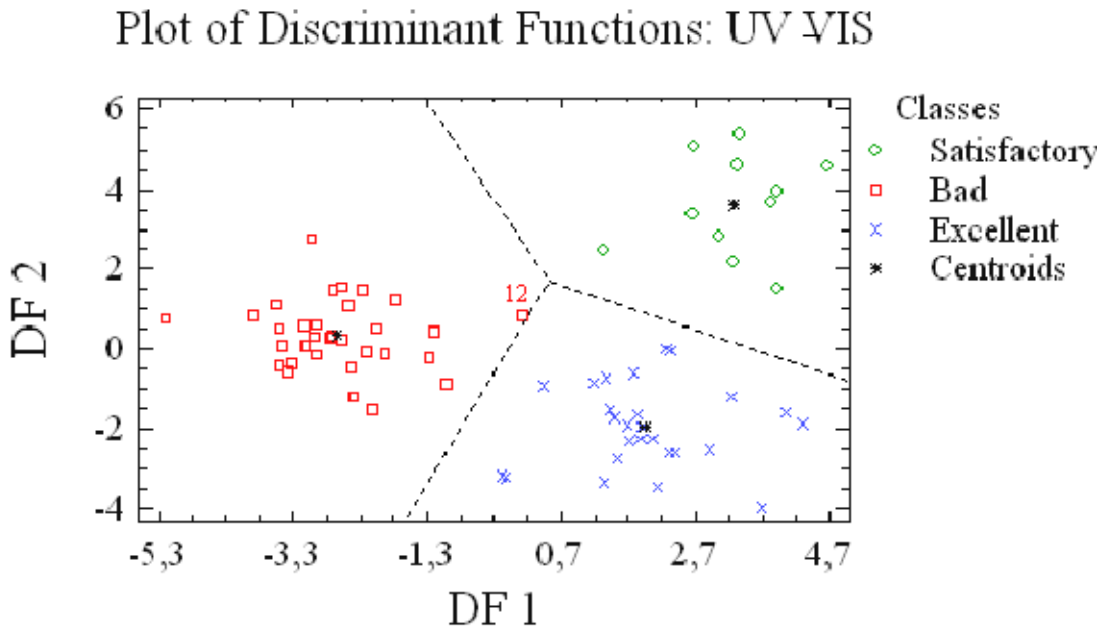

Fig. 1. Scatterplot in the plane of discriminant functions DF2 vs. DF1 illustrating classification of 70 pumpkin oil samples categorised into three classes according to the sensorial quality. UV-Vis absorbance signals were measured at 57 wavelengths optimally chosen by the backward selection technique. Sample number 12 is located at the class border.

The results of cross validation are summarized in Table 1. The optimal number of variables (absorbances at the selected wavelength or wave number) after the feature reduction process, which leads to most successful classification in cross validation, is different compared to the number of originally selected variables. Using optimal number of descriptors the validation is most successful at the expense of less successful categorization of the samples belonging to training set (which is less important). Among absorption spectral methods, the best classification of pumpkin oils was obtained by FT IR measurements but NIR and UV-Vis results were also significant and good. Nevertheless, the results of fluorimetry are still better even though the sample sets were not fully consistent with the previous cases. Further investigation of the application of the fluorescence spectra is very promising.

Table 1. Classification of pumpkin oils by sensorial quality defined by 2 or 3 classes using various spectral methods by means of linear discriminant analysis

Spectral method	2 classes			3 classes		
	No. of variables*	Success/%		No. of variables*	Success/%	
		Training	Leave-1-out		Training	Leave-1-out
UV-Vis	12 (64)	85.7	80.0	12 (64)	77.1	67.1
NIR	16 (62)	89.0	81.7	9 (61)	76.3	70.0
FTIR	16 (59)	89.0	81.7	14 (62)	78.0	72.0
Fluorimetry	38 (38)	100.0	100.0	38 (38)	98.5	93.9

* First number denotes the number of variables after feature reduction process leading to most successful classification in cross validation, the second number represents the number of originally selected variables (absorbances at the selected wavelength or wave number).

Table 2 brings the summary of results making possible to evaluate the best chemometrical technique for olive oil classification. For classification by variety the most successful was KNN technique, the second was LDA. For classification by sensorial quality the most successful technique was LDA, the second was KNN.

Table 2. Classification of olive oils by sensorial quality (3 classes) and variety (5 classes) UV-Vis absorption spectroscopy and various chemometrical techniques

Chemometrical technique	Criterion: *Variety*			Criterion: *Sensorial*		
	No. of variables*	Success/%		No. of variables*	Success/%	
		Training	Leave-1-out		Training	Leave-1-out
LDA	60	99.5	94.7	37	95.6	89.0
QDA	60	100.0	70.1	37	100.0	65.9
KNN	60	99.5	98.7	37	93.4	87.9
LR	60	100.0	-	37	100.0	-
ANN	-	-	-	37	100.0	-

* The final number of variables represents their optimal selection from originally measured 245 variables (absorbances of the digitized spectrum) by feature reduction process.

