JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Virgin Olive Oil Quality Classification Combining Neural Network and MOS Sensors

DIEGO L. GARCÍA-GONZÁLEZ* AND RAMÓN APARICIO

Instituto de la Grasa, Padre García Tejero 4, 41012 Sevilla, Spain

A model based on neural networks has been designed to detect lampante virgin olive oils, a category of olive oil that cannot be consumed without a previous refining process according to the current regulation of the European Communities. The response of 7 metal oxide sensors analyzing 114 olive oil samples has been used in the design, training, and internal validation of the neural network with only 4.5% error in validation. The designed mathematical model, the equations of which are fully described, has been validated also with an external set of 13 samples of diverse varieties and geographical origins with 100% correct classification.

KEYWORDS: Electronic nose; metal oxide semiconductor sensor; multilayer perceptron; lampante virgin olive oil

INTRODUCTION

The market price of virgin olive oils (VOOs) is determined by their sensory quality. The current European Union (EU) regulation (*1*) classifies VOO into three categories: extra virgin, virgin, and lampante. The last category also includes the old category, so-called "ordinary" (*2*). From these three categories, it is important to distinguish lampante virgin olive oils because they cannot be consumed without refining, whereas the "ordinary" classification can sometimes indicate a risk of consumers' rejection. The current methods based on sensory and chemical evaluations of VOO quality are lengthy, expensive, and sometimes affected by the subjectivity of assessors' or analysts' errors. Thus, the risk of wrong classifications increases even more when one is working with panel tests from different nationalities (*3*), whereas the relationship between chemical compounds and sensory attributes is the main drawback of chemical analyses (*4*, *5*) when olive oils are classified into their categories. Besides, none of these methods can be applied on-line.

The alternative is the use of sensors for the measurement of the foodstuff's entire aroma (*6*, *7*). Sensors do not need any pretreatment and do not use solvents to detect the presence of volatiles; besides their main advantages are their low cost and the rapid evaluation of the aroma. Metal oxide semiconductor sensors (MOS) have been used in the quality classification of edible oils (*8*, *9*). Thus, sensors would be useful in distinguishing the highest quality VOOs from the lowest ones if they were able to cluster together all low-quality VOOs whichever the off-flavor, or combination of off-flavors, presented in the oil. However, most of the MOS show a nonlinear response to a given chemical compound or odor. Therefore, this property makes the nonlinear mathematical procedures attractive for MOS analyses. Thus, artificial neural network (ANN), a statistical method appropriate for handling highly complex and nonlinear data, has allowed satisfactory results to be obtained when applied to sensor responses (*6*). Among the different types of ANNs, the multilayer perceptron (MLP) is nowadays considered to be one of the most common in regression and classification tasks (*10*). This paper analyzes the possibilities of MOS and MLP after working with the samples of a data set and verification of the results with an external validation set.

MATERIALS AND METHODS

Materials. A Spanish VOO var. Farga spiked with 60 mg/kg acetic acid was the standard used in the repeatability studies. This standard was still frozen and did not alter during the whole time of the experiments. Blends of refined olive oil and VOO var. Hojiblanca at various percentages (5, 10, 25, 50, and 75%) were used to determine the limits of detection (LOD).

One hundred and fourteen samples of VOO (var. Hojiblanca) were used for designing and training the MLP procedure. This set of 114 samples was split into three subsets: 68 samples (training set), 22 samples (test set), and 24 samples (validation set). The samples were supplied by an association of cooperatives (Hojiblanca SCA, Málaga, Spain) that represents 4% of total Spanish olive oil production. Fifty-six of these samples (49%) were qualified as lampante by the assessors of the cited association. On the other hand, another set (external validation) of 13 samples (var. Arbequina, Cornicabra, and Picual) was used to check the mathematical model. These samples were supplied by Aceites del Sur SA (Sevilla, Spain), four samples being classified as lampante by trained assessors. All of the samples were collected in different geographical regions and analyzed for 11 months to check the effect of drift in the sensor baseline.

