7206 J. Agric. Food Chem. 2006, 54, 7206–7212

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Preliminary Chemometric Study on the Use of Honey as an Environmental Marker in Galicia (Northwestern Spain)

JUAN CARLOS RODRÍGUEZ GARCÍA,† ROBERTO IGLESIAS RODRÍGUEZ,§ ROSA MARÍA PEÑA CRECENTE,† JULIA BARCIELA GARCÍA,† SAGRARIO GARCÍA MARTÍN,† AND CARLOS HERRERO LATORRE*,†

Departamento Química Analítica, Nutrición y Bromatología, Facultad de Ciencias, Universidad de Santiago de Compostela, Augas Férreas s/n, Campus Universitario, 27002 Lugo, Spain, and Departamento Electrónica y Computación, Grupo de Sistemas Inteligentes, Facultad de Física, Universidad de Santiago de Compostela, Campus Universitario, 15706 Santiago de Compostela, Spain

Thirteen metal elements were determined in 40 honey samples from Galicia with different environmental origins: rural, urban, and industrial areas. The data set of the honey metallic profiles was studied with a double purpose: first, to make a preliminary evaluation of honey as an environmental indicator in Galicia with the aim of monitoring pollution and, second, to compare the different capabilities of diverse pattern recognition prediction procedures for modeling the environmental surrounding of the hive. A certain level of similarity for urban and industrial samples was obtained using principal component analysis and cluster analysis, whereas significant differences for urban and industrial honeys were found in relation to rural honey samples. Different classification rules to associate metal content of honeys with their environmental surrounding were obtained by chemometric pattern recognition procedures. In general, the classification methods developed by neural networks provided better results than the traditional pattern recognition procedures. The metal profiles of honey seem to provide sufficient information to enable categorization criteria for classifying samples according to their environmental surrounding. Thus, honey could be a potential pollution indicator for the Galician area.

KEYWORDS: Honey; environmental indicator; heavy metals; pattern recognition

INTRODUCTION

Honey is a natural product produced by *Apis mellifera* bees from nectar (nectar honey) and secretions of the living parts of plants and from excretions of plant-sucking insects (honeydew honey) (*1*). Honeybees forage in a great diversity of places and strongly interact with the environment surrounding the hive. Therefore, honey reflects the pollutants in or on the forage plants, the atmosphere, and the soil of the area in which the hive is located. Thus, bee products, especially honey, are considered as possible materials for monitoring environmental pollution (*2*).

Honey contains mainly sugars (fructose, glucose, sucrose, and others) and water, but different groups of substances such as flavonoids, organic acids, vitamins, hormones, flavor compounds, enzymes, and mineral elements are present in honey at lower concentrations (*3*). Minerals seem to be good candidates for honey characterization with monitoring purposes due to their stability in an acidic matrix (such as honey) and because their content in the product depends on their availability in the environment.

Under these considerations, in the past few years, investigation on the metal content in honey has been carried out with three aims. The first objective was to provide an assessment of the levels of the different elements in honey, with special interest in heavy metals. Different authors have measured the mineral content of honey from different botanical and country origins (*4*−*7*) using diverse analytical techniques such as flame and electrothermal absorption atomic spectrometry (*8*, *9*), inductively coupled plasma atomic emission spectrometry (*10*, *11*), ion chromatography (*12*), total reflection X-ray fluorescence spectroscopy (*13*), and neutron activation analysis (*14*).

The second main purpose of the metal determination in honey was to process the metallic data obtained with multivariate statistical techniques to establish the botanical or geographical origin of the product (*15*−*17*). This is a promising approach in which the use of chemometric techniques for data reduction and visualization (such as principal component analysis or hierarchical cluster analysis) has been successfully employed to reveal the complex structure of the metal data matrix studied. In addition, in other works (*18*−*22*), classification and prediction systems supported by pattern recognition procedures (such as

* Author to whom correspondence should be addressed (telephone +34 982285900; fax +34 982285872; e-mail cherrero@lugo.usc.es).
† Universidad de Santiago de Compostela, Lugo.
§ Universidad de Santiago de Compostela, Santiago de Compostela.

10.1021/jf060823t CCC: $33.50
Published on Web 08/23/2006

discriminant analysis, *K*-nearest neighbors, soft independent modeling of class analogy, and different types of neural networks) were developed to categorize honey samples (according to their geographical or botanical origin) on the basis of the chemical information contained in the metallic profile analyzed.

