JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Supervised Pattern Recognition Procedures for Discrimination of Whiskeys from Gas Chromatography/Mass Spectrometry Congener Analysis

DOMINGO GONZÁLEZ-ARJONA,† GERMÁN LÓPEZ-PÉREZ,†
VÍCTOR GONZÁLEZ-GALLERO,‡ AND A. GUSTAVO GONZÁLEZ*,‡

Departments of Physical Chemistry and Analytical Chemistry, University of Seville, 41012 Seville, Spain

The volatile congener analysis of 52 commercialized whiskeys (24 samples of single malt Scotch whiskey, 18 samples of bourbon whiskey, and 10 samples of Irish whiskey) was carried out by gas chromatography/mass spectrometry after liquid−liquid extraction with dichloromethane. Pattern recognition procedures were applied for discrimination of different whiskey categories. Multivariate data analysis includes linear discriminant analysis (LDA), k nearest neighbors (KNN), soft independent modeling of class analogy (SIMCA), procrustes discriminant analysis (PDA), and artificial neural networks techniques involving multilayer perceptrons (MLP) and probabilistic neural networks (PNN). Classification rules were validated by considering the number of false positives (FPs) and false negatives (FNs) of each class associated to the prediction set. Artificial neural networks led to the best results because of their intrinsic nonlinear features. Both techniques, MLP and PNN, gave zero FPs and zero FNs for all of the categories. KNN is a nonparametric method that also provides zero FPs and FNs for every class but only when selecting $K = 3$ neighbors. PDA produced good results also (zero FPs and FNs always) but only by selecting nine principal components for class modeling. LDA shows a lesser classification performance, because of the building of linear frontiers between classes that does not apply in many real situations. LDA led to one FP for bourbons and one FN for scotches. The worse results were obtained with SIMCA, which gave a higher number of FPs (five for both scotches and bourbons) and FNs (six for scotchs and two for bourbons). The possible cause of these findings is the strong influence of class inhomogeneities on the SIMCA performance. It is remarkable that in any case, all of the methodologies lead to zero FPs and FNs for the Irish whiskeys.

KEYWORDS: Discrimination; whiskey; pattern recognition

INTRODUCTION

Discrimination issues are of utmost importance for a wide variety of products including foods and beverages according to the requirements of countries' laws. Within the realm of alcoholic beverages, the consumption of spirits in the world as appetizers, digestives, and now, especially, as long drinks is increasing. From the aqua vitae of Arnaldo de Vilanova and the aqua ardens of Raimon Llull (*1*) until the most recent products, distilled alcoholic beverages have been present throughout time in all towns and societies of the world. Accordingly, discrimination and authentication procedures have been applied to spirituous beverages by taking into account their elaborate features. Besides, considering that a majority of the compounds selected as descriptors for discrimination purposes are volatile in nature, the analytical technique more prone to volatile congener determination is gas chromatography (GC). Thus, GC has been suitably used for discrimination purposes of several kinds of spirits and liquors such as gins (*2*), brandies and cognacs (*3*, *4*), rums (*5*), tequilas (*6*), and whiskeys (*7*−*9*). Our research, now in progress, is focused on whiskey discrimination, especially to distinguish among single malt Scotch whiskey, bourbon whiskey, and Irish whiskey. Whiskey was legally defined in the European Community Council Regulation No. 1576/89 (*10*). Scotch whiskey is a distilled spirit made in Scotland. (Generally, although not always, the Scottish, Japanese, and Canadian spirits are spelled "whisky"; the Irish and American ones are spelled "whiskey".) The name whiskey is a transformation of the word uisquebaugh, itself a transformation of the Scottish Gaelic uisge beatha, spelled uisce beatha in Irish Gaelic, literally meaning the "water of life". In North America, the abbreviated term scotch is usually used for Scotch whiskey. In England, Scotland, and Wales, the term whiskey almost always refers to Scotch whiskey, and the term scotch is rarely used by itself (*11*). To legally be called Scotch whiskey,

* To whom correspondence should be addressed. Tel: +34-954557173. Fax: +34-954557168. E-mail: agonzale@us.es.
† Department of Physical Chemistry.
‡ Department of Analytical Chemistry.

10.1021/jf0517389 CCC: $33.50
Published on Web 02/10/2006

the spirit must conform to the standards of the Scotch Whiskey Act 1988 (*12*). There are mainly two distinct types of Scotch whiskey, malt whiskey and grain whiskey, but most of the well-known brands of Scotch whiskey are blended from many individual malt and grain whiskeys.

Bourbon is an American form of whiskey, made from at least 51% but not more than 79% corn or maize (typically about 70%) with the remainder being wheat, rye, and malted barley. It is distilled to no more than 160 (U.S.) proof and aged in new charred white oak barrels for at least 2 years (usually much longer). In this way, it is similar to Scotch whiskey, which is also aged in charred barrels. Most of the time, it is then adjusted to 80−100 proof and bottled, although some are bottled at "cask strength" (*12*). Legal definitions are in agreement with the Code of Federal Regulations (*13*).

Irish whiskey is made either from malted barley or from mixtures of malted and unmalted barley and other cereals but with not less than 25% of malted barley. The malt is dried in closed kilns, which avoids the smoky taste and ensures a smooth and natural flavor. In Ireland, whiskey is obtained after three separate distillations. First, a pot still distillate called "low wines" is obtained. This full-flavored product is then distilled in another pot still. The resulting product, called "feints", requires one further distillation that is carried out in a patent Coffey still. According to this triple distillation, a final spirit of light and delicate character is obtained. The product is stored at not more than 63% alcohol (v/v) in oak casks, some of which have been used previously for sherry, for a period of 5 years or more (*14*−*16*). The use of a mixture of partially malted barley, the absence of peat fire smoking, and the application of a triple distillation procedure confer to Irish whiskey peculiar features. Irish whiskey distilleries that comply with this typical elaboration joined together in 1966 to constitute the Midleton Centre in County Cork. Aside from the Midleton Centre, Old Bushmill distilleries, located in North Ireland, produces its own whiskey by following a different Scottish recipe (*14*) using pure malted barley, discontinuous distillation using three pot stills, and maturation in sherry and American oak casks and in port wine pipes. The result is a kind of single malt Scotch whiskey.

