JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Self-Organizing Maps and Learning Vector Quantization Networks As Tools to Identify Vegetable Oils

JOSÉ S. TORRECILLA,* ESTER ROJO, MERCEDES OLIET, JUAN C. DOMÍNGUEZ, AND FRANCISCO RODRÍGUEZ

Departamento de Ingeniería Química, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040-Madrid, Spain

Self-organizing map (SOM) and learning vector quantification network (LVQ) models have been explored for the identification of edible and vegetable oils and to detect adulteration of extra virgin olive oil (EVOO) using the most common chemicals in these oils, viz. saturated fatty (mainly palmitic and stearic acids), oleic and linoleic acids. The optimization and validation processes of the models have been carried out using bibliographical sources, that is, a database for developing learning process and internal validation, and six other different databases to perform their external validation. The model's performances were analyzed by the number of misclassifications. In the worst of the cases, the SOM and LVQ models are able to classify more than the 94% of samples and detect adulterations of EVOO with corn, soya, sunflower, and hazelnut oils when their oil concentrations are higher than 10, 5, 5, and 10%, respectively.

KEYWORDS: Kohonen neural network; adulteration; competitive neural networks; extra virgin olive oil; seeds oil

INTRODUCTION

The adulteration of food products with cheaper and more available substitutes is a typical worldwide problem that has existed for centuries. In recent decades, due to the high price of extra virgin olive oil (EVOO), an appreciable incidence of adulteration has been detected. The substitution or adulteration of EVOO with cheaper ingredients is not only an economic fraud but may also on occasions have severe health implications for consumers, an example being the Spanish toxic oil syndrome (*1, 2*).

Extra virgin olive oil is subject to two types of adulteration. In the first group, the EVOO is blended with low-grade olive oils (olive−pomace oil, virgin olive oil obtained by a second centrifugation of the olives, or refined olive oils). The second consists of mixing EVOO with other cheaper similar products such as seed oils (hazelnut, sunflower oil, maize, corn, soybean, palm, etc.) (*2*−*4*). Given the chemical similarities of EVOO and hazelnut oil, this adulteration is difficult to detect, especially when its concentration is less than 20% (*2, 3*).

Although EVOO quality can be checked by chemical indices and organoleptic assessment (*5*), there is no single analytical index to determine the protected denomination of origin, their geographical origin, or even the olive fruit variety (*6*). That is why the determination of the adulteration of EVOO or the classification of the vegetable oils, concentrations of chemicals present in the oils (acids, sterols, polyphenols, etc.), and their physicochemical properties (density, refractive index, saponification value, etc.) should be quantified (*5, 7*). These can be determined using a wide number of chemometric tools which are based on techniques such as nuclear magnetic resonance spectroscopy (NMR) (*8*), Fourier transform Raman spectroscopy (*9*), gas chromatography (GC), high performance liquid chromatography (HPLC) (*10*), fluorescence spectroscopy (*11*), etc.

To extract the most relevant information from those huge databases of the characteristics, composition, and concentration of chemicals of each vegetable oil, statistical techniques are required. Some of the most important techniques are linear algorithms such as principal component analysis (PCA) (*2, 4*), multivariate regression techniques (*3, 8*), or nonlinear algorithms such as artificial neural networks.

Two of the most used competitive neural networks algorithms used are self-organizing maps (SOMs) and learning vector quantization networks (LVQs) models. SOM models have been successfully used to identify crude oils (*12*) and the plant communities in Pangquangou Nature Reserve, North China (*13*), as well as to plan irrigation strategies in places where water resources are scarce (*14*). LVQ models have been used to recognize facial expression (*15*) and to classify nucleic acid and protein sequences (*16*). For instance, SOM models have been used in the discrimination of wines (*17*) or between brands of milk (*18*). To the best of our knowledge, in the vegetable oil field, SOM models have hardly been used and the LVQ model applications are even scarcer. One of the few applications of SOM models was developed by Brodnjak-Voncinaet et al., who

* Corresponding author. Tel.: +34 91 394 42 40; fax: +34 91 394 42 43; e-mail: jstorre@quim.ucm.es.

10.1021/jf803520u CCC: $40.75
Published on Web 03/06/2009

used a Kohonen neural network to cluster vegetable oil samples (*10*). Marini et al. used this type of model to select the adequate training, verification, and validation samples used by a feed forward neural network in order to resolve binary blends of monocultivar Italian olive oils (*19*). Given the successful results achieved in other scientific fields and the recognition capability of groups with similar characteristics (using SOM models) (*20, 21*) and that even these groups can be in addition classified in target classes defined by the user (using LVQ models) (*21, 22*), linear (PCA) and nonlinear algorithms (SOM or LVQ models) have been applied here in the vegetable oil area.

The objective of this work is the application of a principal component analysis technique, self-organizing map and learning vector quantification networks to identify 13 edible vegetable oils (hazelnut, sunflower, corn, soybean, sesame, walnut, rapeseed, almond, palm, groundnut, safflower, coconut, and extra virgin olive oils) and detect adulterations of EVOO with seed oils (corn, soya, sunflower and hazelnut oils) using only their saturated fatty (mainly palmitic and stearic acids), oleic, and linoleic acid concentration values.