The third target of honey metal determination was to evaluate the opportunity of using honey for monitoring contamination in rural, industrial, or urban areas. Honeybee products, especially honey, have been considered as appropriate indicators for a variety of environmental pollutants due to the ability of bees to collect materials that reflect their immediate environmental conditions. In the past few years, researchers have investigated the relationship between the honey's metallic content (particularly heavy metals) and the presence of these metals in the environment. Barisic et al. (*23*) demonstrated the relevance of honeydew honey as an indicator for Cr, Cs, Rb, Cu, Pb, and Ni pollution in Croatia. Also, Przybylowski and Wilczyñska (*24*) proposed honey as a useful matrix for assessing the presence of environmental contaminants such as heavy metals in the Pomeranian region. Üren et al. (*25*) used the content of 9 metals in 74 Turkish honey samples during a 2-year period to evaluate the effect of a thermoelectric power plant in their surroundings. These authors showed the utility of honey as a monitoring indicator, but they also indicated that honeys with low mineral content are easily affected by vicinity contamination sources, whereas honeys with high amounts of total minerals have a certain resistance to the effects of pollution sources. Porrini et al. (*26*) investigated the use of honeybees for environmental monitoring in Italy with interesting results for pesticides, heavy metals (Pb, Ni, Cr), and radionuclide emissions from nuclear power plants. However, Devillers et al. (*27*, *28*), after the determination of 17 metallic and nonmetallic elements in 150 French acacia honeys from various identified polluted and unpolluted environments, judged that honey was not a good bioindicator for heavy metals and related elements. In spite of the fact that generally the most contaminated honeys correspond to samples produced in hives located in polluted areas, contamination can also be found in samples coming from apparently unpolluted regions.

Thus, the objective of this work was double: first, to study the possibility of employing honey samples from Galicia (northwestern Spain) as an environmental indicator to monitor the environmental pollution of the surroundings in which the hive is located; second, to evaluate the different capabilities of diverse pattern recognition procedures (based on different mathematical strategies) to deduce a classification rule that allows knowing the surrounding environmental characteristics (rural, urban, or industrial) from the metallic content of honeys. The classification systems were developed on the information enclosed in a reduced 40 × 13 data set formed by the concentrations of 13 elements measured in 40 Galician honey samples by atomic spectrometry.

MATERIALS AND METHODS

Honey Samples. Forty representative honey samples of guaranteed Galician origin and processed using the traditional procedures in the producing region were provided by the Certification Origin Council of the Certified Brand of Origin "Miel de Galicia" and by two Galician associations of beekeepers (APLA and AGA). Samples were harvested during 2004, and all of them were unprocessed honeys in order to avoid contamination. Twenty-two samples were produced in urban or industrial areas (for identification purposes, these were coded respectively 1 and 2), whereas the 18 remaining samples came from unpolluted rural areas (coded 3). The location chosen for hive placement was selected by taking into account information concerning Galician pollution condition. Thus, industrial samples were purchased at hives located in industrial areas affected by industrial emissions. Urban honeys were obtained from hives placed in metropolitan areas that have been proved to be influenced by city pollution. Rural samples were obtained from country and natural regions far from contamination sources and unaffected by them. All of the examined samples were honeys of random (mixed) floral type. Samples (1 kg) were collected in glass bottles and stored in the dark at 3−4 °C until analysis, and they did not undergo any treatment that could alter their composition.

Table 1. Mean, Standard Deviation (SD), Maximum, and Minimum Values for the Metals Determined in Galician Honeys

	mean	SD	maximum	minimum
[Ca], μg g^{-1}	91	31	185	41
[Cd], ng g^{-1}	2.5	1.0	6.2	0.8
[Cr], ng g^{-1}	12	5.6	41	5.6
[Cu], μg g^{-1}	0.8	0.4	2.0	0.08
[Fe], μg g^{-1}	2.0	2.0	21	nd[a]
[K], mg g^{-1}	1.4	0.6	3.0	0.3
[Li], μg g^{-1}	8.1	6.6	38	nd
[Mg], μg g^{-1}	103	69	308	18
[Mn], μg g^{-1}	7.8	4.2	20	0.5
[Na], μg g^{-1}	93	29	194	47
[Ni], ng g^{-1}	54	36	172	12
[Pb], ng g^{-1}	15	7.5	49	2.0
[Zn], μg g^{-1}	2.2	1.1	7.1	nd

[a] Not detected.

