Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 148 (2015) 131–137

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

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

Review Article

Determination of geographical origin of alcoholic beverages using ultraviolet, visible and infrared spectroscopy: A review

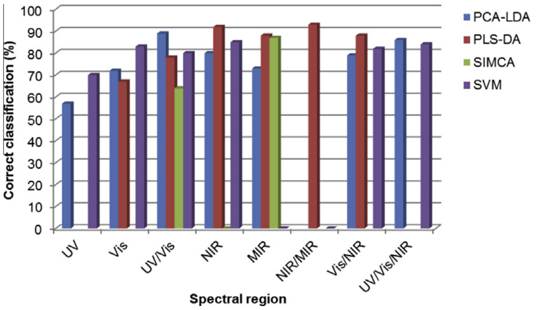
Veronika Uríčková *, Jana Sádecká

Institute of Analytical Chemistry, Faculty of Chemical and Food Technology, Slovak University of Technology, Radlinského 9, 812 37 Bratislava, Slovak Republic

HIGHLIGHTS

- Geographical classification of wines and distilled drinks.
- UV, Vis, NIR and MIR spectroscopic techniques were used.
- PCA–LDA, SVM, SIMCA and PLS-DA was applied.
- Average correct classification higher than 82%.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 January 2015
Received in revised form 10 March 2015
Accepted 27 March 2015
Available online 3 April 2015

Keywords:
Beverages
Ultraviolet and visible spectroscopy
Infrared spectroscopy
Pattern recognition methods
Geographical origin

ABSTRACT

The identification of the geographical origin of beverages is one of the most important issues in food chemistry. Spectroscopic methods provide a relative rapid and low cost alternative to traditional chemical composition or sensory analyses. This paper reviews the current state of development of ultraviolet (UV), visible (Vis), near infrared (NIR) and mid infrared (MIR) spectroscopic techniques combined with pattern recognition methods for determining geographical origin of both wines and distilled drinks. UV, Vis, and NIR spectra contain broad band(s) with weak spectral features limiting their discrimination ability. Despite this expected shortcoming, each of the three spectroscopic ranges (NIR, Vis/NIR and UV/Vis/NIR) provides average correct classification higher than 82%. Although average correct classification is similar for NIR and MIR regions, in some instances MIR data processing improves prediction. Advantage of using MIR is that MIR peaks are better defined and more easily assigned than NIR bands. In general, success in a classification depends on both spectral range and pattern recognition methods. The main problem still remains the construction of databanks needed for all of these methods.

Contents

---

* Corresponding author.
E-mail address: veronika.urickova@stuba.sk (V. Uríčková).

http://dx.doi.org/10.1016/j.saa.2015.03.111
1386-1425/

Introduction

The European Union Regulation establishes the rules for protecting designations of origin and geographical indications for foodstuffs. A protected designation of origin (PDO) indicates that the product must be both produced and processed within the defined geographic area. To attain protected geographical indication (PGI), the product must be produced or processed in the geographical area (either or both). Therefore, the PGI allows greater flexibility than the PDO. In many cases, the EU PDO/PGI system works parallel with the system used in the specified country.

A powerful method for determining the geographical origin of food products is pattern recognition methods of the data provided by analytical instruments. Chromatographic methods are relatively expensive, time-consuming and require highly skilled operators. Common approach is multi-elemental analysis followed by stable-isotope ratio methods. The best results come from combining the two techniques; however, implementation of this strategy is quite difficult in routine analysis. Recently, attention has focused on the development of non-invasive and non-destructive instrumental techniques such as ultraviolet (UV), visible (Vis), near infrared (NIR) and mid infrared (MIR) spectroscopy [1–8].

