Food Chemistry 184 (2015) 214–219

Contents lists available at ScienceDirect

Food Chemistry

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

Modeling excitation–emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety

Silvana M. Azcarate [a], Adriano de Araújo Gomes [b], Mirta R. Alcaraz [c], Mário C. Ugulino de Araújo [b], José M. Camiña [a,*], Héctor C. Goicoechea [c,*]

[a] Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa, and Instituto de Ciencias de la Tierra y Ambientales de La Pampa (INCITAP), Av. Uruguay 151, (6300) Santa Rosa, La Pampa, Argentina
[b] Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA), Universidade Federal da Paraíba, CCEN, Departamento de Química, Caixa Postal 5093, CEP 58051-970 João Pessoa, PB, Brazil
[c] Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Cátedra de Química Analítica I, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral-CONICET, Ciudad Universitaria, Santa Fe (S3000ZAA), Argentina

ARTICLE INFO

Article history:
Received 18 October 2014
Received in revised form 21 March 2015
Accepted 23 March 2015
Available online 28 March 2015

Keywords:
White wine
Excitation–emission matrices
SIMCA
U-PLS-DA
N-PLS-DA
SPA–LDA

ABSTRACT

This paper reports the modeling of excitation–emission matrices for classification of Argentinean white wines according to the grape variety employing chemometric tools for pattern recognition. The discriminative power of the data was first investigated using Principal Component Analysis (PCA) and Parallel Factor Analysis (PARAFAC). The score plots showed strong overlapping between classes. A forty-one samples set was partitioned into training and test sets by the Kennard–Stone algorithm. The algorithms evaluated were SIMCA, N- and U-PLS-DA and SPA–LDA. The fit of the implemented models was assessed by mean of accuracy, sensitivity and specificity. These models were then used to assign the type of grape of the wines corresponding to the twenty samples test set. The best results were obtained for U-PLS-DA and SPA–LDA with 76% and 80% accuracy.

1. Introduction

Wine is a fermented alcoholic beverage containing various compounds of different type, with polyphenols being one of the most important components which influence their quality parameters (Fernández-Pachón, Villaño, García-Parrilla, & Troncoso, 2004). Polyphenols contribute to the organoleptic characteristics, such as color, astringency and bitterness and also exert antimicrobial and antioxidant properties. In addition, other substances such as amino acids, anthocyanin and ethanol, are present in wine (Godoy-Navajas, Aguilar-Caballos, & Gómez-Hens, 2015).

Authenticity and commercial value of wines is often linked to its geographical origin and certain countries or regions are known for producing excellent wines of high commercial value (Šelih, Šala, & Drgan, 2014). Due to its composition and worldwide availability, the controlled denomination of origin (CDO) is usually required to demonstrate the provenance of wines. In the same way, due the cost and/or production volume of several varieties of white wines in comparison to others, the control of grape variety is usually required to avoid the adulteration or fraud in fractioned or selected wines, which can involve the use of high cost analytical methods and well trained analysts (de Villiers, Alberts, Lynen, Crouch, & Sandra, 2003; Wang, Geil, Kolling, & Padua, 2003).

According to data from 2013, Chardonnay, Sauvignon blanc and Torrontés were the main white wines varieties produced in Argentina, while the main producer provinces were Rio Negro Mendoza, San Juan and Salta. Torrontés has been the most exported white wine variety over the last year, followed to Chardonnay and Sauvignon blanc in second and third place, respectively: USA, Canada, Russia and United Kingdom were the main importers of these wines (Instituto Nacional de Vitivinicultura, 2014).

Many efforts have been devoted to the development of new analytical methods for the quality control of wines worldwide (Briz-Cid, Figueiredo-Gonzalez, Rial-Otero, Cancho-Grande, & Simal-Gandara, 2015; González-Álvarez, Noguerol-Pato,

* Corresponding authors.
E-mail addresses: jcaminia@exactas.unlpam.edu.ar (J.M. Camiña), hgoico@fbcb.unl.edu.ar (H.C. Goicoechea).

http://dx.doi.org/10.1016/j.foodchem.2015.03.081
0308-8146/

González-Barreiro, Cancho-Grande, & Simal-Gándara, 2013), with molecular absorption spectroscopy and gas or liquid chromatography coupled to mass spectrometry detection (GC–MS and LC–MS) being the most common instrumentation (de Villiers, Alberts, Tredoux, & Nieuwoudt, 2012).

Recently, Markechová and collaborators employed fluorescence excitation–emission matrices (EEM) and Parallel Factor Analysis (PARAFAC) for the determination of brandy adulteration with mixed wine spirit (Markechová, Majek, & Sádecká, 2014). However, the use of fluorescence spectroscopy for quality control of wines regarding the authenticity of the grape type, geographic origin, brand, or fingerprint has been scarcely explored. For example, Dufour et al., employed both excitation and emission spectra for French and German wine classification (Dufour, Letort, Laguet, Lebecque, & Serra, 2006), and Sádecká et al. reported the classification of brandies and wine distillates using total luminescence and synchronous fluorescence spectra (Sádecká, Tóthová, & Májek, 2009), but only first-order data and algorithms were used in an unsupervised manner in both cases. On the other hand, Ariado-Rodrigues and collaborators employed EEM and PARAFAC for fingerprinting of red wines (Airado-Rodríguez, Galeano-Díaz, Durán-Merás, & Wold, 2009) and quality control in the wine industry (Airado-Rodríguez, Durán-Merás, Galeano-Díaz, & Wold, 2011). Interestingly, and different from the present report, an unsupervised approch was used in both reports. Finally, it should be noted that a study of the possibilities of multi-way fluorescence linked to PARAFAC and to classify the different sherry vinegars accordingly to their ageing was presented by Callejón et al. (2012).