Analytical Determinations. Thirteen metals, Ca, Cd, Cr, Cu, Fe, K, Li, Mg, Mn, Na, Ni, Pb, and Zn, were measured in the honey samples. The analytical procedures employed have been described in detail in previous works: Pb, Cd, Cr, Cu, and Ni were determined by means of different electrothermal atomic absorption spectrometry (ETAAS) methods, which were developed for the measurements of these metals in the honey matrix (*8*, *9*, *29*). In all cases, the ETAAS determinations were performed using an atomic absorption spectrometer Varian-SpectrAA-600 (Varian Inc., Palo Alto, CA) with Zeeman correction equipped with a Varian GTA-100 electrothermal atomizer. To determine the remaining elements, 2.0 g of each honey sample was ashed (550 °C); the obtained ash was dissolved in 1 M HCl and made up with ultrapure water to 10 mL in a volumetric flask. This sample solution obtained was measured in an inductively coupled plasma atomic emission spectrometer Perkin-Elmer Optima 4300 DV (Perkin-Elmer Inc., Boston, MA) operating under the following conditions: power, 1450 W; plasma gas flow, 15 L min^{-1}; nebulizer flow, 0.6 L min^{-1}. The wavelengths for each element were the following: Ca, 317.9 nm; Fe, 238.2 nm; K, 766.5 nm; Li, 670.8 nm; Mg, 285.2 nm; Mn, 257.6 nm; Na, 589.6 nm; and Zn, 206.2 nm. All determinations were made in duplicate.

Data Analysis. A matrix (40 × 13) having rows representing the different analyzed honey samples (objects) and columns corresponding to the content of the 13 determined metals (variables or features) was constructed. Chemometric analyses were performed by means of the following statistical software packages: Statgraphics (*30*) was employed for principal component analysis (PCA) and cluster analysis (CA). Parvus (*31*) was used for linear discriminant analysis (LDA) and *K*-nearest neighbors (KNN), and Pirouette (*32*) was used for soft independent modeling of class analogy (SIMCA). All of the different neural network computations were done using different programs written in Matlab (*33*) code.

RESULTS AND DISCUSSION

The results of the 13 elements determined in Galician honey samples are summarized in **Table 1**. The levels obtained in Galician honeys were comparable to those reported by other authors in honey samples from Galicia (*4*) and, in general, they were slightly higher than the levels obtained in honeys from other European regions. Pb content for Galician samples (1.96−

49.4 ng g^{-1} range) was lower than observed in Turkish honeys (*25*) but higher than that measured in Italian samples by Caroli et al. (*10*, *34*). Other authors found a Pb content between 91 and 141 ng g^{-1} for eucalyptus Italian honeys (*35*). However, the range obtained for robinia honeys from the same origin was 23−31 ng g^{-1}. A similar pattern distribution was obtained for the Cd content of Galician honeys in relation with Turkish and Italian samples. Devillers et al. (*27*) reported a mean content of 0.152 μg of Cd g^{-1} for 19 positive responses in 86 French honeys analyzed (range = 0.08−0.25 μg g^{-1}). Chromium levels for Galician samples are in the 5.6−41.5 ng g^{-1} range. For this metal, there exists a great variability compared to the levels found by diverse authors for different geographic locations. Values from 2.73 to 39.0 ng g^{-1} were measured in Italian samples (*10*, *34*). Golob et al. (*13*) reported levels in the range of 0.78−3.55 μg g^{-1} for Slovenian honeys. Devillers et al. (*27*) detected Cr in 33 cases for 150 French acacia samples with a mean content of 0.187 ppm (range = 0.05−0.52). The Mn concentrations determined for Galician honeys were, in general, higher than those obtained for samples from other origins. The range for Galician honey (0.46−20.3 μg g^{-1}) was comparable to only those published for French samples (*27*) (0.11−42.8 μg g^{-1}). For other geographic sources such as Italy (*10*, *34*), Turkey (*25*), Anatolia (*36*), Slovenia (*13*), and Israel (*37*), the values obtained were significantly lower than those reported in this work. This result was conditioned by certain types of Galician soils that are brown mesotrophic soils rich in manganese.

In spite of certain differences in the mean values, there are not significant differences in the metallic levels of the Galician honey samples according to their different origins (rural, industrial, or urban). The examination of box−whisker plots constructed for each individual variable showed an overlap between ranges of concentration for the different origins in every element. Thus, none of the variables measured, by itself, was able to discriminate between the three established origins. Therefore, a multivariate approach must be evaluated.

Data Reduction and Visualization. The first approach when a multivariate data set is analyzed is to reveal its latent data structure. For this goal, two unsupervised chemometric procedures were applied: PCA and CA.