MATERIAL AND METHODS

Edible Vegetable Oils. To design and optimize both SOM and LVQ models a database of saturated fatty, oleic, and linoleic acid concentrations of 13 types of edible vegetable oils (192 samples) was used. Their mean repeatability and reproducibility values are less than 2 and 2.5%, respectively (*8*). In order to test the performance of optimized SOM and LVQ models related to the classification of vegetable oils and detection of adulteration of EVOO with seeds oils, another seven bibliographical references (263 samples) were employed (*1, 6, 8, 10, 23*−*25*).

Principal Component Analysis. Principal component analysis is a classical unsupervised technique based on linear algebra. It involves a mathematical procedure, which transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components (PCs). The principal components are linear combinations of the original variables. The first principal component accounts for most of the variability in the data, and each succeeding component accounts for as much of the remaining variability as possible. This linear transformation has been widely used in data analysis, in exploratory tools to uncover unknown trends in the data, compression, etc. (*26*).

In this work, the PCA technique has been applied to check and select the most important information from the aforementioned database. Then, using the selected information, two different nonlinear models were designed and applied (vide infra). Principal component analysis was carried out by SPSS software version 15.0.

Self-Organizing Maps. Self-organizing maps or the Kohonen neural network is one of the most interesting topics in the competitive neural network field (*20*). SOM models can learn to detect irregularities and correlations in their input and adapt their future responses to that input accordingly; that is, they are able to recognize groups with similar characteristics (*20, 21*). The architecture of SOM models is shown in **Figure 1**. Every circle and arrow represent a neuron and weight, respectively; that is, there are as many weights as arrows and the number of neurons is equal to the product of the width and length of the competitive layer. In this layer, each neuron has as many weights as the input descriptors (SFA, oleic, and linoleic acids), **Figure 1**. Every neuron is represented by a vector of weights.

Given that self-organizing maps classify input data according to how they are grouped in the input space, along the leaning process, in order to adequately represent all input data, its weights are optimized. As every neuron is represented as a weight vector, during this process, the neurons look for the best place to represent the whole input database. The learning process of the SOM involved two steps viz. ordering and tuning phases (*21*). In the former, the ordering phase learning rate (OLr) and neighborhood distance (ND) are decreased from both that rate and the maximum ND between

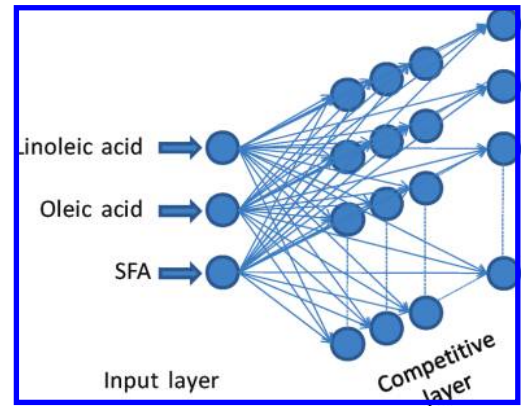

Figure 1. Schematic diagram of self-organizing map model.

two neurons to the tuning phase learning rate (TLr) and the tuning phase neighborhood distance, respectively. The ordering phase lasts for a given number of steps (named ordering phase steps, OP). In the latter, the learning rate is decreased much more slowly than in the ordering phase while the ND remains constant. Therefore, the number of epochs for the tuning phase of the SOM learning process should be much larger than the number of steps in the ordering phase. Given that the SOM model is a competitive and unsupervised neural network, their weight optimization can be summarized in five stages: (i) assign random values to the weights; (ii) a data set from the learning sample is presented to the SOM; (iii) the neuron with the least Euclidean distance between its weights and data set (*D*) is selected, eq 1. It is called the winning neuron. (iv) The weights of the selected neuron are optimized so that they become more similar to the input vector, eq 2; (v) the weight of the neighborhood neurons are also optimized but with proportionally less Euclidean distance to wining neuron eq 3. The process is repeated iteratively. When the whole database has been presented to the SOM an epoch has finalized (*12*). Once the SOM has been trained, this is able to extract the relevant information in order to classify new input vectors (which are interpolated in the learning range).

$$D_j = ||X - W_j|| = \sqrt{\sum_i^S (X_i - W_{ij})^2}; j = 1, 2, ..., N \quad (1)$$

$$W_j(n) = W_j(n-1) + Lr \cdot [X(n-1) - W_j(n-1)] \quad (2)$$

$$W_j(n) = (1 - Lr) \cdot W_j(n-1) + Lr \cdot X(n-1) \quad (3)$$

where *W*, *n*, *X*, *Lr*, *N* and *S* are the weights, iteration of a given epoch, input vector, learning rate, number of weights of the SOM, and number of data set of the learning sample, respectively (*14, 21*). The SOM model used in this work was designed using Matlab version 7.01.24704 (R14) (*21*).