As stated above, the aim of the present paper is to find suitable discrimination rules to differentiate Irish whiskey, single malt Scotch whiskey (scotch), and bourbon whiskey. For these purposes, GC procedures coupled with mass spectrometry detection (GC/MS) are applied because of their powerful performance for discriminating distilled liquors (*17*). The use of fusel oils (1-propanol, 2-methyl-1-propanol, 2-methyl-1-butanol, and 3-methyl-1-butanol) as chemical descriptors seems to be of interest for discrimination (*8, 18*). However, for authentication purposes, it is more advisable to select a higher number of chemical descriptors, such as components (or congeners) formed in the elaboration process and maturation in casks. Whiskey is a very complex product, with several hundreds of congeners including alcohols, aldehydes, acids, esters, phenols, and carbonyl-, nitrogen-, and sulfur-containing products. Aylott et al. (*7*) have pointed out that higher alcohol congener analysis provides a valuable method for checking whiskey brand authenticity, but further confirmatory data are obtained by reference to cask extractive and volatile phenolic congeners. Accordingly, we have performed GC/MS congener analysis on a number of samples of the three different classes of whiskey mentioned above by direct injection of dichloromethane extracts. The chemical descriptors are 12 congeners suitably selected as will be described later. The discrimination of the selected whiskey categories is achieved by applying supervised learning pattern recognition procedures to the results (congener analysis). Since early works until the present time, linear discriminant analysis (LDA) (*19*) and canonical variate analysis (CVA) (*20*) have been the tools of the trade for classification purposes. CVA attempts to find linear combinations of variables from each set that exhibit maximum correlation. These may be referred to as canonical variates, and data can be displayed as a scatterplot of one against the other. The problem of maximizing the correlation can be formulated as an eigenanalysis problem with the largest eigenvalue providing the maximized correlation and the eigenvectors giving the canonical variates. Loadings of original variables in the canonical variates and cumulative proportions of eigenvlaues are interpreted, partly by analogy with principal component analysis (PCA) (*21*). Note that if one set of variables are dummy variables giving group indicators, then CVA is mathematically identical to LDA. Only Aylott et al. (*7*) apply it for whiskey authentication. However, these methods cannot give good results when (i) class borders are of a nonlinear nature, due to the intrinsically linear features of discrimination surfaces in LDA/CVA, and (ii) the descriptors are non-Gaussian distributed. Accordingly, to find the most suitable procedures for discrimination, besides LDA, we call on other different chemometric techniques. Thus, six supervised learning pattern recognition procedures were applied as follows: LDA, K-nearest neighbors (KNN) (*22*), procrustes discriminant analysis (PDA) (*23*−*25*), soft independent modeling of class analogy (SIMCA) (*26*), and methods based on artificial neural networks. These selected methods are very different in nature. The majority are hard-modeling methods against SIMCA, a class-modeled technique (*27*). KNN and artificial neural networks are nonparametric methods, the best choice when the descriptors are not normally distributed. Moreover, parametric methods such as LDA, SIMCA, and PDA work well when the classes present inner similarities and are well-separated from each other (minimization of the ratio of within classes sum of squares and between classes sum of squares) (*28*). Thus, when dealing with data exhibiting nonlinear class structure, these methods are not a good choice; instead, artificial neural networks are the best selection. Multilayer perceptrons trained by back-propagation (MLP) are the best known and most commonly applied artificial neural networks for classification purposes (*29*−*32*). Nevertheless, another kind of artificial neural network is the called probabilistic neural network (PNN) (*33*, *34*), scarcely employed for discrimination purposes and not yet applied in the field of authentication of alcoholic beverages. This procedure conjugates the neural network features and the bayesian estimation of posterior probabilities for evaluating class membership. An outline of each classification procedure, emphasizing PNN, will be given in the Statistical Analysis section.

A preliminary chemometric treatment consists of implementing a PCA on the data matrix for both dimensionality reduction and to select the most promising chemical descriptors as the chemical variables most contributing to the two first principal components (PCs), in our case 12 congeners.

MATERIALS AND METHODS

Standards and Samples. Dichloromethane (99.9%) (Romil, Barcelona, Spain) and acenaphten (99.0%) (Fluka, Buchs, Switzerland) were of analytical quality. A solution of 0.005% acenaphten in dichloromethane was used as an internal standard. Fifty-two commercially available whiskey samples (bottles) were obtained from retail liquor stores. They consisted of 10 Irish whiskeys (class I), 18 bourbon whiskeys (class B), and 24 single malt Scotch whiskeys (class S). An identification code was assigned to each sample: Scotch whiskeys were

labeled from 1S to 24S; bourbons were labeled from 25B to 42B, and Irish whiskeys were labeled from 43I to 52I. The bottles, once opened, were stored in a cupboard at room temperature. Subsampling and analyses were carried out within a 6 month interval.

Chromatographic Equipment and Conditions. A Fisons GC 8000 gas chromatograph coupled to a Fisons Trio 1000 mass spectrometer (Fisons Instrument, Valencia, CA) and fitted with a J&W fused silica capillary column (J&W Scientific, Folsom, CA) of 30 m × 0.32 mm coated with a 1.8 μm film of DB-624 (86% dimethylsilicon and 14% phenylsilicon) stationary phase was used. Chromatographic conditions were adapted from Headley and Hardy (*35*). The oven was operated in programmed temperature mode: initial temperature, 45 °C for 5 min; program rate, 8 °C/min; final temperature, 240 °C; and acquisition time, 30 min. Helium was used as the carrier gas at a 1.5 mL/min flow rate through the column with a 1:70 split ratio, an injector temperature of 240 °C, and an injection volume of 1 μL.

The mass spectrometer (quadrupole) was operated in EI mode at 70 eV. The GC/MS interface was held at 250 °C, and the ionization source was held at 200 °C. A scan mode in the mass/charge range of 20−250 Da with a scan time of 0.90 and an interscan time of 0.01 s was used. Data acquisition began after a solvent delay of 6.6 min.

Sample Treatment. The congeners to be quantitated were preconcentrated by liquid−liquid extraction: Three milliliters of whiskey was poured in a vial and was treated with 1 mL of dichloromethane containing acenaphten (0.005% as internal standard). The vial was shaken for 30 s, and then, 1 μL of dichlorometane phase was injected in the chromatograph.