PCA. PCA is a data displaying technique that allows the visualization of such multidimensional data by reducing the dimensionality of the problem (in the case at hand to three dimensions). The data structure is preserved by retaining the maximum amount of variability present in the data (*38*). PCA transforms the original data matrix (X_{nxm}) into a product of two matrices. One of them contains information about the objects (scores matrix S_{nxm}), whereas the other shows information about the variables (loadings matrix L_{mxm}). Examination of score plots (samples projected in the space defined by principal components) allows a visualization of the latent data structure, whereas evaluation of loadings achieves a description of the relationships between variables and principal components. PCA was performed on the autoscaled data to avoid the effect of different size variables. (Autoscale pretreatment consisted in the subtraction of the variable mean and division by the standard deviation of the variable; thus, a set of new variables with zero mean and unity standard deviation was obtained.) The study of loadings for the variables in the first two principal components shows that Mg, Mn, K, Ca, Ni, and Li were the dominant features in the first principal component (all of them in the positive part), which represented 28.41% of the total data variability. Cu, Cd, Cr, and Ni (the positive part of PC2) and Fe, Zn, and Pb (negative part) dominate the second principal component, which accounted for 18.54% of the total data variance. Chemically, the first principal component or eigenvector can be interpreted in relation to the soil characteristics and agricultural practices, whereas the second principal component can be directly associated with the environmental pollution of the hive surroundings. When the score analysis of the honey samples on the space formed by the first three principal components (58.84% of total variability accounted) was carried out, a natural separation in two groups was detected. As can be seen in **Figure 1**, honey from rural areas formed a separate group in the positive part of PC1. The samples from the two other origins carried out just one group (in the negative part of PC1) including urban and industrial samples. This means that, in the 13-dimensional space, urban and industrial samples are distributed in a different area in relation to rural ones. Assignation of an unknown sample in one or the other class implies the deduction of a classification rule by applying supervised pattern recognition techniques (see Pattern Recognition Analysis).

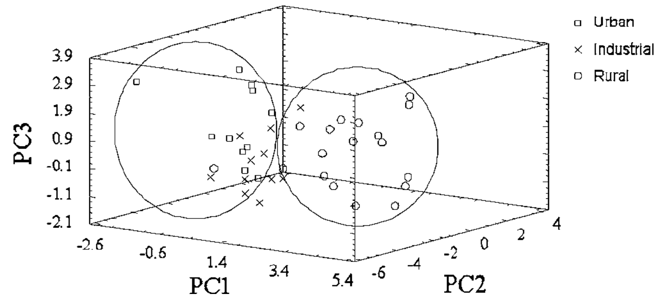

Figure 1. PCA score plot of the Galician honey samples.

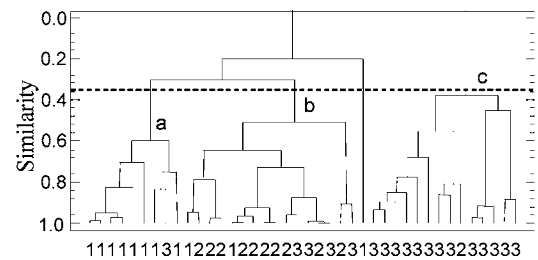

Figure 2. Dendrogram obtained from cluster analysis (Ward's method) of the Galician honey samples: 1, urban samples; 2, industrial samples; 3, rural samples.

CA. Due to its unsupervised character, CA is a complementary technique commonly applied together with PCA to describe the structure of a data set before other multivariate pattern recognition procedures are employed. It reveals the natural groupings existing between the samples characterized by the values of a set of measured variables. Thus, CA was applied to the autoscaled data to achieve this objective. The similarities between samples were calculated on the basis of the squared Euclidean distance, whereas Ward's method was used as linkage procedure to establish the clusters (*39*). The results achieved are presented in a graphic mode as the dendrogram of **Figure 2**. At the indicated similarity level of 0.45, three clusters (coded a−c) can be identified. From the left, the first cluster (cluster a) is a group made of urban samples that includes one rural sample. The second cluster (b) is mainly formed by industrial honey samples, in spite of the inclusion of certain samples from the other two origins. The third cluster (c) includes only samples of rural origin. CA confirmed the separation between the rural and other samples shown previously by PCA. In addition, the visualization of the dendrogram supports the overlapping of the categories urban and industrial in the 13-dimensional space of the variables. It can be concluded that the available metal data may provide enough information to develop classification systems to differentiate rural honey samples representing unpolluted areas from urban and industrial honey samples representing polluted zones.

The two unsupervised displaying procedures used, PCA and CA, produced consistent results for the study of latent structures in the honey data set. Both chemometric techniques indicated the separation between two classes, rural unpolluted versus urban and industrial polluted honey samples. Therefore, classification procedures were applied. Different pattern recognition techniques, which essentially differ in the way they define the classification rule, were developed to assign each of the studied honey samples to one of the two above-defined categories.

