Self-organizing maps based on chaotic parameters to detect adulterations of extra virgin olive oil with inferior edible oils ☆

José S. Torrecilla [a,*], John C. Cancilla [a], Gemma Matute [a], Pablo Díaz-Rodríguez [a], Ana I. Flores [b]

[a] Department of Chemical Engineering, Faculty of Chemistry, University Complutense of Madrid, 28040 Madrid, Spain
[b] Regenerative Medicine Group, Research Center, Instituto de Investigación Hospital 12 de Octubre, Avda. Córdoba s/n, 28041 Madrid, Spain

ABSTRACT

A nonlinear algorithm based on chaotic parameters (CPs) has been employed to determine the nature of different output signals obtained from UV–vis spectrophotometer (UV) measurements. These signals come from UV scans of adulterated samples of extra virgin olive oil (EVOO) with refined olive oil or refined olive pomace oil, or from pure samples of EVOO with white random or sinusoidal white random noises. The data collected from this equipment was used to calculate CP values. Then, a self-organizing map was used to detect different types of signals. Using this method, the signals can be identified and classified into five groups depending on their type, the percentage of noise added, and the concentration of adulterant agents, with a misclassification rate of less than 1.3%.

1. Introduction

Presently, the adulteration of foods is becoming a more and more usual fraud in Mediterranean countries, especially in highly priced products with a strong demand, such as extra virgin olive oil (EVOO). The adulteration of EVOO with cheaper and insalubrious ingredients could have serious health implications for consumers. A clear and dramatic example of this was the Spanish toxic oil syndrome resulting from the consumption of aniline denaturalized rapeseed oil, which has affected at least over 20,000 people, causing serious illness and in some cases even death (Torrecilla, 2010; Dong-Sun et al., 1998; Mildner-Szkudlarz and Jelen, 2008). This is why the detection of adulterated EVOO is an important issue that attracts the interest of many research groups. Due to this growing attention, a large number of cases of adulterated olive juice have been detected recently.

Although the EVOO quality can be checked by chemical indices and organoleptic assessment (Torrecilla, 2010), there is no single analytical index that determines the protected denomination of origin, their geographical origin, or even the olive variety (Marini et al., 2004). Because of this, to quantify the adulteration of EVOO,

the concentrations of chemicals present in the oils (acids, sterols, polyphenols, and others) and/or their physicochemical properties (density, refractive index, saponification value, and so on) should be determined and, in nearly all cases, this data must be mathematically treated (Torrecilla, 2010; Commission Regulation, 2003; Food and Agriculture organization of the United Nations, 2003–2005). Statistically, to extract the most relevant information (mathematical relations between independent and dependent variables) from huge databases generated from analytical or measuring equipment, powerful statistical techniques are required. Some of the most important techniques are based on linear algorithms such as principal component analysis (Mildner-Szkudlarz and Jelen, 2008; Dourtoglou et al., 2003; Torrecilla et al., 2008) or multivariate regression techniques (Peña et al., 2005). On the other hand, others are based on nonlinear algorithms such as artificial neural networks (NNs), self-organizing maps (SOMs), and algorithms based on chaotic parameters (CPs) (Torrecilla 2010; Torrecilla et al., 2011a; Torrecilla et al., 2011b; Torrecilla et al., 2009a).

In the last few years, clear examples where SOMs have successfully been applied is in the food field. For instance, to classify carrots by their colors, which in the end relate to their quality (Janaszek and Trajer, 2011) or to distinguish different soluble coffees using information obtained from electronic noses (Bona et al., 2012). When looking into what has been done with extra virgin olive oil, Torrecilla et al. successfully designed two different types of maps to detect and differentiate samples of extra virgin olive oil

adulterated with other edible oils using their chemical composition (Torrecilla et al., 2009a). Therefore, these facts led towards using nonlinear algorithms to analyze complex systems rather than employing linear algorithms like principal component analysis for instance. The nonlinear algorithm of choice is based on SOMs, and they will be tested in order to define the best possible model to distinguish noisy signals from clean signals, and from signals from pure or adulterated EVOO samples.

Since adulteration of EVOO is not only an economical fraud, but can also be involved in human health issues, a reliable, fast, simple, and cheap method to detect these undesirable adulterating agents is required. Currently, chemometric tools based on nonlinear algorithms have presented the most successful results (Torrecilla et al., 2009a; Marini et al., 2007). For example, due to the fact that tools based on CPs can detect slight variations in initial experimental conditions (Vayenas and Pavlou, 2001; Torrecilla et al., 2009b), these models can determine the present adulteration even in low concentrations of low grade olive oils (Torrecilla et al., 2011a). Undoubtedly, one of the most important enemies in the interpretation of databases, and therefore, in identifying and classifying adulterated foods, is the presence of noise in the signals from analytical equipment. The determination of adulterants will be approached by the use of algorithms based on CPs in order to distinguish and classify signals from adulterated samples and signals from pure samples with noise, Fig. 1.