Learning Vector Quantization Networks. LVQ models can classify any set of input vectors, not only linearly separable sets of input vectors. The only requirement is that the competitive layer must have enough neurons, and each class must be assigned enough competitive neurons (*22*). LVQ models classify input vectors into target classes by using a competitive layer to find subclasses of input vectors, and then, combining them into the target classes defined by the user. Therefore, LVQ networks consist of two layers viz., unsupervised (competitive) and supervised (linear) layers (*21*).

The competitive layer learns to classify input vectors in much the same way as the competitive layers of self-organizing maps (vide supra). The linear layer transforms the competitive layer's classes into target classifications defined by the user. The linear layers have one neuron per class (*21*). The LVQ model used in this work was designed using Matlab version 7.01.24704 (R14).

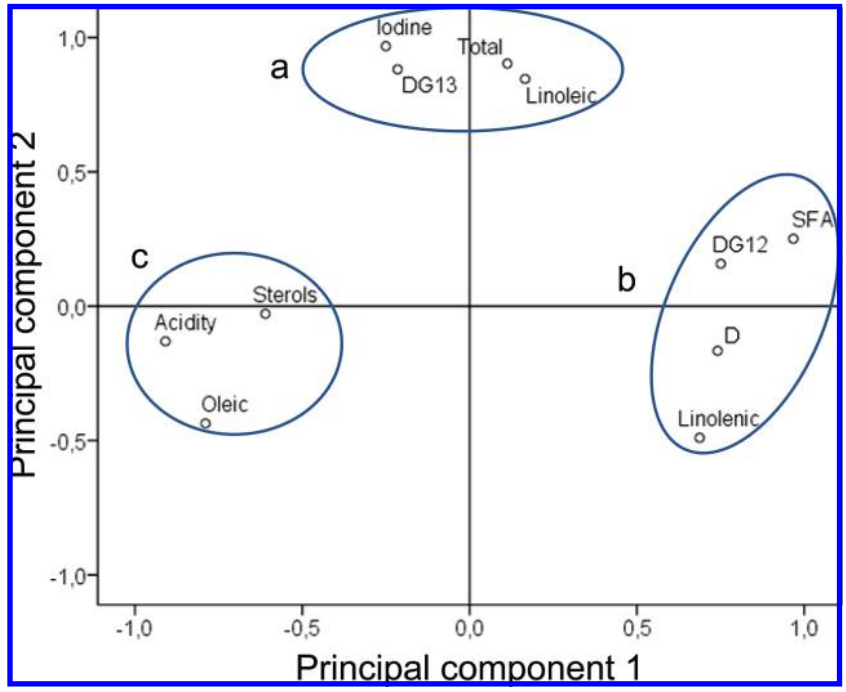

Figure 2. Principal component score plot.

Table 1. Correlation Coefficient (R^2) between 13 Concentrations and Index for 192 Oil Samples (*8*)[a]

	DG12	DG13	total	*D*	sterols	acidity	linolenic	linoleic	oleic	SFA	iodine
DG12	1										
DG13	0.006	1									
total	0.118	**0.903**	1								
D	0.444	0.075	0.000	1							
sterols	0.057	0.006	0.001	0.132	1						
acidity	0.370	0.007	0.031	0.264	0.170	1					
linolenic	0.114	0.179	0.069	0.195	0.135	0.408	1				
linoleic	0.026	0.316	0.373	0.008	0.024	0.092	0.125	1			
oleic	0.301	0.015	0.125	0.163	0.265	0.679	0.060	0.316	1		
SFA	0.607	0.000	0.114	0.476	0.016	**0.904**	0.282	0.138	0.757	1	
iodine	0.002	**0.917**	0.728	0.124	0.170	0.010	0.398	0.594	0.048	0.000	1

[a] DG12, 1,2-diglycerides; DG13, 1,3-diglycerides; total, ratio of 1,2-diglycerides to the total diglycerides; *D*, total diglycerides; sterols, sterols concentration; linolenic, linolenic concentration; linolenic, linolenic concentration; oleic, oleic concentration; SFA, saturated fatty acids; iodine, iodine value.

RESULTS AND DISCUSSION

Learning, Verification and Validation Samples. The learning, verification, and validation samples were used to carry out the optimization, internal, and external validation of the models, respectively. The optimization and internal validation processes of the SOM and LVQ models have been developed using data from the literature (*8*). These data consist of values of the acidity, iodine value, ratio of 1,2-diglycerides to the total diglycerides and the concentrations of total sterols, total diglycerides, 1,2-diglycerides, 1,3-diglycerides, saturated fatty (SFA), oleic, linolenic, and linoleic acids determined by analysis of the respective ^{1}H NMR and ^{31}P NMR spectra (*8*). These properties were calculated in 192 samples corresponding to 13 types of vegetables oils (hazelnut, sunflower, corn, soybean, sesame, walnut, rapeseed, almond, palm, groundnut, safflower, coconut, and extra virgin olive oils). In order to test the adulteration detection capability of SOM and LVQ models, the aforementioned properties were also measured and calculated in 28 samples which consist of the mixture of EVOO/corn oil (6 samples), EVOO/soya (6 samples), EVOO/sunflower oil (6 samples), and EVOO/hazelnut oil (10 samples) (*8*).

