LWT - Food Science and Technology 65 (2016) 565–571

Contents lists available at ScienceDirect

LWT - Food Science and Technology

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

Linear and non-linear modeling to identify vinegars in blends through spectroscopic data

José S. Torrecilla*, Regina Aroca-Santos, John C. Cancilla, Gemma Matute

Department of Chemical Engineering, Faculty of Chemistry, University Complutense of Madrid, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 6 May 2015
Received in revised form
23 July 2015
Accepted 8 August 2015
Available online 14 August 2015

Keywords:
Vinegar
Blends
Identification
Artificial neural network
Spectroscopy

ABSTRACT

The identification of vinegars produced from six different raw materials (red wine, white wine, cider, apple, molasses, and rice) in blends has been accomplished through their UV–vis spectra and different mathematical models: partial least squares discriminant analysis (PLS-DA) and artificial neural networks (ANNs). The registered spectra were mathematically treated following a linear (PLS-DA) approach and a non-linear one (ANN) based on multilayer perceptron models with different training functions. The average correct classification rate of a series of comparable internal validations was around 55% and 90%, for the PLS-DA and the ANN models respectively, which heavily favors the non-linear approach. Therefore, an accurate chemometric tool with the ability to detect specific vinegars in mixtures in an inexpensive and straightforward fashion has been designed and optimized.

1. Introduction

Vinegar is vastly used in Europe and Asia as a condiment, acidulant, and food preservative (De la Haba, Arias, Ramírez, López, & Sánchez, 2014; Chen, Ding, Cai, & Zhao, 2012). Recent studies have shown that the particular composition of vinegar, especially regarding organic acids, polyphenols, and other phenolic compounds may have a positive effect for human health due to their antioxidant activity (Budak et al., 2011; Nakamura et al., 2010). Therefore, the consumption of vinegar may play a role in the prevention of several diseases like hypertension, cardiovascular illnesses, cancer, diabetes, and others (Budak, Aykin, Seydim, Greene, & Guzel-Seydim, 2014; Naziroglu et al., 2014; Wu et al., 2013; Budak et al., 2011; Nakamura et al., 2010).

The final characteristics of a vinegar mainly depend on the presence and amount of these compounds which are fundamentally related to the raw material used to produce the vinegar and the acetification system employed (De la Haba et al., 2014; Guan, Zhao, Lin, & Zou, 2014). Even though wine vinegar is the most traditionally consumed in Mediterranean countries (De la Haba et al., 2014), nowadays vinegar is produced from several fruits and other food-related materials, such as rice or molasses. As it was said before, depending on the raw material used, the antioxidant effects of the vinegar will vary, as the phenolic profile is different for each one (Guan et al., 2014). Therefore, the mixture of different vinegars would cause a modification on the final composition of the product. Although this blending does not necessarily involve a safety risk to human health, it does constitute a potential fraud (Saiz-Abajo, González-Sáiz, & Pizarro, 2005; Saiz-Abajo, González-Sáiz, & Pizarro, 2004). Consequentially, these practices are prohibited by law, for instance, in countries like Spain (BOE, 2012).

These facts result in the need to control the quality and origin (both botanical and geographical) of vinegar, as well as develop analytical techniques able to determine these frauds or illegal practices (Saiz-Abajo et al., 2005, 2004). Thereby, a series of different methods and techniques have already been proposed to reach a reliable evaluation of the quality of vinegar, such as headspace microextraction coupled with gas chromatography (Pizarro, Esteban-Díez, Sáenz-González, & González-Sáiz, 2008), high performance liquid chromatography (Cerezco et al., 2008), near-infrared spectroscopy (De la Haba et al., 2014), fluorescence spectroscopy (Castellón et al., 2012), and electronic nose systems (Guan et al., 2014). Among all of them, spectroscopic techniques have demonstrated to be an excellent alternative to other more time- and reagent-consuming methods, offering great results in the discrimination of vinegars (De la Haba et al., 2014; Castellón et al., 2012; Saiz-Abajo et al., 2004, 2005). In this sense, UV–vis and near-infrared absorption spectroscopy are simple and straightforward

* Corresponding author.
E-mail address: jstorre@ucm.es (J.S. Torrecilla).

http://dx.doi.org/10.1016/j.lwt.2015.08.027
0023-6438/

techniques that can be employed to register the absorbance bands of the different compounds (organic acids, polyphenols, flavonoids, and others) that might differ between vinegars from diverse botanical origins.

However, when these instrumental techniques are applied, a lot of information is generated. In order to extract the most representative and desired data, several mathematical procedures can be used, such as principal component analysis (Guan et al., 2014; Chen et al., 2012), linear discriminant analysis (Guan et al., 2014) and partial least squares analysis (De la Haba et al., 2014; Guan et al., 2014) among others. In many cases, this last method, partial least squares discriminant analysis (PLS-DA), has provided reliable and accurate models, and for this reason it has been selected to mathematically treat the data in this research (Yang et al., 2014; Izuka & Aishima, 1997). Nevertheless, the characteristics these linear models offer are sometimes insufficient to manage the huge loads of data obtained that in fact may be non-linearly related. In these cases, more complex, non-linear statistical tools like artificial neural networks (ANNs) are required. These non-linear algorithms are mathematical systems able to determine non-linear relations that exist between variables (Tanajaura da Silva, Filardi, Pepe, Chaves, & Santos, 2015).