Equipment. A Fox 4000 with ACU 500 humidifier supplied by AlphaMOS SA (Toulouse, France) was used. This instrument is equipped with 18 metal oxide sensors, inside three chambers, 6 of them being undoped metal oxide sensors and 12 being metal oxide sensors doped with noble catalytic metals in order to shift the selectivity spectrum toward different chemical compounds. The temporary and

* Author to whom correspondence should be addressed (e-mail dluisg@cica.es; telephone +34 954611550; fax +34 954616790).

10.1021/jf021217a CCC: $25.00 © 2003 American Chemical Society
Published on Web 05/09/2003

Table 1. Mean Values of RSD (Percent) and LOD[a]

sensor	% RSD A	% RSD B	% RSD C	LOD
S1	1.86	2.07	9.28	2.23
S2	2.96	1.14	19.91	9.03
S3	2.20	1.53	15.72	22.15
S4	2.44	1.15	10.42	16.10
S5	2.49	1.74	14.85	8.57
S6	1.15	2.33	3.10	1.70
S7	9.14	13.74	14.07	15.71
S8	9.20	16.08	16.73	6.82
S9	6.15	11.33	8.38	5.58
S10	7.75	12.72	12.82	9.85
S11	8.75	13.05	13.41	19.13
S12	11.92	20.98	27.46	11.81
S13	10.89	19.78	18.49	74.82
S14	10.02	10.93	17.48	11.83
S15	3.90	4.80	12.59	0.22
S16	9.93	12.16	10.20	16.06
S17	0.94	1.92	1.96	2.62
S18	0.91	1.88	2.23	2.65

[a] Repeatability study for within-day (A), within-week (B), and 6 months (C).

reversible adsorption of volatile-reducing compounds at the sensor surface changes its electrical resistance in a nonlinear manner (*11*). The response is characteristic of each sensor and depends on the concentration and the profile of the volatile compounds.

The air conditioning unit (ACU 500) consists of a thermostat tank containing distilled water where the carrier gas bubbles continually. When a valve is opened during the injection time, a controlled mixture of dry and humid industrial airs sweeps the headspace of the sampling chamber, the temperature of which is controlled automatically.

Industrial air, from an air compressor, was used as carrier gas after being filtered through two columns. The first column was filled with molecular sieve 8/12 mesh (Supelco, Bellefonte, PA) to remove the moisture, whereas the second column was filled with activated carbon (Supelco) to remove hydrocarbons and other undesirable volatile compounds.

The analytical parameters (sample amount, headspace generation time, sample temperature, flow rate, and injection time) were determined after an optimization process based on the evaluation of three desirability functions by fuzzy algorithms (*8*).

Samples were analyzed in duplicate. Standards for calibration of the sensor array were measured at programmed times to check that the aging of the sensors did not affect the measurement.

Measurements of Repeatability and Limit of Detection. Repeatability studies, either within-day or within-week or between-days, were investigated by consecutively collecting the sensor results of the same sample of VOO (var. Farga spiked with 60 ppm of acetic acid) (*8*). The within-week repeatability was determined by analyzing the sample for 5 consecutive days. Finally, the between-days repeatability study was carried out for 6 months. The relative standard deviation (% RSD) was the parameter used to analyze the repeatability.

The LOD of the sensors were calculated by using a calibration line of refined oil samples spiked with VOO at six different percentages. Three replicates of each level were analyzed (*12*). Once a linear regression analysis was carried out for each sensor, the LOD calculation was based on the standard deviation of the regression line and the slope (*13*).

Data Processing. The response of the sensors presents an exponential-like shape, but not all of this information is useful. After different methods of data preprocessing had been tested, raw data (non-preprocessing data) were selected because they showed the best differential properties (*9*). Windowed time slicing (WTS) (*14*) was used to reduce the information to a reasonable data set. The number of windowing functions was 4, each one applied to a different region of the sensor response (*8*).

A standard was analyzed before and after each series of analyses with the objective of minimizing the effect of sensor aging and environmental conditions. The information was used to standardize WTS data.

The detection of multivariate outliers was carried out by applying principal component analysis (PCA). Mahalanobis distance, evaluated as χ^2, was used to discover outliers among samples and with respect to the solution, whereas outliers among variables (WTS) were detected by the squared multiple correlation (*15*).