Finally, in order to test the competence of the optimized models, an external validation process was carried out using other bibliographical references (*1*, *6*, *10*, *23*−*25*).

The learning, verification, and validation samples are composed of data that characterize the classification process. The applicability domain of the data used in the learning, internal and external validation processes was evaluated following the calculation process described in the literature (*27*, *28*), which consists of determining the compounds with cross-validated standardized residuals greater than three standard deviation values. The widest dispersion is presented in the SFA data (<3.6 standard deviation), where the coconut and almond oils present the highest and the lowest ranges of values, respectively. As these oils form two of the 13 types of oils, these samples are not considered as outliers.

In order to select the most important variables, the underlying information of the database used should be revealed. With this objective, the PCA technique has been applied. As can be seen in **Figure 2**, there are three independent groups of important chemicals (a, b, and c). At least one chemical of each group should be selected to represent adequately the input information. In addition, the SFA (palmitic and stearic acids), oleic, and linoleic acids are present in all oil types used here and in most edible vegetable oils (*5*, *7*). As the selection of SFA, oleic, and linoleic acids compounds fulfilled the aforementioned condition, these compounds were selected.

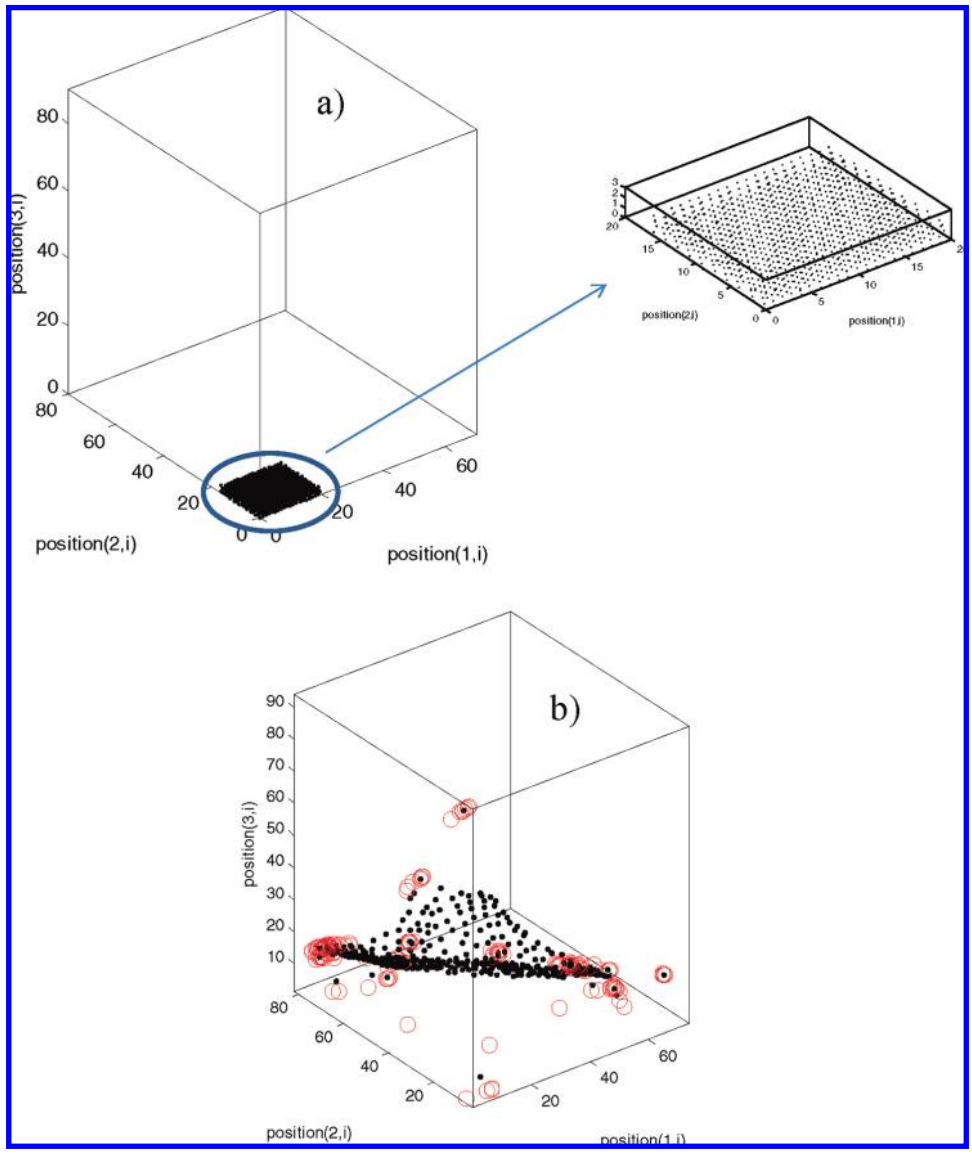

Figure 3. Position of the neurons of the self-organizing map along its learning process: **(a)** initial distribution (hexagonal topology); **(b)** final positions of the neurons (red open dots) and the learning and verification samples (○) (*8*).