Pattern Recognition Analysis. In this section, metal data of the honeys were processed by chemometric pattern recognition procedures (such as LDA, KNN, SIMCA, and different types of neural networks) aimed at developing different classification rules to relate the metallic content of the analyzed honeys with their environmental surrounding. According to the results achieved under Data Reduction and Visualization, two different categories were established. The first category is composed of rural honey samples from unpolluted areas (rural unaffected, RU); the second group is made of honeys from urban and industrial polluted zones (urban industrial polluted, UIP). Several supervised pattern recognition methods (which essentially differ in the way they define the classification rule) were applied to the autoscaled data matrix $X_{40\times13}$ to obtain categorization procedures for RU and UIP groups. In all cases, the training set (used to derive the decision rule) consisted of a subset performed by 75% of the samples. The test set (used to validate the decision rule obtained in the learning step) was composed of the remaining 25% samples. Both sets were obtained randomly. With this set arrangement, an adequate number of samples were utilized in the training procedure and a representative number of members formed the test set used to validate the derived classification model as well. To perform a cross-validation procedure, the same process was repeated four times with different constitutions of both sets, to ensure that all samples were included in the evaluation set at least once. The success of the different classification systems was expressed as recognition ability (percentage of training set members correctly classified) and prediction ability (percentage of test set members adequately classified by using the rules developed in the training step).

LDA. The LDA searched for optimal linear boundaries between categories in the 13-dimensional space (equal to the number of variables). The discriminant function (which is a linear combination of the original variables) is established to achieve the maximum discrimination among the given categories, maximizing the ratio of between-class variance to the within-class variance in any particular data set to guarantee the maximal class separability (*40*). The application of LDA for the different data matrix described above developed an unsuccessful discriminant function; the achieved percentages of correct recognition and prediction are summarized in **Table 2**. Satisfactory results in the 92–99% range were attained for recognition ability; however, the classification rule obtained is not adequate for generalization, because poor results for UIP and RU classes with a prediction percentage of success between 71.8 and 81.6% were achieved. The inappropriate results reached by this pattern recognition procedure could be expected by taking into account the LDA assumption that classes must be linearly separated. In the case at hand, they are not, as can be shown in the score plot of **Figure 1**.

Table 2. Classification Results for the Compared Supervised Pattern Recognition Procedures

	category	recognition ability (%)	prediction ability (%)
LDA	UIP	99.2	81.6
	RU	92.4	71.8
KNN	UIP	88.2	87.5
	RU	70.0	94.1
SIMCA	UIP	85.3	85.0
	RU	94.6	93.7
MLF-ANN	UIP	100	90.0
	RU	100	90.1
AVQ-ANN	UIP	100	95.0
	RU	100	82.5
VQBCP-ANN	UIP	100	80.1
	RU	100	77.5
LVQ-ANN	UIP	100	100
	RU	100	82.5

KNN. KNN classifies an unknown object in the category to which the *K* nearest known objects of the training set belong (*41*). The *K*-nearest neighbor classifier labels an unknown object within the class of the majority of the *K* nearest neighbors. A neighbor is deemed to be nearest if it has the smallest distance in feature space (in the Euclidean sense). Thus, the existence of a local metric related to the similarity between objects constitutes the basis of this pattern recognition system. In this work, the Euclidean distance was used as local metric and the importance of a given neighbor in the class assignation was weighted on the basis of the inverse squared Euclidean distance. The value of *K* was selected by optimization determining the classification ability with different values of *K* between 1 and 5; the selected value of *K* found to be optimal for further applications of KNN was 3. Under these specifications, KNN was also applied to the four data sets as described above. Unsuccessful results of correct recognition and prediction abilities were attained for the two classes, especially for the RU category in which the rule derived was demonstrated not to be stable because a higher value was obtained for prediction than for recognition ability.

SIMCA. SIMCA is a modeling pattern recognition procedure. Classification is performed in the SIMCA approach to identify local models for possible groups and to predict a probable class membership for new observations. At first, this approach runs a global PCA regression (according to the available data structure) on the whole dataset to identify groups of observations. Local models are then estimated for each class. Finally, new observations are classified into one of the established class models on the basis of their best fit to the respective model (*42*). In this work, SIMCA (using the cross-validation procedure in four steps as described above) afforded three component models for the two categories considered. These results were studied in terms of recognition and prediction abilities (see **Table 2**). The classification model developed by SIMCA produced good classification and prediction for the RU class; in practice, most of the RU samples (94%) were assigned to their category. However, the RU model developed by SIMCA also accepted 5% of UIP samples as RU. On the other hand,

Table 3. MLF-ANN Architectures Assayed

MLF network architecture	root-mean-square error	
	training set	test set
13–5–1	0.0331	0.330
13–5–2	0.0279	0.250
13–7–2	0.0277	0.182
13–9–2	0.0273	0.142

the UIP model achieved a minor level of hits, and there exists a certain probability that the model could show RU sample as UIP.