Data Acquisition. Because the aim of this work was the discrimination of whiskeys by using the volatile congeners as descriptors rather than the analysis of such congeners, instead of evaluating the concentration of each compound in the whiskey, its relative amount was considered. The relative amount of each congener was taken as the height ratio of the analyte peak and the internal standard (acenaphten) peak. Accordingly, no calibration standards were used.

STATISTICAL ANALYSIS

Multivariate analysis was performed with a number of packages. LDA and artificial neural networks algorithms were carried out by using the STATISTICA package (*36*) fitted with the neural network module. SIMCA P9 (*37*) was utilized for SIMCA classification. PDA was performed by using the Holmes program (*23*, *24*). KNN classification was achieved with a homemade program written in QuickBasic.

To obtain suitable classification rules for assigning categories to whiskey samples, supervised learning pattern recognition methods were applied. For validation purposes, the whole data set is then split randomly into two sets, the training and the evaluation set, each containing about 50% samples of every class. Once the classification rule is developed, some workers consider as validation parameters the recalling efficiency (rate of training samples correctly classified by the rule) and, especially, the prediction ability (rate of evaluation samples correctly classified by the rule). However, these parameters could be misleading because they do not consider the number of false positives (FPs) and false negatives (FNs) for each class. These two concepts provide a deep knowledge of the classes' space. Accordingly, it seems to be more advisable to use the terms sensitivity (SENS) and specificity (SPEC) (*23*, *24*) for validating the decision rule. The SENS of a class corresponds to the rate of evaluation objects belonging to the class that are correctly classified, and the SPEC of a class corresponds to the rate of evaluation objects not belonging to the class that are correctly considered as belonging to the other classes. This may be explained in terms of the first and second kind of risks associated with prediction. The first kind of errors (α) corresponds to the probability of erroneously rejecting a member of the class as a nonmember (rate of FN). The second kind of errors (β) corresponds to the probability of erroneously classifying a nonmember of the class as a member (rate of FP). Accordingly, for a given class A and setting n_A as the number of members of class A, $\bar{n}_A$ as the number of nonmembers of class A, $\langle n_A \rangle$ as the number of members of class A correctly classified as belonging to class A, and $\langle \bar{n}_A \rangle$ as the number of nonmembers of class A classified as not belonging to class A, we have

$$\text{SENS} = \frac{\langle n_A \rangle}{n_A} = 1 - \alpha = 1 - \frac{\text{FN}}{n_A}$$

$$\text{SPEC} = \frac{\langle \bar{n}_A \rangle}{\bar{n}_A} = 1 - \beta = 1 - \frac{\text{FP}}{\bar{n}_A} \quad (1)$$

Values close to unity for both concepts imply both high SENS and SPEC. The supervised learning pattern recognition techniques utilized in this work for classification purposes are presented in the following.

KNN. KNN is a nonparametric method that classifies a test sample according to the class of the majority of its K-nearest neighbors in the training set, by using the Euclidean distance as a similarity measurement (*22*).

LDA. LDA is a linear and parametric method with hard-modeling features (*19*). Discriminant functions are obtained as a linear combination of descriptors that maximize the *F*-ratio of between classes sum of squares and within classes sum of squares. If we have p descriptors and g classes, the number of uncorrelated discriminant functions is either p or $g - 1$, whichever is smaller. In this discriminant space, the classes are separated by decision hyperplanes. Evaluation samples are classified according to the proximity to these hyperplanes, according to the minimum Mahalanobis distance rule, or from the estimation of an a posteriori probability of class membership using the Bayes estimation.

SIMCA. SIMCA is a parametric method with soft-modeling characteristics (*26*). The basic idea is the construction of a PC model for each class separately in the training set. The perpendicular distance of any evaluation sample (considered as a vector) to the hyperplane defined by the eigenvectors of the PC-modeled class is used for classification purposes.

PDA. This method is a parametric and linear one (*23*−*25*), where a PCA of the training set is carried out and the scores are transformed via a procrustes transformation into the true target matrix of class membership (constructed with ones and zeros: scotches are 100, bourbons are 010, and Irish whiskeys are 001). Procrustes transformation is a nonorthogonal oblique transformation involving rotation, translation, and stretch of the score matrix of each class. From the training set and the class membership matrix, the procrustes matrix is obtained and then it is used for predicting the class membership of the evaluation set.

MLP. Artificial neural networks such as feed forward MLP trained by back-propagation of errors are very efficient tools for classification (*29*−*32*). The MLP consists of formal neurons and connections (weights) between them. The neurons are commonly arranged in three layers: an input layer, one hidden layer (sometimes plus a bias neuron), and an output layer according to an architecture as depicted in **Figure 1**. The connections are unidirectional from the input to the output. Adjacent layers are fully connected, and no connections between neurons within the same layer exist. A formal neuron sums up incoming signals, multiplied by the connection weights, subtracts a threshold value (called bias), and calculates output signals by using a transfer function. Input neurons simply distribute the

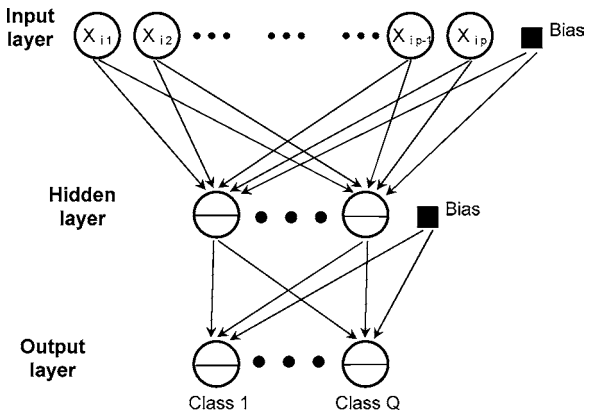

Figure 1. General architecture of a three layer MLP with bias.

elements of the data matrix row to the hidden layer neurons without any further computation. Hidden layer neurons commonly have a sigmoidal transfer function:

$$f(\text{input}) = \frac{1}{1 + \exp(-\text{input})} \tag{2}$$

This limits the neuron's output signal to values between 0 and 1. In addition, output neurons usually have also a sigmoidal transfer function. Training samples are taken at random. After each input, all of the weights are reevaluated according to the called "δ rule". In the back-propagation scheme, each time all of the training samples pass through the network, it is called an iteration cycle or epoch. An important remark for validation purposes concerning the neural network for avoiding overtraining is the use of an additional monitoring set (aside from the training and evaluation sets) to stop the learning process at the suitable number of epochs to avoid the learning of idiosyncrasies of the training samples that may lead to overfitting (*38*).