Table 2. Database Used to Carry out the Internal Validation of SOM and LVQ Models (*8*)

		no. of misclassifications	
type of oil	no. of samples	SOM model	LVQ model
almond	3	0	0
coconut	2	0	0
corn	2	0	0
EVOO	5	0	0
groundnut	5	0	0
hazelnut	**3**	**1**	**1**
palm	2	0	0
rapeseed	2	0	0
safflower	2	0	0
sesame	2	0	0
soya	2	0	0
sunflower	4	0	0
walnut	4	0	0
total	**38**	**1**	**1**

In order to remove the possible data redundancy in the input data, mutual correlation coefficients between 1,2-diglycerides (DG12), 1,3-diglycerides (DG13), the ratio of 1,2-diglycerides to total diglycerides (total), acidity, iodine value, and fatty acid composition (SFA) were calculated, **Table 1**. Correlation coefficients higher than 0.9 were found in three cases viz. DG13 and total, DG13 and iodine value and acidity and SFA. These results are in agreement with those found using the PCA technique (all three characteristics belonging to group a, **Figure 2**). Finally, as acidity and SFA belong respectively to groups b and c and there is a mathematical correlation between them ($R^2 = 0.904$), only one should be chosen. Therefore, there are no redundancies in the input data.

All three samples (learning, verification, and validation) have the same format. These have as many rows as variables necessary to characterize the process (concentrations of SFA, oleic and linoleic acids) and the same number of columns as the number of vectors to describe the system to be studied. Whole database has been distributed randomly into learning (80%) and verification (20%) samples. The learning, verification, and validation samples dimensions are 154 × 3, 38 × 3, and 263 × 3, respectively. The dimension of the validation sample is studied in more detail below.

Self-Organizing Map Optimization. A nonlinear mapping method was used to classify the aforementioned 13 types of oils. The output neurons were arranged in three different

Table 3. Databases Used to Carry out the External Validations of SOM and LVQ Models

type of oil	no. of samples	ref	no. of misclassifications: SOM model	no. of misclassifications: LVQ model
coconut	2	Lee et al., 1998 (*1*)	0	0
	1	Gan et al., 2005 (*23*)	0	0
corn	3	Lee et al., 1998 (*1*)	0	0
	1	Gan et al., 2005 (*23*)	0	0
	2	Brodnjak-Voncina et al., 2005 (*10*)	0	0
EVOO	27	Spangenberg et al., 1998 (*24*)	2	3
	15	Marini et al., 2004 (*7*)	1	1
	137	Dais et al., 2007 (*25*)	6	7
	1	Gan et al., 2005 (*23*)	0	0
groundnut	1	Spangenberg et al., 1998 (*24*)	0	0
hazelnut	1	Spangenberg et al., 1998 (*24*)	0	0
	1	Gan et al., 2005 (*23*)	0	0
palm	1	Gan et al., 2005 (*23*)	0	0
rapeseed	2	Lee et al., 1998 (*1*)	0	0
	1	Gan et al., 2005 (*23*)	0	0
	10	Brodnjak-Voncina et al., 2005 (*10*)	0	0
sesame	10	Lee et al., 1998 (*1*)	0	0
	1	Gan et al., 2005 (*23*)	0	0
sunflower	2	Spangenberg et al., 1998 (*24*)	0	0
	1	Gan et al., 2005 (*23*)	0	0
	13	Brodnjak-Voncina et al., 2005 (*10*)	2	2
walnut	1	Spangenberg et al., 1998 (*24*)	0	0
	1	Gan et al., 2005 (*23*)	0	0
total	**235**		**11**	**13**

Table 4. Database Used To Determine the Capacity to Detect Adulteration of EVOO by SOM and LVQ Models (*8*)

EVOO + type of oil (%)	no. of misclassifications: SOM model	no. of misclassifications: LVQ model
	corn	
5	1	1
10	0	0
15	0	0
20	0	0
35	0	0
50	0	0
	soya	
5	0	0
10	0	0
15	0	0
20	0	0
35	0	0
50	0	0
	sunflower	
5	0	0
10	0	0
15	0	0
20	0	0
35	0	0
50	0	0
	hazelnut	
5	1	1
10	0	0
15	0	0
20	0	0
35	0	0
50	0	0

topological grids viz. grid (G), hexagonal (H), and random (R) topologies. In addition, three different methods to calculate the distances were used, viz. link (the number of links, or steps, which must be taken to reach the neuron under consideration, L), Euclidean (vide supra, E) and Manhattan (vide infra, M) distances (*21*). To classify the edible oil samples in the most reliable way possible, both topologies and distance were combined and the best pair was selected. Following the manufacturer's indications, throughout the selection process, all other parameters were maintained constant and fixed by default (OLr, OP, TLr, ND and the dimension of the network was equal to 0.9, 1000, 0.02, 1, and 5 × 8, respectively) (*21*). In most of the nine possible combinations, between 10 to 12% of EVOO and hazelnut oil samples from verification sample were misclassified as hazelnut, sunflower, and EVOO oils. The combination of hexagonal topology and Manhattan distance, eq 4, was the best combination, misclassifying only less than 5% of the EVOO samples (verification sample). Therefore, this combination was selected.