A genetic algorithm was applied to select the optimal set of variables (WTS) for the neural network. An ANN was used because of its ability to handle nonlinear data and to compensate for the drift of the sensor array (*6*). MLP, which is perhaps the most popular network architecture, was used to study the differences between lampante and nonlampante VOO. A conjugate gradient descent algorithm (*16*) was used to minimize the prediction error made by the ANN. The training algorithm used the sum-squared error function to train the network and to report the root-mean-square (RMS) error. The weights and threshold were calculated by applying this algorithm to the training set. The best network was selected by means of the samples of the test set. This set is also used to stop the training procedure in case of overlearning. ANN performance was determined by the validation set.

Statistica (*17*) was used to perform the data processing and to implement neural network analyses.

RESULTS AND DISCUSSION

Study of Repeatability and Limit of Detection. Because the main limitations of sensor systems are related to drift and

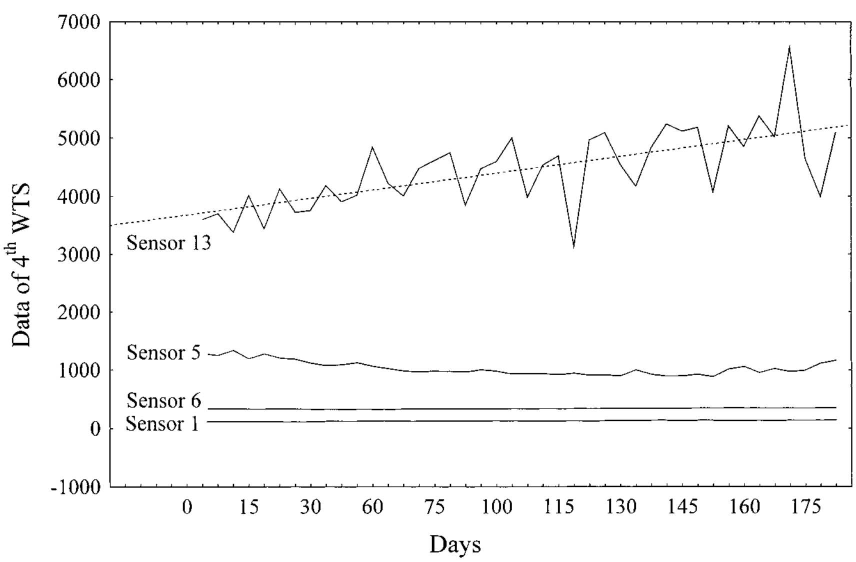

Figure 1. Values of the fourth windowed time slicing (WTS) of various sensors for 6 months.

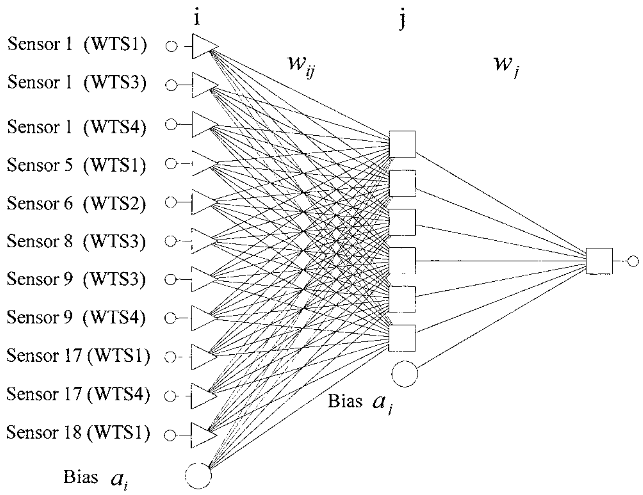

Figure 2. Structure of the multilayer perceptron (MLP) used to distinguish between lampante and nonlampante VOO.

low repeatability (*6*), it is important to understand whether any signal variation is due to sample change or an inherent signal drift. On the other hand, an estimation of the detection limits is required to know the minimum amount of volatile compounds of a VOO sample that can be detected by each sensor.