Multilayer Feed-Forward Artificial Neural Networks (MLF-ANN). MLF-ANN is a supervised system that builds a model based on a set of samples with known input−output (training set). The learning strategy consists of updating the weights of the connections between neurons to reach the appropriate output for each input. Once the network has been trained and validated, MLF-ANN can be used for classification purposes; the final weights provide enough information to predict the target output (category) based on the inputs (chemical variables) (*43*). For the case at hand, a MLF-ANN network was employed to predict the honey category (UIP or RU) on the basis of an input pattern consisting of the 13 autoscaled chemical variables. The first step in the use of a MLF-ANN must be the selection of an adequate network. Different network architectures were assayed using the complete data set (see **Table 3**). All of the structures presented 13 neurons in the input layer, different numbers of neurons in the hidden layers, and an output layer with 1 or 2 neurons. When the output layer was composed by only one neuron, the target output was written $y = (1)$ for class UIP and $y = (0)$ for the RU category. In the case of networks with structures in which the output layer was performed by two neurons, the target was codified $y = (1,0)$ for UIP category and $y = (0,1)$ for the RU class. The best result was obtained using a three-layer MLF-ANN having an input layer with 13 neurons, a hidden layer with 9 neurons, and an output layer consisting of 2 neurons with a binary target output as described above. The transfer function used was sigmoidal $f(x) = 1/(1 + [\exp(-x)])$. The initial values of the weights associated with the connections between neurons were selected randomly in the range from −3 to 3. The maximum number of epochs was limited to 2000 to avoid overfitting. Validation of the MLF-ANN classification results was carried out by the cross-validation process in four steps above indicated. As can be seen in **Table 2**, the classification rule developed by MLF-ANN exhibited a very good result for recognition performance (100%) in the two groups, achieved in 400−500 epochs. The validation evaluated by means of test sets showed some errors in prediction for both classes (10%).

Vector Quantization Classification Systems. Three different neural networks based on vector quantization were employed. Vector quantization is a classical signal approximation method (*44*) that usually forms an approximation to the probability density function of a pattern distribution, using a finite number of reference vectors $W = \{w_1, ..., w_N\}$. These N-reference vectors permit the partition of multidimensional space into a set of un-overlapped regions, known as Voronoi regions. Each Voronoi region includes all of the space points having a common nearest reference vector. In the present case, a neural network based on the accurate vector quantization (AVQ) algorithm of the training set was applied (*45*). In fact, the algorithm used was capable of obtaining a set of reference vectors W in such a way that each of the Voronoi regions includes only training set objects belonging to one class. Because of this approach, an unknown object will be classified (as UIP or RU class) according to its associated Voronoi region. The procedure was applied to the autoscaled data matrix; learning and validation steps were based on the division between training and test sets described under Pattern Recognition Analysis. When this algorithm was applied, all of the samples in the training set were correctly classified in their corresponding classes (100% of recognition ability). The results in the test sets were successful for the UIP class (95.0% correct prediction), but unsatisfactory for the RU category, in which only the 82.5% of the samples were correctly assigned to their class.

At this point, and to improve the performance of this classifier through a better drawing of the boundaries between classes, a modified vector quantization-based classification procedure (VQBCP) was applied. The neural network supported on the VQBCP method has been described in detail in a previous work (*46*). By means of this modification, information concerning the class separability criteria is taken into account. The classification system was completed with two other phases that seek better adaptation to the specific characteristics of each of the regions of the 13-dimensional space. Thus, the new learning procedure consisted of three steps. In the first one, the previously described partition of the input 13-dimensional space $\mathscr{R}^{13}$ was made by AVQ. In the second step, information from the first level (relevant to the identification of the space zones in which the separation between classes is lower: higher risk of incorrect classifications) is used to analyze the distribution of the training set objects in the Voronoi regions obtained in the first step. Finally, in the third step, the information learned in step 2 was used to obtain a more suitable repartition of $\mathscr{R}^{13}$ in Voronoi regions (each of them again associated with only one of the specified classes, UIP or RU). In spite of this more sophisticated approach, the results obtained, in terms of prediction, were worse than the achieved by AVQ (see **Table 2**). It can be concluded that this classification strategy strongly depends on the quality of the training set. In our experience, VQBCP does not produce good classifications when there is a certain noise level in the training set or when the categories are partially overlapped in the multidimensional space.

A third and last classification procedure based on the vector quantization principle was applied by means of a neural network based on the learning vector quantization (LVQ) algorithm (*47*). In this case, the partition of the multidimensional space into the Voronoi regions is not focused in order to obtain the correct assignation of all the training set; the LVQ strategy is determined with the aim of minimizing the classification error for the training set members with higher values for the probability density function $p(x)$. The boundaries between classes were computed following the Bayes decision surfaces for the minimum classification error probability (*48*). This procedure produced very good results for the UIP class, achieving a total percentage of prediction for this category (100%). However, the prediction ability for the RU class was unsatisfactory (82.5%). This procedure was able to recognize all polluted surroundings, but there is a certain risk of classifying an unpolluted area as contaminated.

