Available online at www.sciencedirect.com
SCIENCE DIRECT®

Analytica Chimica Acta 552 (2005) 13–24

ANALYTICA
CHIMICA
ACTA

www.elsevier.com/locate/aca

Classification of olive oils using high throughput flow ^{1}H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks[☆]

Serge Rezzi[a], David E. Axelson[b], Károly Héberger[a,c,*],
Fabiano Reniero[a], Carlo Mariani[d], Claude Guillou[a]

[a] *European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Physical and Chemical Exposure Unit, BEVABS T.P. 740, I-21020 Ispra (VA), Italy*
[b] *MRi_Consulting, 8 Wilmot St., Kingston, Ont., Canada K7L 4V1*
[c] *Chemical Research Center, Hungarian Academy of Sciences, Institute of Chemistry, H-1525 Budapest, P.O. Box 17, Hungary*
[d] *Stazione Sperimentale Olii e Grassi, Milano, Italy*

Received 27 May 2005; received in revised form 24 June 2005; accepted 19 July 2005

Abstract

The combination of ^{1}H NMR fingerprinting with multivariate analysis provides an original approach to study the profile of olive oil in relation to its geographical origin and processing. The present work aims at illustrating the relevance of ^{1}H NMR fingerprints for assessing the geographical origin and the year of production for olive oils from various Mediterranean areas. Multivariate (chemometric) techniques are able to filter out the most relevant information from a spectrum, e.g. for a classification. Principal component analysis (PCA) was carried out on the ∼12,000 variables (chemical shifts) and four data sets were defined prior to PCA. Linear discriminant analysis (LDA) of the first 50 PC's was applied for classification of olive oil samples (97 or 91) according to the geographic origin and year of production. The data analysis has been carried out with and without outliers, as well. Variable selection for LDA was achieved using: (i) the best five variables and (ii) an interactive forward stepwise manner. Using LDA on the external validation sets the correct classification varied between 47 and 75% (random selection), and between 35 and 92% (Kennard–Stone selection (KS)) depending on geographic origin (country) and production years. A similar success rate could be achieved using partial least squares discriminant analysis (PLS DA). The success rate can be considerably improved by using probabilistic neural networks (PNN). Correct classification by PNN varied between 58 and 100% on the external validation sets. Other chemometric techniques, such as multiple linear regression, or generalized pair-wise correlation, did not give better results.

Keywords: NMR; Authenticity; Multivariate methods; Linear discriminant analysis; Principal component analysis; Chemometrics; Artificial neural networks

1. Introduction

Olive oil is a very important agricultural product for most of the countries of the Mediterranean basin. According to recent estimations on olive oil markets, the European union (EU) produces 78% of the world production followed by Turkey (6%), Syria (6%), Tunisia (3%) and Morocco (2%). The EU also dominates world consumption (73%) while the rest of the production is absorbed by USA (8%), Japan, Canada and Australia (1% for each). Within the EU, Spain, Italy and Greece are mainly responsible for olive oil production with around 865, 590 and 375 thousands of tons reached in 2003, respectively. The high added value of olive oil makes its control an important goal for EU producers and consumers. There is thus a need in developing reliable analytical methods to ensure compliance with labeling, i.e. the control of geographical origin giving also support to the

---

[☆] The paper was originally presented at the 9th International Conference on Chemometrics in Analytical Chemistry, 20–23 September 2004, Lisbon.
[*] Corresponding author. Tel.: +36 1 438 11 03; fax: +36 1 438 11 43.
E-mail address: heberger@chemres.hu (K. Héberger).

0003-2670/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2005.07.057

denominated protected origin (DPO) policy, and the determination of the genuineness of the product by the detection of eventual adulterations. Despite the wide range of modern analytical tools, the development and the validation of methods able to tackle authenticity controls can be very challenging due to the broad range of olive oil qualities (extra virgin, virgin, lampante, refined, etc.), the natural variability of the product depending on pedoclimatic and process conditions, and the various cultivars.

Chemometric methods for classification of olive oils have become of increasing importance [1–11]. The principal aim is differentiation among cultivars and geographic origin as well as unraveling adulteration, and authentication. Similarly, different analysis methods, and different chemometric tools are applied in various laboratories.