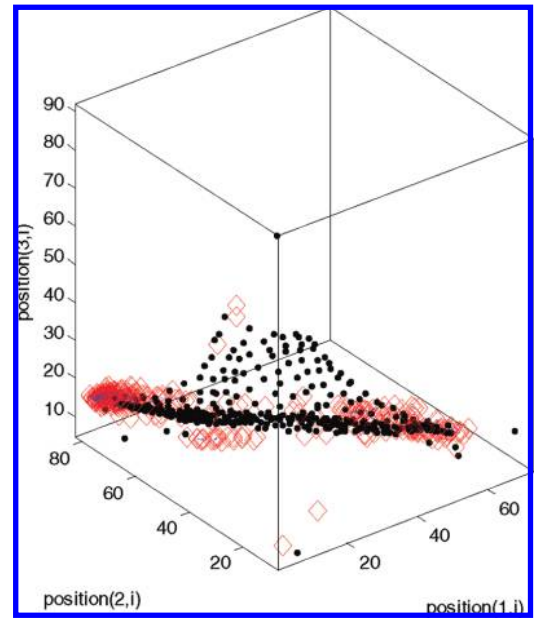

Figure 4. Final positions of the neurons (•) and the validation sample (red open diamonds) (235 samples) (*1, 6, 10, 23−25*).

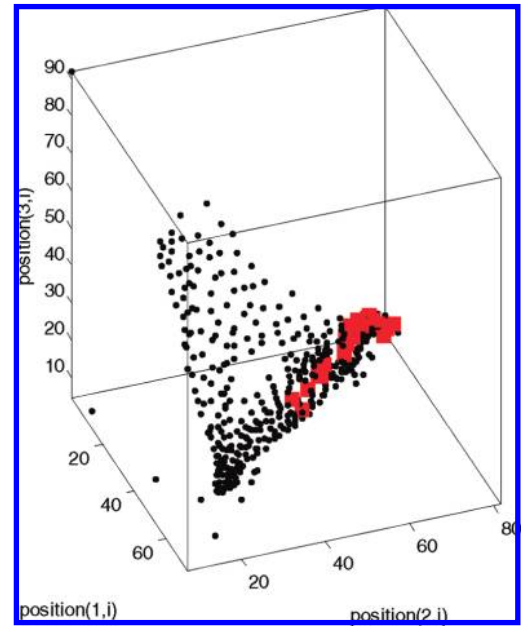

Figure 5. Final positions of the neurons (•) and the validation sample (solid red squares) corresponding to EVOO adulteration with seed oil (28 samples) (*8*).

$$M_j = \sum_i^S |X - W_j| \qquad (4)$$

Once the topology and distance of the SOM were selected, the dimension of the network was optimized. Networks of sizes ranging from 18 × 18 to 26 × 26 were tried (*12*). From 20 × 20 up to 26 × 26 dimensions, the performance of these maps (number of misclassifications) was somewhat similar. Therefore the 20 × 20 dimension was selected.

With the selected topology, distance and the optimized network dimension, the parameters of the SOM model were optimized by a Central Composite Design 2^5 + star experimental design, where the variables analyzed were OLr (from 0.1 to 1), OP (from 500 to 1500), TLr (from 0.01 to 0.03), ND (from 0.5 to 1.5) and the number of epochs in the learning process (from 10000 to 30000 epochs). The response of the experimental design was the number of incorrect classifications of the oil samples from the verification sample. In order to reach the least number of misclassifications, the optimum parameter values have been fixed at 0.1, 1500, 0.01, 0.5, and 30000 to OLr, OP, TLr, ND and the number of epochs necessary in the learning process, respectively.

The weight vectors used to classify the input data are shown in **Figure 3**. As can be seen, throughout the learning process, the competitive neurons have been adequately distributed in whole three-dimensional space (from **Figure 3a**,**b**), and therefore, every data set would be classified by one or a group of neurons, **Figure 3b**. The number of misclassifications depending on the oil type is shown in **Table 2**. The only misclassification consisted of a hazelnut sample which was classified as EVOO. This mistake is based on their similar chemical composition (*29*). As the verification sample was formed and fixed at 38 samples and only three of them correspond to this oil, the percentage of misclassified hazelnut oil samples is high (>34%). Given that a wider database is required to study this point in more detail, the SOM model's capacity to carry out classification will be tested in the external validation process. Nevertheless, overall, the number of misclassifications is less than 3%. Therefore, in the light of these results, this type of map is able to classify nearly all input data used in the internal validation.

Learning Vector Quantification Network Optimization. The LVQ model consists of unsupervised and supervised layers. The former is a competitive network similar to the SOM model. This part, formed by the input and hidden neurons (also called the number of competitive neurons), is fixed at three and 20 × 20, respectively, as has been described in the Self-Organizing Map Optimization section. The supervised layer consists of an output layer with 13 neurons, one for each oil type.

In the light of these considerations, the only LVQ parameter to optimize is the learning rate (Lr). In the optimization process, Lr was tested between 1×10^{-3} to 1. Taking into account that the minimum number of misclassifications is required, the best Lr value was 0.01. This is in agreement with the literature (*21*). As was mentioned above, although the number of misclassifications of hazelnut oil with respect to the whole verification sample is less than 3%, the LVQ capability to classify samples of hazelnut oil cannot be assumed. This point will be studied in the external validation.