The flexibility, parallel process, and high tolerance to noisy data demonstrated by the neural networks are advantages compared to the classical chemometrical classification techniques; however, the poor chemical information provided is a serious disadvantage. Of the neural networks assayed, those developed on the basis of of multilayer feed-forward with learning by back-propagation (MLF) and learning vector quantization (LQV) were the best. In both cases, the prediction abilities are satisfactory, and, in addition, the number of parameters to be optimized is lower.

In conclusion, it can be seen that the use of honey as an environmental pollution marker in Galicia seems to be a promising approach to differentiate between industrial–urban polluted and rural unpolluted areas. The chemical information provided by the honey metal profiles processed by different neural networks allows the development of successful classification models that reflect the environmental conditions of the surroundings in which the hive is located.

Although Galician honeys presented slightly higher metal levels than other honeys from other European countries, it was observed that Galician honey samples were differently affected according to their origin (rural unpolluted or urban–industrial polluted). From this point of view, a more comprehensive study with more honey samples and during a longer period is necessary for a definitive evaluation of the role of honey as a suitable bioindicator as well as for an estimation of the mineral content effect on the stability from the effects of contamination sources, as discussed by Üren et al. (*25*).

LITERATURE CITED

(1) Council Directive 2001/110/EC of 20 December 2001 relating to honey. *Off. J. Eur. Communities* **2002**, *10*, 47–52.

(2) Devillers, J.; Pham-Derlegue, M. H. *Honey Bees: Estimating the Environmental Impact of Chemicals*; Taylor and Francis: London, U.K., 2002.

(3) White, J. W. Composition of honey. In *Honey: A Comprehensive Survey*; Crane, E., Ed.; Heinemann: London, U.K., 1975; pp 180–194.

(4) Rodríguez-Otero, J.; Paseiro, P.; Simal, J.; Cepeda, A. Mineral content of the honeys produced in Galicia (North-West Spain). *Food Chem.* **1994**, *49*, 169–171.

(5) Pietra, R.; Serra, M. A.; Fortaner, S.; Edel, J.; Minoia, C. Application of nuclear, radiochemical and spectrochemical techniques in metal toxicology. *J. Trace Microprobe Technol.* **1993**, *11*, 217–233.

(6) Stein, K.; Umland, F. Trace analysis of lead, cadmium and manganese in honey and sugar. *Fresenius' J. Anal. Chem.* **1986**, *323*, 176–177.

(7) Fodor, P.; Molnar, E. Honey as environmental indicator: effect of sample preparation on trace element determination by ICP-AES. *Microchim. Acta* **1993**, *112*, 113–118.

(8) Rodríguez, J. C.; Barciela, J.; Herrero, C.; García, S.; Peña, R. M. Comparison of palladium–magnesium nitrate and ammonium dihydrogenphosphate modifiers for lead determination in honey by ETAAS. *Food Chem.* **2005**, *91*, 435–442.

(9) Rodríguez, J. C.; Barciela, J.; Freire, M.; Herrero, C.; García, S.; Peña, R. M. Comparison of palladium–magnesium nitrate and ammonium dihydrogenphosphate modifiers for cadmium determination in honey by ETAAS. *Talanta* **2003**, *61*, 509–517.

(10) Caroli, S.; Forte, G.; Iamiceli, A. L.; Galoppi, B. Determination of essential and potentially toxic trace elements in honey by inductively coupled plasma based techniques. *Talanta* **1999**, *50*, 327–336.

(11) Ioannidou, M. D.; Zachariadis, G. A.; Anthedimis, A. N.; Stratis, J. A. Direct determination of heavy and transition metals in honey and sugars using inductively coupled plasma atomic emission spectrometry. *Talanta* **2005**, *65*, 92–97.

(12) Buldini, P. L.; Cavalli, S.; Mevolli, A.; Sharma, J. L. Ion chromatographic determination of heavy and transition metals in honey. *Food Chem.* **2001**, *73*, 487–495.

(13) Golob, T.; Dobersek, U.; Kump, P.; Necemer, M. Determination of trace and minor elements in Slovenian honey by total reflection x-ray fluorescence spectroscopy. *Food Chem.* **2005**, *91*, 593–600.

(14) Sevimli, H. N.; Bayulgen, A.; Varinlouglu, J. Determination of trace elements in honey by instrumental neutron activation analysis. *J. Radioanal. Nucl. Chem.* **1992**, *165*, 319–325.

(15) Feller, M.; Vincent, B.; Beaulieu, F. Teneur en minéraux et origine géographique des miels du Canada. *Apidologie* **1989**, *20*, 77–91.