Marini et al. have authenticated extra virgin olive oil varieties using linear discriminant analysis (LDA) and back propagation artificial neural networks (BP-ANN) [1]. Cichelli et al. have distinguished different cultivars measuring pigments (such as chlorophyll) and using multivariate analyses [2]. Christy et al. have detected and quantified adulteration in olive oil by near-infrared spectroscopy and using chemometric techniques: principal component analysis (PCA), partial least squares regression (PLS) and applied methods for data pretreatments such as multiplicative signal correction [3]. Aranda et al. have measured triglycerides, total and 2-position fatty acid composition by HPLC and achieved 90% correct classification using PCA and LDA in differentiating Spanish olive oil cultivars [4]. Casas et al. have determined the fatty acid composition of eight different regions within Extremadura (Spain) and achieved 89% correct classification on the prediction set using LDA [5]. Brescia et al. have distinguished virgin and pure olive oils by excitation-emission matrix fluorescence spectroscopy using PCA and parallel factor analysis (PARAFAC) [6]. Downey et al. have studied the visible and NIR spectra of extra virgin oils and classified the 94% of the samples correctly according to the geographic origin and using PLS, LDA and k-nearest neighbor method (KNN) [7]. Jakab et al. have applied HPLC-MS and MALDI-TOF-MS methods and distinguished vegetable oils with 97% correct classification using LDA [8,9]. The potential of multivariate statistics as a promising tool to authenticate and classify food products according to their geographical origin and variety has also been reviewed critically [10].

Several authors have also used nuclear magnetic resonance (NMR) spectroscopy for the characterization of olive oils [11–17]. The amount of information available in a NMR spectrum and the easy sample preparation render this spectroscopic technique very attractive for the assessment of olive oil composition. Among the numerous analytical methods developed during the last decade, ^{1}H NMR spectroscopy of oils has demonstrated the possibilities for providing information, in a fairly rapid manner, on lipid classes, unsaturation level, molar fractions of specific fatty acids (linoleic, linolenic acids) and several minor compounds (sterols, squalene, terpenes, oxidized products, etc.) [11–14]. Applying LDA to ^{1}H NMR generated data Fauhl et al. have separated hazelnut, sunflower and olive oils (as well as their mixtures) [11]. On the other hand, ^{13}C NMR gives unique information on the position of fatty acids on glycerol and the stereochemistry of unsaturation [13,14]. More recently, ^{31}P NMR applied to oils previously derivatized with a phosphorous reagent showed the ability to do quantitative analysis of their phenolic compounds and diglycerides [15,16]. The combination of ^{1}H NMR fingerprinting with multivariate analysis provides an auspicious approach to study the profile of olive oils in relation to its geographical origin and its history from extraction to processing and storage. Until now, most of the published studies on the application of NMR to olive oil report the use of this technique in a manner that one may interpret as a targeted analysis, since components to be assessed have to be defined a priori.

The present work aims at using the NMR profile of olive oils taken as a whole in combination with multivariate statistics for the classification of olive oil samples from various Mediterranean areas. We report a methodological approach based on the high throughput acquisition of NMR profiles statistically processed in a holistic way in order to provide information regarding the geographical origin and the year of production.

2. Methods

2.1. ^{1}H NMR

Olive oil samples were provided by Stazione Sperimentale Olii e Grassi (Milano, Italy).

NMR spectra were measured in the stop-flow mode using a Bruker (Rheinstetten, Germany) DRX-500 instrument operating at 500.13 MHz for ^{1}H observation using a 3 mm single cell ^{1}H/^{13}C inverse detection flow probe with an active volume of 60 μl. Sample transfer from the 96-well plate to the NMR flow probe used a Gilson (Middleton, WI, USA) XL233 automatic sample handling system interfaced to the NMR data system for control and timing [17,18].

For each sample, oil was diluted in 1 mL of deuterated chloroform. ^{1}H NMR spectra were registered at 300 K using 64 K data points and the standard Bruker pulse sequence (zgps). The residual $CHCl_3$ signal was suppressed with an off-resonance presaturation generated by a shape pulse of 80 ms duration at 7.28 ppm. Eight scans were accumulated for each spectrum. The profiles of oils were acquired in 1.30 min under experimental conditions allowing quantitative measurements. The total experimental time per sample including injection, transfer into the NMR flow-cell, optimization of the field homogeneity, data acquisition and the inter-sample washing procedure takes around 5 min. Under such conditions, one could perform the ^{1}H NMR analysis of around 300 oils per day. The FIDs were zero filled and Fourier transformed applying a line broadening of 0.1 Hz. All spectra were phased manually and baseline correction was applied.