Table 1 shows the % RSD (mean of the four WTS) of all the sensors for the within-day, within-week, and between-day repeatability studies. The WTS values of the sensors with values >10% were not used in the following studies. It is important to remark that within-day and within-week results displayed in **Table 1** are the worst of all the studies carried out for 6 months. **Figure 1** shows the response of some sensors for 6 months to study their aging. The signals of some sensors (sensors 1, 5, and 6) show no change during this period. The response of sensor 13, however, increased over time, which means that a significant part of the information is caused by a change of the sensor sensitivity. Thus, the use of reference standards seems to be very important when any analysis with sensors showing this behavior is carried out (*18*). We selected a VOO spiked with acetic acid because the best material for monitoring performance and providing data for calibration should be similar to the product that is being tested (*18*) and acetic acid qualifies the undesirable sensory attribute of "winey/vinegary" (*8*).

Table 1 also shows the values of LOD of all the sensors. An LOD of ≤10% was considered to be appropriate for providing satisfactory results. Thus, sensors 1, 5, 6, 8, 9, 17, and 18 showed the best behaviors according to the repeatability and LOD values.

Classification of Virgin Olive Oils. The profile of volatile compounds is responsible for the aroma of low-quality VOOs (lampante) and, hence, for the off-flavors. The hypothesis that the sensor response depends on the amount and composition of volatile compounds has already been demonstrated by the authors, who analyzed some negative attributes (fusty, rancid, and vinegary) by canonical correlation (*8*). On the basis of these promising results, a study with the following sequence was planned: detection of multivariate outliers by means of PCA, design and training of MLP procedure, implementation of a neural network with a discriminant model, and a second validation of the model by an external set of samples.

The raw information was first clustered into four WTS and then standardized to avoid hypothetical sensor aging. The collected information was checked for the detection of outliers.

The study of outliers is extremely necessary as they can greatly affect the magnitudes of the decision equation coefficients. This study was carried out by multivariate procedures, as the problems come mostly from multivariate outliers among variables and cases. Thus, five multivariate outliers among cases (two nonlampante and three lampante VOOs) and three multivariate outliers among variables (sensors 2, 12, and 13) were detected by PCA and removed prior to the following studies.

Once the outliers were removed, a genetic algorithm was applied to the WTS of the selected sensors in order to choose the optimal variables for designing the network. The procedure selected 11 inputs from an initial set of 28 variables. Several ANNs were designed with these inputs, using the set of samples var. Hojiblanca (training set). After several network architectures had been trained and validated, an MLP 11:11-6-1:1 (**Figure 2**) was selected with a classification rate of 95%. This MLP was obtained after a training process with the conjugate gradient descent algorithm (58 epochs) until obtaining the minimum RMS error (**Figure 3**). The RMS errors of the three subsets, in which the data set was divided, were 0.16 (training set), 0.19

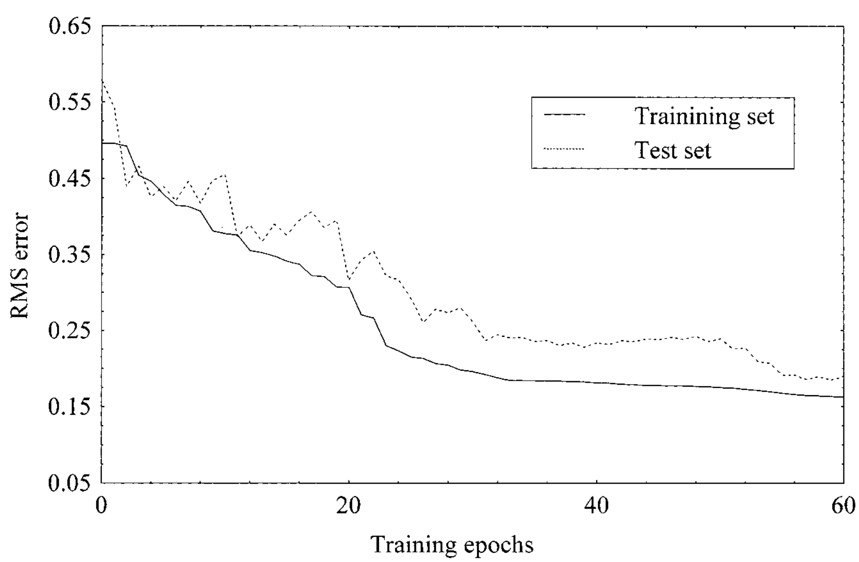

Figure 3. Evolution of root-mean-square (RMS) error during training by the conjugated gradient descent algorithm.