PNN. PNN is another important classification procedure based on feed forward artificial neural networks without back-propagation that implements a Bayesian decision strategy (*33*, *34*). Let Q equal the number of classes in the training set, namely, C_1, C_2, ..., C_Q; and let n_1, n_2, ..., n_Q equal the corresponding number of pattern vectors that belongs to the aforementioned classes. The prior probability for that a given pattern vector $\mathbf{x}$ belongs to class C_K can be estimated from Laplace's rule:

$$P(C_K) = \frac{n_K}{\sum_{j=1}^{Q} n_j} \tag{3}$$

However, if we know the components x_1, x_2, ..., x_p, this information can be added to the network and the posterior probability can be calculated by using Bayes' theorem (*39*):

$$P(C_k/\mathbf{x}) = \frac{p(\mathbf{x}/C_K)\, P(C_K)}{p(\mathbf{x})} \tag{4}$$

P refers to the probability and p refers to the probability density function (PDF). Thus, $p(\mathbf{x}/C_K)$is the conditional PDF of the pattern vector once it belongs to class C_K. $p(\mathbf{x})$ is the PDF of pattern vectors and plays the role of a scale normalization factor. The posterior probability $P(C_K/\mathbf{x})$ is the probability that the pattern belongs to class C_K once their components are known. This probability can be evaluated by choosing the class

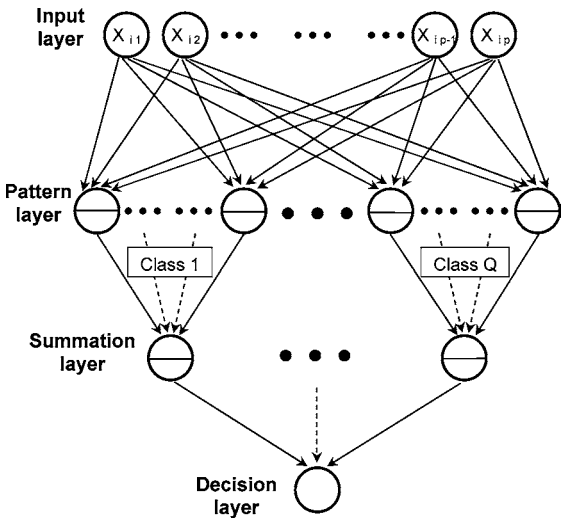

Figure 2. Architecture of a typical PNN.

having the high posterior value:

$$P(C_K/\mathbf{x}) = \max_{j\in\{1,2,\ldots Q\}} P(C_j/\mathbf{x}) \tag{5}$$

Taking into account that $p(\mathbf{x})$ is a normalization factor, we can write

$$p(\mathbf{x}/C_K)\, P(C_K) = \max_{j\in\{1,2,\ldots Q\}} p(\mathbf{x}/C_j)\, P(C_j) \tag{6}$$

The estimation of conditional PDF for every class can be easily done by applying the modified Parzen's estimator (*40*) that is the activation function of PNN and acts as a multivariate Gaussian operator:

$$p(\mathbf{x}/C_K) = \frac{1}{(2\pi)^{p/2}\, \sigma_K^p\, n_K} \sum_{i=1}^{n_K} \exp\left[-\sum_{j=1}^{p} \left(\frac{x_j - x_{ij}^{(k)}}{\sigma_K} \right)^2 \right] \tag{7}$$

x_j and x_{ij} are the components of any pattern vector $\mathbf{x}$ and of the pattern vector $\mathbf{x}_i^{(k)}$ belonging to class K. σ_K is the called smoothing factor (*41*), which is optimized during training.

PNNs are arranged into four layers: the input layer, pattern layer, summation layer, and decision layer. **Figure 2** shows the corresponding architecture. The input layer is used to store the new samples of the validation set. Pattern vectors of the training set are used to optimize the smoothing factor in the training step (*42*). The pattern layer contains as many neurons as pattern vectors of the training set grouped by classes. PNN training is accomplished by simply copying each pattern in the training set to the neurons of the pattern layer. The summation layer consists of one neuron for each class and sums the outputs from all pattern neurons. This gives a measure of the posterior probability density function for each class when an input vector is processed. The decision layer consists of one neuron that searches for the maximum posterior PDF and assigns to the input vector the class with the highest probability.

RESULTS AND DISCUSSION

Congener quantitation in whiskey samples has been carried out, as indicated above, by GC/MS. Chromatographic peaks have been identified by using the NIST library of mass spectra (*7*) with the Lab-Base software (*43*). Typical chromatograms obtained for Scotch, bourbon, and Irish whiskeys are depicted in **Figures 3**−**5**, respectively.