Application of Optimized SOM and LVQ Models to Others Databases. In order to validate the optimized SOM and LVQ models, the external validation process has been carried out using six bibliographical databases (235 samples), **Table 3** (*1*, *6*, *10*, *23*−*25*). In addition, analytical values from mixtures of EVOO and edible oils (28 samples) were also used to test the reliability of the aforementioned models in the EVOO adulteration detection (*8*).

As mentioned previously, in order to guarantee the reliability of the classifications carried out by these models, the applicability domain has been evaluated selecting the compounds with cross-validated standardized residuals greater than three standard deviation values (*27*, *28*). In this evaluation applied to validation sample, no response outlier was found. Then, the validation sample was input into the SOM and LVQ models, **Figure 4**. As can be seen, overall, only one input data set is not adequately represented by the optimized SOM and two competitive neurons are not used to classify some input data sets. As can be expected, the number of misclassifications in the external validation process is higher than those in the internal validation. Nevertheless, as the misclassification percentage is less than 5%, the tested models are able to classify vegetable oils adequately. Given that the LVQ model is partially based on the SOM model, the results are similar. Although the misclassification percentage is slightly higher (<5.5%), the LVQ model has the advantage in that the classifications are organized by the user. To recap, the models tested are able to classify all oil types, even the hazelnut oil samples.

Finally, the capacity to detect adulteration of EVOO with seeds oils has been tested (*8*). As can be seen in **Figure 5**, all 28 samples are adequately represented by the optimized SOM or LVQ models. In particular, using these methods, the adulteration with corn, soya, sunflower, and hazelnut oils can be detected when their respective concentrations are higher than 10, 5, 5, and 10%. In the case of hazelnut adulteration, these results are similar to those published in the literature (*11*). But here, thanks to the nonlinear models applied, the required information is notably less. It is certain that the result would have been notably improved if a SOM or LVQ model had been specifically designed for oil mixtures and had been focused on determining adulteration. In particular, if the optimization of both models focus on detecting one type of adulteration (e.g., with hazelnut oil) in a narrower range of concentrations (less than 10%), the results will be notably improved.

Conclusions. In this work, two mathematical approaches based on self-organizing maps and learning vector quantifications networks models have been designed to classify samples in 13 classes of vegetable oils and detect adulterations of extra virgin olive oil. These points have been developed using only three of the chemicals present in most vegetable oils. To test these models internal and external validations were carried out. In the internal and external validations, less than 3 and 5.5% of the samples were misclassified. The adulteration of EVOO with corn, soya, sunflower, and hazelnut oils was detected when their concentration was higher than 10, 5, 5, and 10%, respectively. In the light of these results, both models are adequate to classify these studied samples in 13 types of vegetable oils. Although the results could be improved by specifically designed models for the adulteration databases, the results reached here are promising.

The results show a way to identify vegetable oils or to determine the protected denomination of origin, and in addition, the techniques proposed are suitable to detect adulteration at relatively low concentrations.

LITERATURE CITED

(1) Lee, D. S.; Noh, B. S.; Bae, S. Y.; Kim, K. Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics. *Anal. Chim. Acta* **1998**, *358*, 163–175.