The main objective of this work is to create a SOM based on chaotic parameters to group signals, with specific common traits, into five sets. These sets are basically described in two groups: (i) UV–vis signals from pure extra virgin olive oil and their adulterated samples, with different quantities of refined olive oil and refined olive–pomace oil (less than 2.1 w/w%); (ii) and two types of digital noisy signals added (white random and sinusoidal white random signals) changing the noisy/signal rate from 0% to 10%.
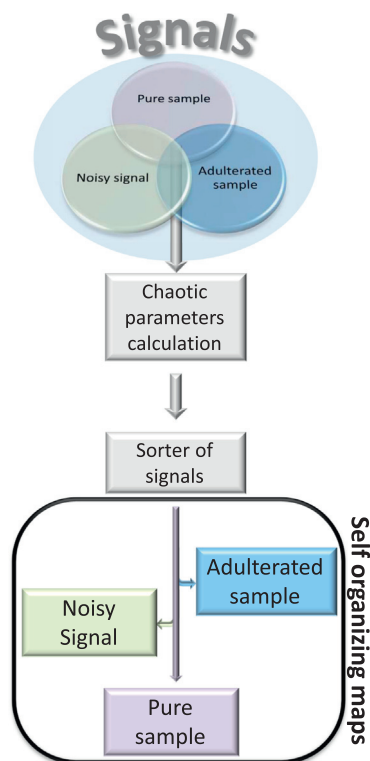

Fig. 1. Calculation diagram to determine the nature of the different signals and main results obtained.

2. Materials and methods

2.1. Instrumentation and oil samples

A Varian Cary 1E UV–vis spectrophotometer was employed for absorbance measurements between 190 and 900 nm using quartz cells of a path length of 1 cm. The expanded uncertainty in the experimental measurements was lower than 0.03. All stock solutions were prepared using an AG 245 Mettler Toledo analytical balance (precision of 0.01 mg).

The samples of refined olive oil (ROO) and refined olive–pomace oil (ROPO) used have been provided by different Spanish companies, Table 1. The botanical origin and quality of all samples of extra virgin olive oil were guaranteed by the supplier. All were stored in the dark at room temperature until the day of analysis, which was prior to their expiration date. To estimate and detect the adulteration of EVOO with other low cost olive oils, mixtures containing EVOO and refined olive oil or refined olive–pomace oil were prepared (*vide infra*). Following the procedure shown in the Official Journal of the European Union (Commission Regulation (EC) No. 640/2008, Annex IX), all samples were prepared and diluted in iso-octane ($C_8H_{18} \geqslant 99.5\%$ purity, from MERCK).

2.2. Independent variables used

To detect other low grade olive oils in the EVOO and distinguish these from other noisy signals (signal of pure EVOO with different percentages of white random noise and sinusoidal white random noise) lag-k autocorrelation coefficients have been calculated by UV–vis scans of pure and adulterated EVOO samples. These calculations are mentioned as "chaotic parameters calculation" in Fig. 1.

Lag-k autocorrelation coefficient ($R_{\Delta\lambda}$) are real numbers that measure linearly how strongly on average each data point is correlated with the wavelength lag ($\Delta\lambda$), Eq. (1) (Kant and Schreiber, 2005). These are the ratios of autocovariance/variance of the data. In general, $R_{\Delta\lambda}$ is between 1 ($\Delta\lambda = 0$) and 0 (large $\Delta\lambda$) (Drazin, 1992; Sprott, 2003):

$$R_{\Delta\lambda} = \frac{\sum_{n=1}^{N-k}(X_n - \bar{X})(X_{n-k} - \bar{X})}{\sqrt[2]{\sum_{n=1}^{N-k}(X_n - \bar{X})\sum_{n=1}^{N-k}(X_{n-k} - \bar{X})}} \qquad (1)$$

where k, X, $\bar{X}$, and N represent wavelength lag, the absorbance set of the measurements by UV–vis spectrophotometer, their average, and the total number of data sets, respectively. Given that the λ values range between 0 and 650, groups of $\Delta\lambda = 50$ nm (14 parameters) have been selected. For instance, in the case of $\Delta\lambda = 100$, which would result from the sum of the two first groups, throughout the work, its $R_{\Delta\lambda}$ has been referred to as R_{100}.

2.3. Learning, verification, and validation sample

Every data point of the learning and verification samples is composed of fourteen autocorrelation parameters (Eq. (1)). These were calculated from the UV–vis scans from pure and binary mixtures composed of EVOO and ROO (0–2.1 w/w%) or ROPO (0–2.1 w/w%). As an example, in Fig. 2, experimental data from noisy signals of pure and adulterated EVOO samples are shown. In addition, the

Table 1
Botanical origin and brand of oil samples used.

Type of olive oil	Brand
Extra virgin olive oil	Aceites Borges Pont SAU
Refined olive oil	KOIPE, SOS Cuétara SA
Refined olive–pomace oil	DIA