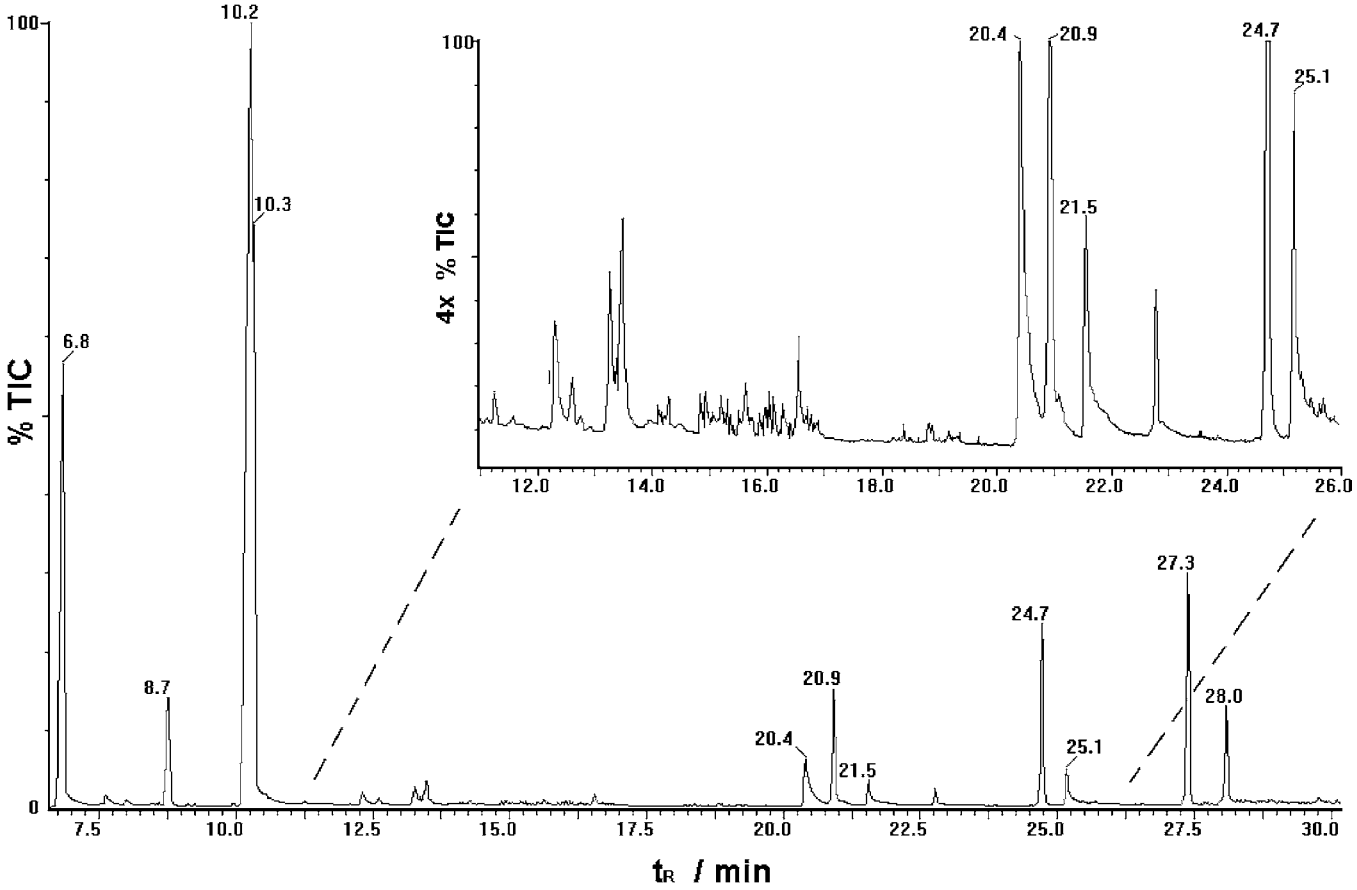

Figure 3. Chromatogram of a Scotch sample zooming at the region of analytical interest. The internal standard peak has a retention time of 27.3 min.

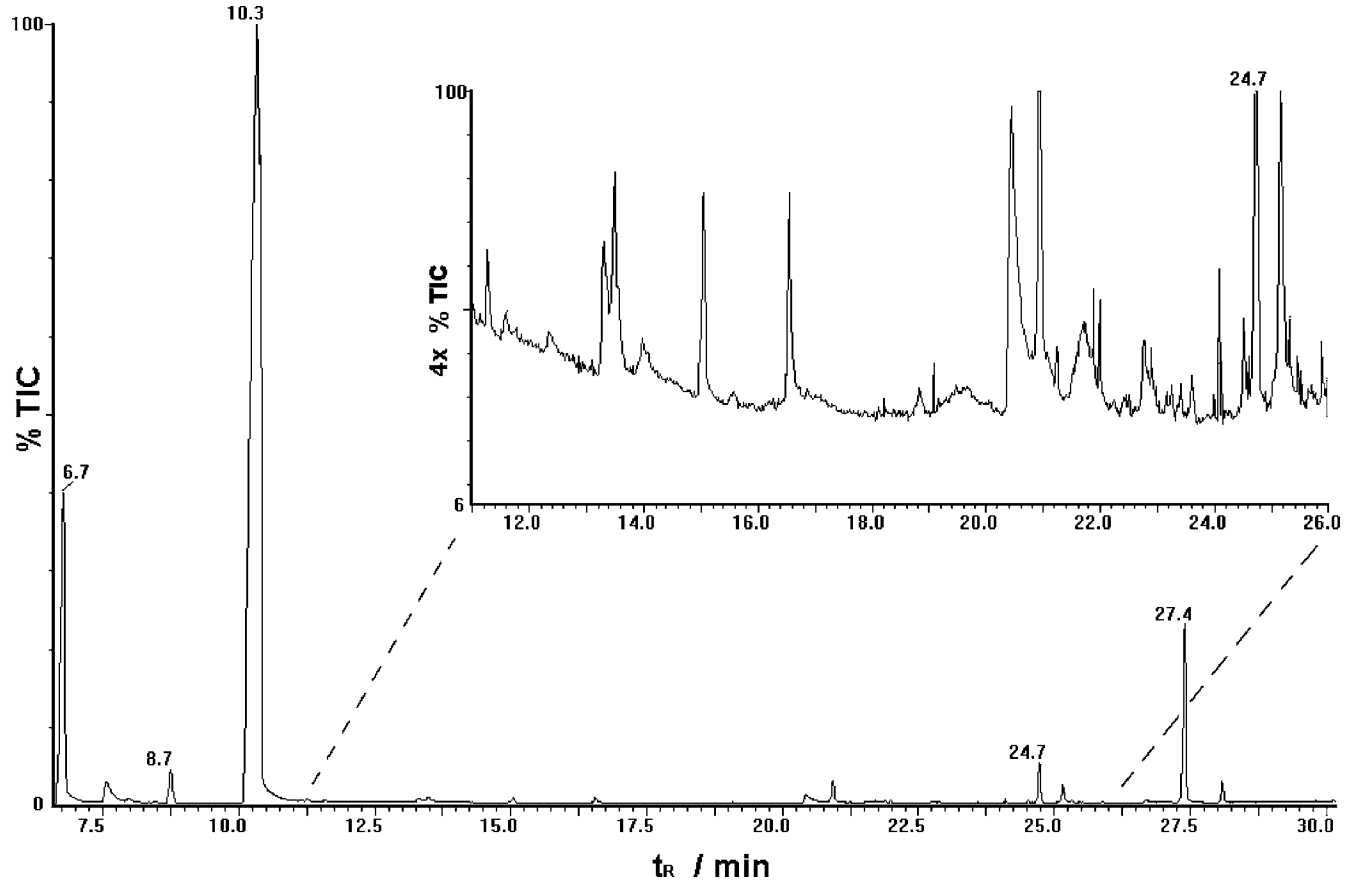

Figure 4. Chromatogram of a Bourbon sample zooming at the region of analytical interest. The internal standard peak has a retention time of 27.4 min.