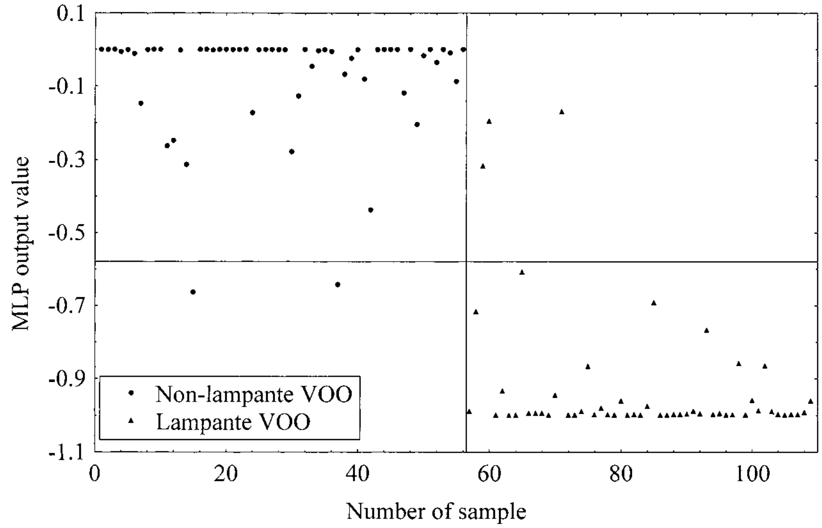

Figure 4. Results of applying the multilayer perceptron (MLP) to the samples of the data set.

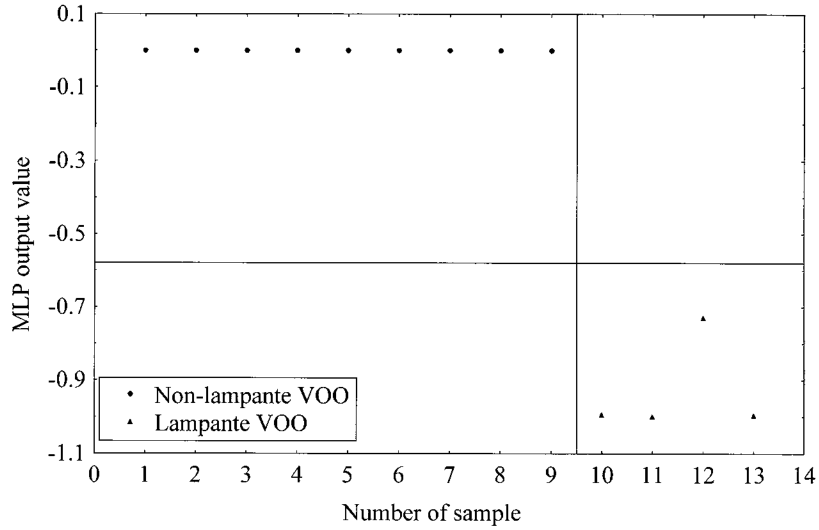

Figure 5. Results of applying the multilayer perceptron (MLP) to the samples of the external validation set.

(test set), and 0.27 (validation set). The designed MLP of three layers is defined by the equation

$$y = f[\sum_j w_j\, f(\sum_i w_{ij}\, x_i + a_i) + a_j] \qquad (1)$$

where y is the output variable, x_i is the input variable, w_{ij} and w_j are the weights for the connections from the input layer to the hidden one and from the hidden layer to the output, respectively, and a_i and a_j are constants that operate as "bias" values in the network. The values for these parameters are shown in **Table 2**. The output values for each node use the sigmoid activation function (f):

$$f(x) = 1/(1 + e^{-x}) \qquad (2)$$

The set of sensors selected for MLP (x_i in eq 1) shows that the most relevant information concerns to the processes of adsorption (WTS1) and desorption (WTS4 and WTS3) of volatiles and the steady state (WTS2). The second kind of information is related with sensor characteristic because sensors 1, 5, and 6 are undoped, whereas sensors 8, 9, 17, and 18 are doped (*6*). Finally, no discrimination was detected in terms of the order of sensors evaluating the samples because four belong to the first chamber (sensors 1, 5, and 6) and the others are placed inside the second (sensors 8 and 9) and third (sensors 17 and 18) chambers.