In the present work, a methodology is presented for exploring the feasibility of discrimination of Argentinean white wines varieties by using second-order data obtained recording excitation–emission matrices in samples with minimum pre-treatment. Different algorithms for pattern recognition were implemented: PCA, PARAFAC, soft independent modeling of class analogy (SIMCA), discriminant analysis by unfolded partial least squares (U-PLS-DA), discriminant analysis by multi-way partial least squares (N-PLS-DA) and successive projection algorithm (SPA–LDA). All of them were evaluated by using independent sample sets. As will be shown, the best results in terms of accuracy were obtained for U-PLS-DA and SPA–LDA, while the latter algorithm allowed building extremely parsimonious models, i.e. models that accomplish a desired level of explanation or prediction with as few predictor variables as possible.

2. Experimental

2.1. Samples

Forty-one different varieties of commercial white wine samples from four wine-producing provinces of Argentina (Mendoza, San Juan, Salta, and Rio Negro) were included in this study: 12 Torrontés wine (from Mendoza, San Juan, Salta, and Río Negro), 14 Chardonnay wine (from Mendoza and San Juan), and 15 Sauvignon Blanc wine (from Mendoza, San Juan, and Río Negro). Wines samples were selected from the 2011 to 2013 vintages. The alcoholic content ranged from 12.2% to 13.8% vol/vol ethanol. All were purchased from a local supermarket.

2.2. Apparatus

All spectrofluorimetric measures were acquired on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Waldbronn, Germany) using a 1x1 cm quartz fluorescence cell, xenon flash lamp and CaryEcplise software package to control the instrument, data acquisition and data analysis. Excitation–emission matrices were saved in ASCII format, and transferred to a PC for subsequent manipulation.

2.3. Experimental procedure

Prior to the measurements, samples were equilibrated at room temperature. Immediately after opening the bottle, 10 mL of each white wine sample was filtered with 0.45 μm Nylon filter. Each sample was prepared in triplicate and measure using the fluorescence spectrophotometer.

2.4. Data modeling

All calculations carried out in this work were done in Matlab environment (MATLAB, 2010). PARAFAC and N-PLS were applied using the *N-way Toolbox* (Andersson & Bro, 2000) developed BrO and available in http://www.models.life.ku.dk/algorithms. The Kennard–Stone (KS) Algorithm (Harrop Galvao et al., 2005) and SPA–LDA (Carreiro Soares, Gomes, Rodrigues Galvao Filho, Ugulino Araujo, & Harrop Galvao, 2013) were written by the authors. U-SIMCA and U-PLS-DA calculations were conducted in graphical interface *classification-toolbox 3.1* available in http://michem.disat.unimib.it/chm/download/softwares.htm.

2.5. Chemometric tools

The chemometric algorithms applied in the present work, and their corresponding references are listed and briefly at the following:

(a) SIMCA is a method of pattern recognition widespread and used to solve classification problems in chemistry (Wold, 1976). Briefly, the SIMCA method assumes the measurement values for groups of similar samples have a uniform distribution and malleable, and this modeling is based on Principal Component Analysis (PCA) (Bro & Smilde, 2014; Forina, Oliveri, Lanteri, & Casale, 2008).

(b) U- and N-PLS-DA was originally proposed for multivariate calibration, and then used in classification problems. Their mathematical foundations have been described in the literature (Indahl, 2014). In essence both U-PLS and N-PLS for discriminant analysis (Barker & Rayens, 2003; Ouertani, Mazerolles, Boccard, Rudaz, & Hanafi, 2014) are equal for calibration purposes.

(c) The SPA–LDA algorithm is aimed at selecting a subset of variables with small collinearity and suitable discriminating power for use in classification problems involving $Q \geqslant 2$ different classes. For this purpose, it is assumed that a training set of N objects with known class labels is available to guide the variable selection process. In the case of spectroscopic data, for example, each object consists of a spectrum recorded over K wavenumbers (or wavelengths) (Carreiro Soares et al., 2013).

(d) Validation of classifiers: as in calibration, in classification the performance of the model for the training sample set or an independent test set can be accessed using some figures of merit. In that way, for a classification model, the most common figures of merit are accuracy (*AC*), sensitivity (*S*) and specificity (*SP*) (Lavine, 2009, chap. 3). With the purpose of illustration, consider a two-class classification problem (A_1 positive class and A_2 negative class), in which a_1 and a_2 are the quantities of objects A_1 assigned to A_1 and quantities of objects A_2 assigned to A_2, respectively. On the other hand b_1 are objects de A_1 assigned to A_2 and b_2 are objects de A_2