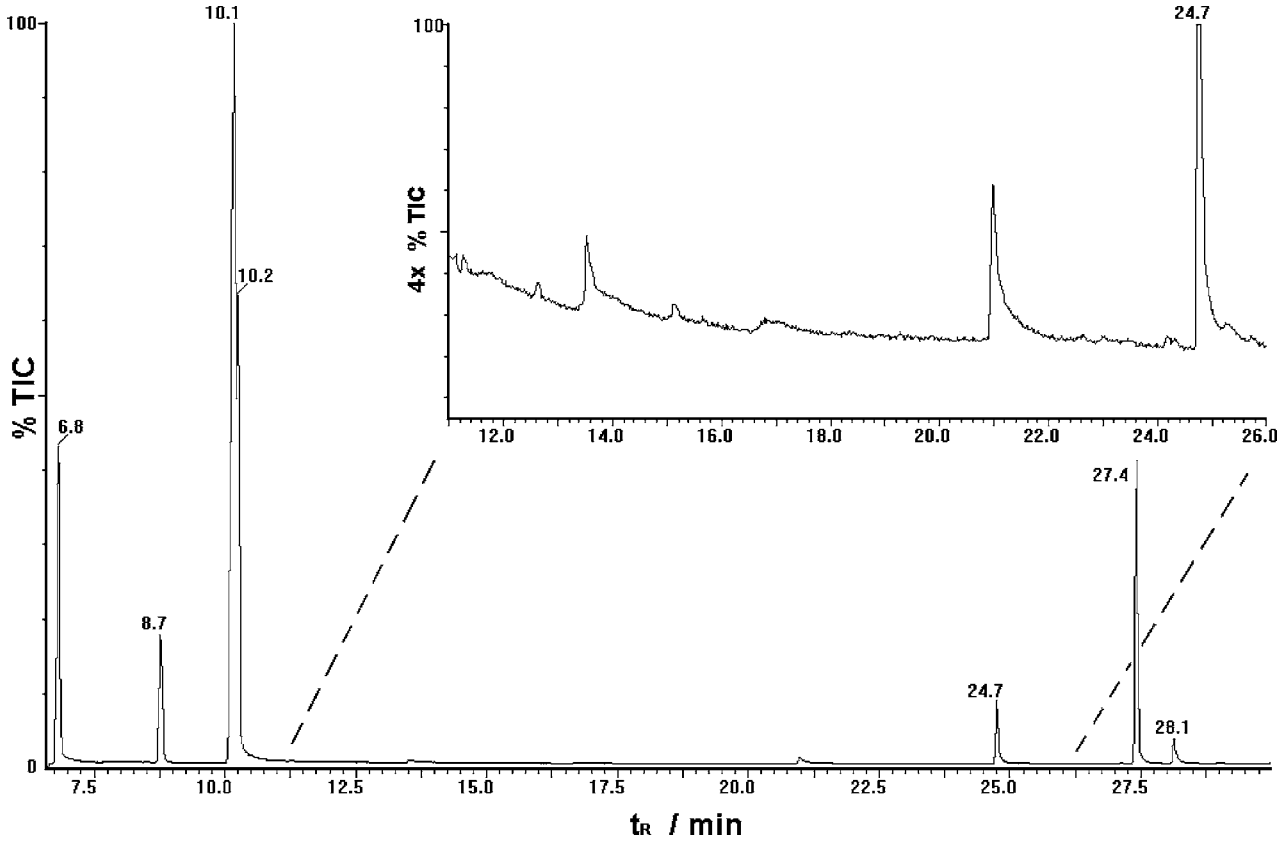

Figure 5. Chromatogram of an Irish sample zooming at the region of analytical interest. The internal standard peak has a retention time of 27.4 min.

Table 1. Classification Performance of Different Pattern Recognition Tecniques According to the SENS and SPEC of the Evaluation Set

	discrimination technique											
	3NN		LDA		SIMCA		PDA		MLP		PNN	
category	sens.	spec.	sens.	spec.	sens.	spec.	sens.	spec.	sens.	spec.	sens.	spec.
Scotch	1.00	1.00	0.92	1.00	0.67	0.70	1.00	1.00	1.00	1.00	1.00	1.00
Bourbon	1.00	1.00	1.00	0.94	0.82	0.77	1.00	1.00	1.00	1.00	1.00	1.00
Irish	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Accordingly, 28 congeners were identified in the different processed whiskey samples. Not all of these congeners can be detected in every sample. As can be observed in chromatograms, different chemical profiles are found for the three whiskey classes, and some congeners are specific compounds for a given class of whiskey, whereas other congeners can be detected in any kind of whiskey. Those congeners with high discriminating power are considered as descriptors or features of the system. To ascertain whether any congener by itself could distinguish among the three classes of whiskeys, a feature selection was made by performing a PCA of cases and retaining the congeners with higher contributions to the first PCs. Thus, the entire data set was subjected to eigenanalysis. The data matrix has 52 rows (whiskey samples) and 28 columns (the relative amounts of the selected chemical descriptors). PCA leads to six significant PCs, validated according to the Kaiser's criterion (*44*), which accounted for 78.6% variance. The features most contributing to the two first PCs were selected as suitable descriptors, namely, 1,1-diethoxybutane, ethyl exanoate, ethyl octanoate, heptanoic acid, ethyl decanoate, decanoic acid, ethyl dodecanoate, dodecanoic acid, hexadecanol, propanoic acid, ethyl formate, and 4-hydroxy-*e*-methoxy-benzaldehyde. Accordingly, the final data matrix has 52 rows and 12 columns. The results obtained for each classification procedure are shown in the following.