Table 2. Values of the Parameters in the Neural Network Equation[a]

	hidden layer						output layer	
	w_{i1}	w_{i2}	w_{i3}	w_{i4}	w_{i5}	w_{i6}		output unit
a_i	−4.85	−0.63	1.41	5.48	−0.13	2.03	a_j	0.01
x_1	2.45	2.68	−2.25	−12.12	0.62	−0.75	w_1	8.66
x_2	−0.21	1.19	−0.82	4.36	1.91	0.44	w_2	0.08
x_3	−4.81	1.99	−3.08	16.09	1.41	0.59	w_3	−1.27
x_4	2.05	−0.84	1.17	4.48	0.34	−2.59	w_4	−20.11
x_5	3.91	−1.15	1.09	−3.77	−0.26	−1.86	w_5	−1.52
x_6	1.95	0.32	−0.70	−2.90	−1.84	−1.08	w_6	0.23
x_7	0.84	0.06	−2.01	3.20	−0.04	−1.53		
x_8	1.26	1.45	−2.25	1.06	−0.31	−1.41		
x_9	0.72	1.02	−0.52	−1.46	0.53	−0.89		
x_{10}	−1.42	1.29	−1.57	−1.49	0.03	−0.32		
x_{11}	1.44	1.19	0.46	0.76	−0.29	−1.68		

[a] x_i is the input variable; w_{ij} and w_j are the weights for the connections between layers; and a_i and a_j are the bias values.

After design of the network, the MLP was applied to the whole data set with the idea of testing the model in a larger number of samples. The neural model was able to classify correctly 96.4% of the nonlampante and 94.3% of the lampante

VOOs for a threshold value of −0.58. Furthermore, the RMS error was slightly higher in the lampante VOOs (0.22) than in the nonlampante (0.16). The diversity of possible off-flavors (rancid, winey-vinegary, fusty, muddy sediment, cucumber, etc.) explains why sensors have difficulty in clustering together all of them. This problem does not appear in the case of nonlampante VOOs, which lack off-flavors, and hence the neural network interprets this absence of off-flavors in an olive oil as being a nonlampante VOO.

Figure 4 shows the output values of the MLP against the number of samples. The *Y*-axis indicates the quality level of the samples in relation to the sensory assessment, because this axis shows the values obtained after application of the neural network equation to each sample of the data set. Values of the network equation close to 0 correspond to nonlampante VOOs, whereas values close to −1 indicate that the analyzed samples are lampante VOOs. The threshold centered at −0.58 was established in such a way that the errors of a wrong classification were minimized (*17*). The threshold was then used as a limit between lampante and nonlampante VOOs. The samples classified next to this threshold indicate that their quality would correspond to the former "ordinary" VOOs, a category that is between the nonlampante and lampante categories, according to a previous EU regulation (*2*). In this zone is where the risk of a wrong classification is quite high due to the absence of discontinuity in the sensory evaluation by assessors, and hence it can be considered as a transition zone between the two main groups (lampante versus nonlampante VOOs). When this threshold was applied, only five samples (two nonlampante and three lampante VOOs) were classified incorrectly. Two of these lampante samples erroneously classified as nonlampante VOO had been previously declared to be "ordinary" VOO according to the cited former regulation for the sensory assessment (*2*). The other, evaluated with a median of defects (Md) of 6.4, was evaluated again by assessors of an official panel (Instituto de la Grasa), and it was classified with Md = 5.8, what indicates that the sample would be within the former "ordinary" category (*2*). This fact demonstrated the difficulty in obtaining a full consensus among different panel tests when they evaluate VOOs (*3*). The two nonlampante VOOs that were misclassified were also evaluated by the panel test previously mentioned, and a slightly winey flavor was found in one of them. However, no deficiency was found in the other sample.