(2) Mildner-Szkudlarz, S.; Jelen, H. H. The potential of different techniques for volatile compounds analysis coupled with PCA for the detection of the adulteration of olive oil with hazelnut oil. *Food Chem.* **2008**, *110*, 751–761.
(3) Peña, F.; Cárdenas, S.; Gallego, M.; Valcárcel, M. Direct olive oil authentication: Detection of adulteration of olive oil with hazelnut oil by direct coupling of headspace and mass spectrometry, and multivariate regression techniques. *J. Chromatogr. A* **2005**, *1074*, 215–221.
(4) Dourtoglou, V. G.; Dourtoglou, T.; Antonopoulos, A.; Stefanou, E.; Lalas, S.; Poulos, C. Detection of olive oil adulteration using principal component analysis applied on total and regio FA content. *J. Am. Oil Chem. Soc.* **2003**, *80*, 203–208.
(5) Commission Regulation (EC) No 1989/2003 of 6 November 2003 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-pomace oil and on the relevant methods of analysis.
(6) Marini, F.; Balestrieri, F.; Bucci, R.; Magrì, A. D.; Magrì, A. L.; Marini, D. Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. *Chemom. Intell. Lab. Syst.* **2004**, *73*, 85–93.
(7) Food and Agriculture organization of the United Nations. Codex alimentarius commission. Codex Standard for named vegetable oils. CODEX Stan 210, 1999 (revision and amendments: 2003, 2005).
(8) Vigli, G.; Philippidis, A.; Spyros, A.; Dais, P. Classification of edible oils by employing ^{31}P and ^{1}H NMR spectroscopy in combination with multivariate statistical analysis. A proposal for the detection of seed oil adulteration in virgin olive oils. *J. Agric. Food Chem.* **2003**, *51*, 5715–5722.
(9) Baeten, V.; Hourant, P.; Morales, M. T.; Aparicio, R. Oil and fat classification by FT-Raman spectroscopy. *J. Agric. Food Chem.* **1998**, *46*, 2638–2646.
(10) Brodnjak-Voncina, D.; Kodba, Z. C.; Novic, M. Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids. *Chemom. Intell. Lab. Syst.* **2005**, *75*, 31–43.
(11) Sayago, A.; García-González, D. L.; Morales, M. T.; Aparicio, R. Detection of the presence of refined hazelnut oil in refined olive oil by fluorescence spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 2068–2071.
(12) Fonseca, A. M.; Biscaya, J. L.; Aires-de-Sousa, J.; Lobo, A. M. Geographical classification of crude oils by Kohonen self-organizing maps. *Anal. Chim. Acta* **2006**, *556*, 374–382.
(13) Zhang, J.; Yang. H. Application of self-organizing neural networks to classification of plant communities in Pangquangou Nature Reserve, North China. *Front. Biol. China*, DOI 10.1007/s11515-008-0061-7.
(14) Raju, K. S.; Kumar, D. N.; Duckstein, L. Artificial neural networks and multicriterion analysis for sustainable irrigation planning. *Comput. Oper. Res.* **2006**, *33*, 1138–1153.
(15) Bashyal, S.; Venayagamoorthy, G. K. Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Eng. Appl. Artif. Intell.* **2008**, *21*, 1056–1064.
(16) Wu, C.; Chen, H. L.; Chen, S. C. Counter-propagation neural networks for molecular sequence classification: supervised LVQ and dynamic node allocation. *Appl. Intell.* **1997**, *7*, 27–38.
(17) Masouma, S.; Bouveresse, D. J. R.; Vercauteren, J.; Jalali-Heravi, M.; Rutledge, D. N. Discrimination of wines based on 2D NMR spectra using learning vector quantization neural networks and partial least squares discriminant analysis. *Anal. Chim. Acta* **2006**, *558*, 144–149.
(18) Ciosek, P.; Wróblewski, W. The analysis of sensor array data with various pattern recognition techniques. *Sens. Actuator B-Chem.* **2006**, *114*, 85–93.
(19) Marini, F.; Magrì, A. L.; Bucci, R.; Magrì, A. D. Use of different artificial neural networks to resolve binary blends of monocultivar Italian olive oils. *Anal. Chim. Acta* **2007**, *599*, 232–240.
(20) Kohonen, T. *Self-Organization and associative Memory*, 2nd ed.; Berlin: Springer-Verlag, 1987.
(21) Demuth, H.; Beale, M.; Hagan, M. *MATLAB User's Guide, v 4.0*; Neural Network Toolbox, MathWorks Inc., Mass., USA, 2005.
(22) Lloyd, G. R.; Brereton, R. G.; Faria, R.; Duncan, J. C. Learning vector quantization for multiclass classification: application to characterization of plastics. *J. Chem. Inf. Model.* **2007**, *47*, 1553–1563.
(23) Gan, H. L.; Che Man, Y. B.; Tan, C. P.; NorAini, I.; Nazimah, S. A. H. Characterisation of vegetable oils by surface acoustic wave sensing electronic nose. *Food Chem.* **2005**, *89*, 507–518.
(24) Spangenberg, J. E.; Macko, S. A.; Hunziker, J. Characterization of olive oil by carbon isotope analysis of individual fatty acids: Implications for authentication. *J. Agric. Food Chem.* **1998**, *46*, 4179–4184.
(25) Dais, P.; Spyros, A.; Christophoridou, S.; Hatzakis, E.; Fragaki, G.; Agiomyrgianaki, A.; Salivaras, E.; Siragakis, G.; Daskalaki, D.; Tasioula-Margari, M.; Brenes, M. Comparison of analytical methodologies based on ^{1}H and ^{31}P NMR spectroscopy with conventional methods of analysis for the determination of some olive oil constituents. *J. Agric. Food Chem.* **2007**, *55*, 577–584.
(26) Torrecilla, J. S.; Rojo, E.; García, J.; Rodríguez, F. Principal component analysis/UV spectroscopy for the determination of 1-ethyl-3-methylimidazolium ethylsulfate ionic liquid and toluene concentrations in aqueous solutions. *Ind. Eng. Chem. Res.* **2008**, *47*, 4025–4028.
(27) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694–670.
(28) Gramatica, P.; Giani, E.; Papa, E. Statistical external validation and consensus modeling: A QSPR case study for Koc prediction. *J. Mol. Graph.* **2007**, *25*, 755–766.
(29) Ozen, B. F.; Mauer, L. J. Detection of hazelnut oil adulteration using FT-IR spectroscopy. *J. Agric. Food Chem.* **2002**, *50*, 3898–3901.

Received for review November 10, 2008. Revised manuscript received January 29, 2009. Accepted February 5, 2009. The authors are grateful to the Spanish "Ministerio de Ciencia e Innovación" for financial support for project CTQ2006−04644 and a Ramón y Cajal research contract of José S. Torrecilla.

JF803520U