KNN. The optimum number leading to a minimum of misclassification is $K = 3$, with FN = 0 and FP = 0 for all classes.

LDA. Figure 6 shows the discriminant scatterplot corresponding to the whiskey samples. The scores were estimated using the selected variables on the training set (closed symbols) and used in the projection to the evaluaation set (open symbols). As can be observed, some class overlapping appears that reflects the nonsuitability of linear frontiers between categories. In our case and using the Bayesian estimation, the results were as follows: scotches, FP = 0 and FN = 1; bourbons, FP = 1 and FN = 0; and Irish whiskeys, FP = 0 and FN = 0.

SIMCA. By applying the SIMCA P9 software, the following results were obtained for each class: scotches (model with six PCs), FP = 5 and FN = 6; bourbons (model with one PC), FP = 5 and FN = 2; and Irish whiskey (model with two PCs), FP = FN = 0. The poor results obtained with SIMCA are sometimes due to the presence of class inhomogeneities (*45*, *46*). Moreover, numerous samples are required for each class to be able to construct meaningful and representative PC models (*47*).

PDA. Selecting nine PCs as the key number of factors in the PCA of the training set, the results for all classes were FP = FN = 0.

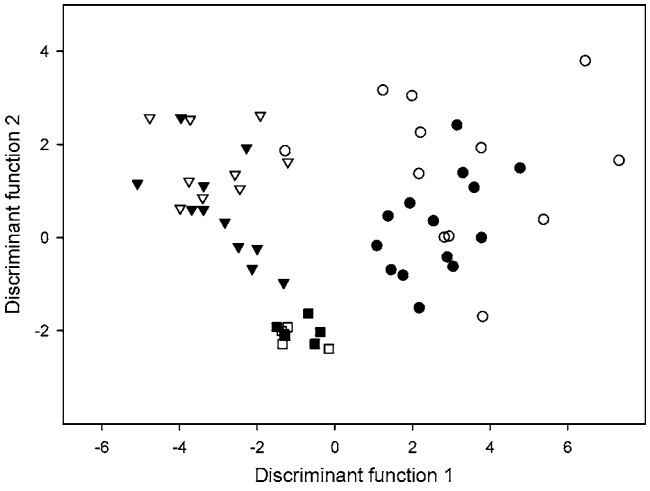

Figure 6. Discriminant scatter plot of whisky samples: Scotch (●), Bourbon (▼), and Irish (■). Scores were estimated using the training set (closed symbols) and projected over the evaluation set (open symbols).

MLP. Initial weights are taken randomly within −0.1 and 0.1. The architecture was 12 input neurons (the chemical descriptors), six hidden neurons (heuristically estimated to minimize overfitting), and three output neurons (the three classes). The learning rate and the momentum were fixed to 0.2 and 0.5, respectively. Target outputs were normalized to 0−1 and written in binary form: 100 (scotches), 010 (bourbons), and 001 (Irish whiskeys). Both training and monitoring errors decreased monotonically with the increasing number of iterations up to 1000 epochs without overfitting. The results for all classes were FP = FN = 0.

PNN. In our case, we have 12 input neurons (one for any chemical descriptor), 30 pattern neurons (14 for scotches, 11 for bourbons, and five for Irish whiskeys), three summation neurons (the three classes), and one decision neuron (the class winner). By adjusting the smooth factor to $\sigma = 0.047$, we obtain FP = FN = 0 for all classes.

The results of the different supervised pattern recognition methods are presented in **Table 1**. As can be observed, KNN, PDA, MLP, and PNN lead to the best results (SENS = 1 and SPEC = 1), followed by LDA and last SIMCA, but for the Irish class, FN = FP = 0 always.

It is remarkable the excellent results obtained when applying artificial neural networks either MLP or PNN. Because although KNN and PDA also led to FP = FN = 0 for every class, this is only true if we choose $K = 3$ neighbors in KNN and if we select nine PCs as latent dimensionality in PDA. Otherwise, the occurrence of FP and FN begins. This fact may be due to a nonlinear disposition of the class frontiers that only can be suitably managed by neural networks.