The ANN classifications were not evaluated by leave-one-out technique but by means of the test set containing of 30 % of all samples (out of the training set). The results are summarized in Table 3 containing two different parts. In the left part the composition of training and test sets is reported, the form of five best networks and corresponding classification details are in the right part. Regarding sensorial quality, the ANN results are the best of all. ANN classification by variety is not as much successful. It is worth mentioning that all results are highly significant since the classification success achieved by random choice is 33.3 % in three classes discrimination and 20.0 % when discriminating five classes.

Table 3. Success in olive oils classification (on the right side) for optimally found neural networks using two classification criteria: sensorial quality and variety

Criterion: *Sensorial*				
Training set (70 %)	Test set (30 %)	Network	Correct/All	Success/%
20 *M*	9 *M*	13×16×5	50/56	89.3
24 *K*	10 *K*	22×19×5	51/56	91.1
9 *E*	4 *E*	20×22×5	52/56	92.9
65 *N*	28 *N*	64×27×5	54/56	96.4
13 *T*	5 *T*	64×27×5	54/56	96.4
Total: 131	56			
Criterion: *Variety*				
Training set (70 %)	Test set (30 %)	Network	Correct/All	Success/%
19 best	8 best	41×11×3	21/27	77.8
41 good	17 good	41×11×3	21/27	77.8
4 worst	2 worst	41×11×3	22/27	81.5
Total: 64	27			

The total number of samples was divided using the ratio 70/30 into Training set and Test set. Further details are described in text.

Conclusion

Spectral data on olive oil and pumpkin seed oil samples were used without the need to analyze chemical substances contained in the oil samples and corresponding to the selected spectral characteristics (wavelengths or wave numbers). This procedure does not need to use standards for calibration of the spectral characteristics so that the applied approach demands more calculation effort but requires less instrumental work in the laboratory and is therefore cheaper. With regard to the performed classification FT IR spectroscopy was evaluated as the best spectral technique. The choice of the best chemometrical technique depends on the selected classification criterion, however, generally very satisfactory results were provided by artificial neural networks, KNN technique and linear discriminant analysis.

Acknowledgement

This work was supported by courtesy of the Slovak Grant Agencies (VEGA 1/3584/06 and APVV-0057-06).

References

Brescia MA, Alviti G, Liuzzi V, Sacco A (2003) Journal of the American Oil Chemists' Society 80: 945–950.

Brodnjak-Vončina D, Cencič-Kodba Z, Novič M (2005) Chemometrics and Intelligent Laboratory Systems 75: 31–43.

Christy AA, Kasemsumran S, Du YP, Ozaki Y (2004) Analytical Sciences 20: 935–940.

Dennis MJ (1998) Analyst 123: 151R–156R.

Downey G, McIntyre P, Davies AN (2003) Applied Spectroscopy 57: 158–163.

D'Imperio M, Dugo G, Alfa M, Mannina L, Segre AL (2007) Food Chemistry 102: 956–965.

Ferreira AP, Lopes JA, Menezes JC (2007) Analytica Chimica Acta 595: 120–127.

Khattree R, Naik DN (2000) Multivariate data reduction and discrimination, SAS Institute, Cary, North Carolina.

Kruzlicova D, Mocak J, Balla B, Petka J, Farková M, Havel J (2009) Food Chemistry 112: 1046–1052.

Kruzlicova D, Mocak J, Lankmayr E (2007). In: Staruch L, Morávková H (Ed) Laboralim 2007 (pp. 151–156), STU, Bratislava, ISBN 978-80-227-2628-3.

Lankmayr E, Mocak J, Serdt K, Balla B, Wenzl T, Bandoniene D, Gfrerer M, Wagner S (2004) Journal of Biochemical and Biophysical Methods 61: 95–106.

Mannina L, Dugo G, Salvo F, Cicero L, Ansanelli G, Calcagni C, Segre AL (2003) Journal of Agricultural and Food Chemistry 51: 120–127.

Mannina L, Patumi M, Proietti N, Bassi D, Segre AL (2001) Journal of Agricultural and Food Chemistry 49: 2687–2696.

Marini F, Balestrieri F, Bucci R, Magrí AD, Magrí AL, Marini D (2004) Chemometrics and Intelligent Laboratory Systems 73: 85–93.

Penza M, Cassano G (2004) Food Chemistry 86: 283–296.

Petka J, Mocak J, Farkas P, Balla B, Kovac M Journal of the Science of Food and Agriculture (2001) 81: 1533-1539.

Rezzi S, Axelson DE, Héberger K, Reniero F, Mariani C, Guillou C (2005) Analytica Chimica Acta 552: 13–24.

Vigli G, Philippidis A, Spyros A, Dais P (2003) Journal of Agricultural and Food Chemistry 51: 5715–5722.