Although spectroscopic techniques and ANNs have been already applied to vinegar discrimination (Guan et al., 2014; Chen et al., 2012), their use as a tool to distinguish between the components of blends of two vinegars with different botanical origins, is a new approach. Combining these elements would permit establishing a simple and straightforward procedure to identify fraudulent blends of vinegars or monitor their composition.

Even though both UV–vis and NIR spectroscopy could be useful for this task, we have strictly employed the former approach. Therefore, to summarize, the main goal of this work is the identification of distinct vinegars present in binary blends in different proportions, employing UV–vis absorption spectroscopy coupled with PLS-DA and with ANNs. The use of these two mathematical approaches will allow the comparison of their results in order to obtain a new tool capable of controlling the quality of vinegar without spending elevated amounts of time and/or chemical reagents and also without the need of sophisticated and expensive equipment.

2. Material and methods

In the following subsections, the reagents, instruments, and mathematical procedures utilized are explained.

2.1. Vinegars employed and sample preparation

A total of six commercial vinegars were employed, of which all were purchased in local stores and within their expiration dates. Each one possesses a specific botanical origin: red wine, white wine, cider, apple, molasses, and rice.

The binary blends were prepared mixing known volumes of the vinegars, varying their composition in 5% v/v increases (from 0% to 100%), attaining 21 samples for every possible mixture, which led to a total of 315 samples. These samples were diluted with deionized water in a 20:1 proportion (water: vinegar) to avoid the saturation of the spectrophotometer.

2.2. UV–vis absorption spectroscopic measurements

The absorption spectra of the samples were acquired between 200 and 800 nm using a UV–vis spectrophotometer (Varian Cary 50 Conc), with a scanning speed of 4800 nm/min. The diluted vinegar blends were each measured three times in a quartz cuvette of 1 cm length path. The averages of these measurements were used as the representative spectrum of the corresponding sample.

In order to treat the obtained spectra to discriminate between the components of the blends, it was necessary to mathematically extract the underlying relevant chemical information. The selected procedure, based on the calculation of the area under the curve (AUC) of the absorption bands, is furtherly explained in the "Results and discussion" section.

The analysis of the spectra was carried out using OriginPro v8.0724 (B724) and Matlab 2013b software packages.

2.3. Database

The final database consists of a total of 315 spectra, each one corresponding to a specific blend of two vinegars. The database includes both compositional information of the mixtures prepared and the chemical data contained in their absorption spectra, which are shown in the Supplementary information section.

2.4. Mathematical models

2.4.1. PLS-DA

Partial Least Squares Discriminant Analysis is a regression used to determine dependent binary variables from a group of independent predictor variables, and it is a classic statistical tool which is used to create a linear regression by the projection of both observed and predicted variables into a new space (somewhat related to principal components analysis) (Barker and Rayen, 2003; Palancar, Aragon, Miguens, & Torrecilla, 1996).

The resulting linear models have the next structure (Eq. (1)):

$$y = c + \sum_{n} a_n x_n \tag{1}$$

Where $\boldsymbol{c}$ is the constant of the model (regression coefficient for the intercept), $\boldsymbol{x}$ and $\boldsymbol{y}$ represent the independent and dependent variables, respectively, and $\boldsymbol{a_n}$ is the regression coefficient for the independent variables.

In order to evaluate the spectroscopic information obtained from the vinegar blends, six independent PLS models will be designed, using data extracted from the spectra as the independent variables of the models.

2.4.1.1. Validation of the PLS-DA. The generalization capability of the PLS-DA was tested with four internal validations containing 10 data points each. These 40 data points were not used to create the six lineal models mentioned above. A threshold of 0.5 was set to show the results as "0s-and-1s" vectors (the results which are lower than 0.5 become "0s" (absence of that vinegar type) and the ones higher than 0.5 become "1s") of six elements (e.g. [0, 1, 0, 0, 0, 1]). These simulation vectors were then compared with the real ones to finally obtain the identification rate of the model.

2.4.2. ANN

The other mathematical tool selected to determine the vinegars that compose the prepared blends is based on ANNs, due to the advantages that their non-linear interpolating ability provides (Nedic et al., 2014). However, in order to offer accurate and reliable estimations, ANNs need to be trained accordingly to fulfill a specific task. There are different training or learning possibilities that can be summarized in three main types: supervised, non-supervised, and mixed learning approaches (Cartwright and Curtenau, 2013). When the first ones are employed, it is a requirement to provide both input and target information, so the network can compare the attained results with the already known target to improve the