LITERATURE CITED

(1) Muñoz Ramos, J. *Enciclopedia de los Alcoholes*, 2nd ed.; Planeta, Ed.; Barcelona, Spain, 2002.
(2) Martín Álvarez, P. J.; Herranz, A. Application of multivariate statistical methods to the differentiation of gin brands. *J. Sci. Food Agric.* **1991**, *57*, 263−272.
(3) Schreier, P.; Reiner, L. Characterization and differentiation of grape brandies by multiple discriminant analysis. *J. Sci. Food Agric.* **1979**, *30*, 319−327.
(4) Mazerolles, G.; Vidal, J. P.; Cantagrel, R.; Gaschet, J.; Faure, A.; Lambert, F. *Symposium Scientifique International du Cognac*; Tec. & Doc-Lavoisier Publisher: Paris, France, 1993; pp 410−416.
(5) Herranz, A.; De la Serna, P.; Barro, C.; Martin-Alvarez, P. J. Multivariate statistical methods applied to the determination of rum brands. *J. Sci. Food Agric.* **1990**, *51*, 555−560.
(6) Aguilera-Rojo, D. K.; Martin del Campo, S. T.; Cosio Ramirez, R.; Escalona Buendia, H.; Estarron Espinosa, M. *Proceedings of the 10th Weurman Flavor Research Symposium*; Tec & Doc-Lavoisier Publisher: Beaune, France, 2003; pp 761−764.
(7) Aylott, R. I.; Clyne, A. H.; Fox, A. P.; Walker, D. A. Analytical strategies to confirm Scotch whisky authenticity. *Analyst* **1994**, *119*, 1741−1746.
(8) Wilson, L. A.; Ding, J. H.; Woods, A. E. Gas Chromatographic determination and Pattern recognition analysis of methanol and fusel oil concentrations in whiskeys. *J. Assoc. Off. Anal. Chem.* **1991**, *74*, 248−256.
(9) Martin, G. E.; Burggraff, J. M.; Dyer, R. H.; Buscemi, P. C. Gas−liquid chromatographic determination of congeners in alcoholic products with confirmation by gas chromatography/mass spectrometry. *J. Assoc. Off. Anal. Chem.* **1981**, *64*, 186−190.
(10) ECC Council Regulation No. 1576/89 of 29 May 1989. *Off. J. Eur. Commun.* **1989**, *L160*, 1.
(11) http://en.wikipedia.org/wiki/Scotch_Whisky, last visited Nov. 2005.
(12) http://en.wikipedia.org/wiki/Bourbon_whiskey, last visited Nov. 2005.
(13) *Code of Federal Regulations*; Washington DC, Title 27, Ch. 1, Subpart C, Section 5.22, pp 48−53.
(14) Casamayor, P.; Colombani, M. J. *El Libro del Amante del Whisky*; El Cuerno de la Abundancia: Barcelona, Spain, 1985.
(15) Delos, G. *Les Whiskies du Monde*; Studio: Paris, 1996.
(16) http://www.whiskeytours.ie/howmade/made-3.html, last visited Nov. 2005.
(17) Yoshihiro, H.; Ikeda, H.; Nagata, T.; Kudo, K. Discrimination of distilled liquors by GC/MS. *Jpn. J. Forensic Toxicol.* **1995**, *13*, 49−56.
(18) González-Arjona, D.; González-Gallero, V.; Pablos, F.; González, A. G. Authentication and differentiation of Irish whiskeys by higher-alcohol congener analysis. *Anal. Chim. Acta* **1999**, *381*, 257−264.
(19) Coomans, D.; Massart, D. L.; Kaufman, L. Optimization by statistical linear discriminant analysis in analytical chemistry. *Anal. Chim. Acta* **1979**, *112*, 97−122.
(20) Cole, R. A.; Phelps, K. Use of canonical variate analysis in the differentiation of Swede cultivars by gas−liquid chromatography of volatile hydrolysis products. *J. Sci. Food Agric.* **1979**, *30*, 669−676.
(21) Henrion, R.; Henrion, G. Hauptkomponentanalyse. In *Multivariate Datenanalyse*; Springer-Verlag: Berlin, Germany, 1995; pp 10−43.
(22) Fix, E.; Hodges, J. L. *Discriminatory Analysis: Nonparametric Determinations*; USAF School of Aviation Medicine: Rondolph Field, Texas, 1951; Consistency Properties, Vol. 4.
(23) González-Arjona, D.; López-Pérez, G.; González, A. G. Performing procrustes discriminant analysis with HOLMES. *Talanta* **1999**, *49*, 189−197.
(24) González-Arjona, D.; López-Pérez, G.; González, A. G. Holmes, a program for performing Procrustes Transformations. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 133−137.
(25) Ståhle, L.; Wold, S. Partial least-squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *J. Chemom.* **1987**, *1*, 185−196.
(26) Wold, S. Pattern recognition by means of disjoint principal components models. *Pattern Recognit.* **1976**, *8*, 127−139.
(27) González-Arjona, D.; González, A. G. Adaptation of linear discriminant analysis to second level-pattern recognition classification. *Anal. Chim. Acta* **1998**, *363*, 89−95.
(28) Cameán, A. M.; Moreno, I.; López-Artíguez, M.; Repetto, M.; González, A. G. Differentiation of Spanish brandies according to their metal content. *Talanta* **2001**, *54*, 53−59.
(29) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH: Weinheim, 1993; pp 119−148.

(30) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: New York, 1995; pp 116−163.
(31) García-Parrilla, M. C.; González, A. G.; Heredia, F. J.; Troncoso, A. M. Differentiation of wine vinegars based on phenolic composition. *J. Agric. Food Chem.* **1997**, *45*, 3487−3492.
(32) Herrador, M. A.; González, A. G. Pattern recognition procedures for differentiation of green, black and oolong teas according to their metal content from inductively coupled plasma atomic emission spectrometry. *Talanta* **2001**, *53*, 1249−1257.
(33) Sprecht, D. Probabilistic neural networks. *Neural Networks* **1990**, *3*, 109−118.
(34) Streit, R. L.; Luginbuhl, T. E. Maximum likelihood training of probabilistic neural networks. *IEEE Trans. Neural Networks* **1994**, *5*, 764−783.
(35) Headley, L. M.; Hardy, J. K. Classification of whiskies by principal component analysis. *J. Food Sci.* **1989**, *54*, 1351−1358.
(36) *STATISTICA for Windows (Computer Program Manual)*; StatSoft, Inc.: Tulsa, OK, 1999.
(37) *SIMCA P9 Multivariate Data Analysis Package*; Umetrics: Umea, Sweden, 2001.
(38) González-Arjona, D.; López-Pérez, G.; González, A. G. Nonlinear QSAR modeling by using multilayer perceptron feedforward neural networks trained by back-propagation. *Talanta* **2002**, *56*, 79−90.
(39) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: New York, 1995; pp 23−24.
(40) Cacoullos, T. Estimation of a multivariate density. *Ann. Inst. Stat. Math. (Tokyo)* **1966**, *18*, 179−189.
(41) Roggo, Y.; Duponchet, L.; Huvenne, J. P. Comparison of supervised pattern recognition methods with McNemar's statistical test. Application to qualitative analysis of sugar beet by near-infrared spectroscopy. *Anal. Chim. Acta* **2003**, *477*, 187−200.
(42) Hammond, M. K.; Riedel, C. J.; Rose-Pehrsson, S. L.; Williams, F. W. Training set optimization methods for a probabilistic neural network. *Chemom. Intell. Lab. Syst.* **2004**, *71*, 73−78.
(43) *Lab-Base GC-MS Data System*; Fisons: Manchester, United Kingdom, 1991.
(44) Kaiser, H. F. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141−151.
(45) Dahlberg, D. B.; Lee, S. M.; Wenger, S. J.; Vargo, J. A. Classification of vegetable oils by FT-IR. *Appl. Spectrosc.* **1997**, *51*, 1118−1124.
(46) Branden, K. V.; Hubert, M. Robust classification in high dimensions based on the SIMCA method. *Chemom. Intell. Lab. Syst.* **2005**, *79*, 10−21.
(47) Kempler, M. S.; Luchetta, L. M. A guide to raw material analysis using near-infrared spectroscopy. *J. Near Infrared Spectrosc.* **2003**, *11*, 155−174.

Received for review July 20, 2005. Revised manuscript received November 25, 2005. Accepted January 19, 2006.

JF0517389