After the designing of the ANN using a set of samples of the same variety, the next aim was the validation of the neural network equation with an external validation set to check the generalization ability. The equation was applied to a set of samples of different single varieties (Arbequina, Cornicabra, and Picual) and diverse geographical origins. The aim was to check if the variety and/or the sensory evaluation, in this case carried out by the panel test of the factory Aceites del Sur SA, affected the model based on neural networks. As a result, all of the samples (100%) were correctly classified (**Figure 5**), which indicates the validity of the proposed model.

In conclusion, the model based on neural networks has been able to distinguish lampante virgin olive oils from the other categories with only 4.5% error in the data set and without error in an external validation set. However, the error is even less if we take into account that only one sample could not be explained by a second sensory evaluation carried out with an official panel test. Thus, the designed model avoids the subjective opinions of assessors, classifies a sample in only a few seconds, and has a minimum cost of analysis.

LITERATURE CITED

(1) EC. *Off. J. Eur. Communities* **2002**, May 15, Regulation 796/2002.

(2) EC. *Off. J. Eur. Communities* **1991**, July 11, Regulation 2568/91.

(3) Ranzani, C. Olive oil quality and EEC regulations. *Grasas Aceites* **1994**, *45*, 1−4.

(4) Aparicio, R.; Morales, M. T.; Alonso, M. V. Relationship between volatile compounds and sensory attributes of olive oil by the sensory wheel. *J. Am. Oil Chem. Soc*. **1996**, *73*, 1253−1264.

(5) Morales, M. T.; Tsimidou, M. The role of volatile compounds and polyphenols in olive oil sensory quality. In *Handbook of Olive Oil: Analysis and Properties*; Harwood, J., Aparicio, R., Eds.; Aspen: Gaithersburg, MD, 2000; pp 393−458.

(6) García-González, D. L.; Aparicio, R. Sensors: From biosensors to the electronic nose. *Grasas Aceites* **2002**, *53*, 96−114.

(7) Shurmer, H. V.; Gardner, J. W. Odour discrimination with an electronic nose. *Sens. Actuator B* **1992**, *8*, 1−11.

(8) García-González, D. L.; Aparicio, R. Detection of vinegary defect in virgin olive oils by metal oxide sensors. *J. Agric. Food Chem*. **2002**, *50*, 1809−1814.

(9) García-González, D. L.; Aparicio, R. Detection of defective virgin olive oils by metal-oxide sensors. *Eur. Food Res. Technol*. **2002**, *215*, 118−123.

(10) Gallinari, P.; Cibas, T. Practical complexity control in multilayer perceptrons. *Signal Processing* **1999**, *74*, 29−46.

(11) Mielle, P. "Electronic noses": Towards the objective instrumental characterization of food aroma. *Trends Food Sci. Technol*. **1996**, *7*, 432−438.

(12) Boqué, R.; Maroto, A.; Riu, J.; Rius, X. Validation of analytical methods. *Grasas Aceites* **2002**, *53*, 128−143.

(13) ICH-Q2B Validation of Analytical Procedures: Methodology (International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, Geneva, Switzerland, Nov 1996).

(14) Gutiérrez-Osuna, R.; Nagle, H. T. A method for evaluating data-preprocessing techniques for odor classification with an array of gas sensors. *IEEE Trans. Syst. Man Cyber*. **1999**, *29*, 626−632.

(15) Tabachnick, B. G.; Fidell, L. S. *Using Multivariate Statistics*; Harper and Row Publishers: New York, 1983.

(16) Hagan, M. T.; Demuth, H. B.; Beale, M. Performance optimization. In *Neural Network Design*; Hagan, M. T., Demuth, H. B., Beale, M., Eds.; PWS Publishing: Boston, MA, 1995; pp 9.1−9.40.

(17) *Statistica, User's Guide*, release 5.5; Statsoft: Tulsa, OK, 2000.

(18) Shiers, V.; Adechy, M. Use of multi-sensor array devices to attempt to predict shelf lives of edible oils. *Semin. Food Anal.* **1998**, *3*, 43−52.

Received for review December 16, 2002. Accepted March 25, 2003.

JF021217A