Large amounts of spectral data, containing useful analytical information, noise, variabilities, uncertainties and unrecognized features, are usually obtained from spectroscopic instruments. Thus, pattern recognition methods are required to extract as much relevant information from spectral data as possible. Pattern-recognition methods are subdivided into supervised and non-supervised methods. Non-supervised methods do not require any a priori knowledge about the group structure in the data, but instead produces the grouping, i.e. clustering, itself. This type of analysis is often very useful at an early stage of an investigation and can be performed with simple visual techniques, such as hierarchical cluster analysis (HCA) or principal component analysis (PCA) [2]. When employing HCA, the original data are separated into a few general groups, each of which is further divided into still smaller groups until finally the individual objects themselves remain. The results are presented in the form of dendograms to facilitate the visualization of sample relationships [2].

PCA is usually the first step in spectroscopic data exploration. The aims of performing a PCA are two-such as fold. Firstly, PCA reduces the dimensions of the spectral dataset by explaining a large part of the variance using synthetic factors, called principal components (PCs). Therefore, the whole range of wavelengths can be compressed into the first few PCs. Secondly, PCA performed on spectral data makes it possible to draw similarity maps of the samples and to get spectral patterns. Classification of objects is done by constructing similarity maps of the samples, using PCs chosen by the researcher. The spectral patterns corresponding to the PCs provide information about the characteristic peaks which are the most discriminating for the samples observed on the similarity maps [9]. With supervised methods, the number of groups is known in advance and representative samples of each group are available. This information is used to develop a suitable discrimination rule or discriminant function with which new, unknown samples can be assigned to one of the groups. The commonly used supervised methods are linear discriminant analysis (LDA), support vector machine (SVM), soft independent modeling of class analogy (SIMCA) and partial least squares discriminant analysis (PLS-DA). LDA is concerned with determining the so-called discriminant functions as linear combinations of the descriptors which best separate the groups according to minimization of the ratio of within group and between-group sum of squares. LDA requires that the number of variables (wavelengths) must be smaller than the number of samples in each group. Consequently, large spectral datasets with few samples cannot be analyzed using LDA. Combining LDA with a PCA overcomes this problem [10].

SVM maps the sample data with specific kernel functions to a higher dimensional feature space to linearize the boundary and generate the optimal separating hyperplane. There are number of kernels that can be used in SVM models. These include linear, polynomial, radial basis function and sigmoid. SVM is still effective in cases where the number of variables is greater than the number of samples. If the number of variables is much greater, the method is likely to give poor performances [11].

In PLS-DA, the standard PLS algorithm can be used and group labels can be given for the dependent vector. In the two-group case, usually the values of the dependent variable are given 1 for one group and 0 or -1 for the other group and a PLS1 algorithm is used. In case of more than 2 groups, dummy variables are defined and a PLS2 algorithm is used. PLS-DA is well suited to deal with a much larger number of variables than samples and with correlated variables [12].

SIMCA builds a distinct confidence region around each class after applying PCA. New measurements are projected in each PCs space that describes a certain class to evaluate whether they belong to it or not. The classification of a sample in one or several classes, or in none of them, is possible with SIMCA, while discriminant methods only permits to classify a sample in a unique class. Another advantage of SIMCA is that there is no restriction on the number of variables [13].

This review covers UV, Vis, NIR and MIR spectroscopy applications for determining geographical origin of alcoholic beverages, namely wines and distilled drinks. Table 1 provides a summary of the application of spectroscopic and pattern recognition methods done country by country, while Table 2 shows a prediction results for models developed using different spectral regions, regardless of geographical origin. Regarding supervised methods (Tables 1 and 2), LDA was based on the first PCs, SIMCA was used on PCA classes, while PLS-DA and SVM were applied over the original spectra. Table 3 provides the most important variables or regions related with the differences in the data sets obtained by PCA.

Geographical origin of wines

UV/Vis/IR spectra of wines

The UV/Vis absorption spectra of wine samples are similar in profile, but differ in the absorption at the measurement wavelengths. The highest absorbances are observed at 202 and 230 nm. These are associated with carboxyl groups of organic acids. Moreover, the UV/Vis spectra contain information regarding phenolic compounds, e.g., benzoic acids (235–305 nm),