(16) Rashed, M. N.; Soltan, M. E. Major and trace elements in different types of Egyptian mono-floral and non-floral bee honeys. *J. Food Compos. Anal.* **2004**, *17*, 737–747.

(17) Downey, G.; Hussey, K.; Kelly, J. D.; Walshe, T. F.; Martin, P. G. Preliminary contribution to the characterization of artisanal honey produced on the island of Ireland by palynological and physico-chemical analysis. *Food Chem.* **2005**, *91*, 347–354.

(18) Latorre, M. J.; Peña, R.; Pita, C.; Botana, A.; García, S.; Herrero, C. Chemometric classification of honeys according to their type. II. Metal content data. *Food Chem.* **1999**, *66*, 263–268.

(19) Hernández, O. M.; Fraga, J. M.; Jiménez, A. I.; Jiménez, F.; Arias, J. J. Characterization of honey from the Canary Islands: determination of the mineral content by AAS. *Food Chem.* **2005**, *93*, 449–458.

(20) Latorre, M. J.; Peña, R.; García, S.; Herrero, C. Authentication of Galician (NW Spain) honeys by multivariate techniques based on metal content data. *Analyst* **2000**, *125*, 307–312.

(21) Serrano, S.; Villarejo, M.; Espejo, R.; Jodral, M. Chemical and physico-chemical parameters of Andalusian honey: classification of *Citrus* and *Eucaliptus* honeys by discriminant analysis. *Food Chem.* **2004**, *87*, 619–625.

(22) Nozal, M. J.; Bernal, J. L.; Diego, J. C.; Martín, M. T. Classifying honeys from the Soria province of Spain via multivariate analysis. *Anal. Bional. Chem.* **2005**, *382*, 311–319.

(23) Barisic, D.; Bromeshenk, J. J.; Kezic, N.; Vertacnik, A. The role of honey bees in environmental monitoring in Croatia. In *Honey Bees: Estimating the Environmental Impact of Chemicals*; Devillers, J., Pham-Derlegue, M. H., Eds.; Taylor and Francis: London, U.K., 2002; pp 161–185.

(24) Przybylowski, P.; Wilczyñska, A. Honey as an environmental marker. *Food Chem.* **2001**, *74*, 289–291.

(25) Üren, A.; Serifoglu, A.; Sarikahya, Y. Distribution of elements in honeys and effect of a thermoelectric power plant on the element contents. *Food Chem.* **1998**, *61*, 185–190.

(26) Porrini, C.; Ghini, S.; Girotti, S.; Sabatini, A.G.; Gattavechia, E.; Celli, G. Use of honey bees as bioindicators of environmental pollution in Italy. In *Honey Bees: Estimating the Environmental Impact of Chemicals*; Devillers, J., Pham-Derlegue, M. H., Eds.; Taylor and Francis: London, U.K., 2002; pp 185–247.

(27) Devillers, J.; Dore, J. C.; Marenco, M.; Poirier-Duchene, F.; Galand, N.; Viel, C. Chemometrical analysis of 18 metallic and nonmetallic elements found in honey sold in France. *J. Agric. Food Chem.* **2002**, *50*, 5998–6007.

(28) Devillers, J.; Dore, J. C.; Marenco, M.; Poirier-Duchene, F.; Galand, N.; Subirana, M. Typology of French acacia honeys based on their concentrations in metallic and non-metallic elements. In *Honey Bees: Estimating the Environmental Impact of Chemicals*; Devillers, J., Pham-Derlegue, M. H., Eds.; Taylor and Francis: London, U.K., 2002; pp 248–268.

(29) Rodríguez, J. C.; Barciela, J.; Herrero, C.; García, S.; Peña, R. Direct and combined methods for the determination of chromium, copper, and nickel in honey by electrothermal atomic absorption spectroscopy. *J. Agric. Food Chem.* **2005**, *53*, 6616–6623.

(30) *Statgraphics*, version 5.0; Statistical Graphics Corp.: Rockville, MD, 1991.

(31) Forina, M.; Leardi, R.; Armanino, C.; Lanteri, S. *Parvus: An Extendable Package of Programs for Data Exploration, Classification and Correlation*; Elsevier: Amsterdam, The Netherlands, 1988.

(32) *Pirouette: Multivariate Data Analysis*, version 2.51; Infometrix: Woodinville, WA, 1998.

(33) *MATLAB*, version 7.1; MathWorks: Natick, MA, 2005.

(34) Caroli, S.; Forte, G.; Alessadrelli, M.; Spagnoli, R. M.; D'Ilio, S.; Pauwels, J.; Kramer, G. N. A pilot study for the production of a certified reference material for trace elements in honey. *Mikrochem. J.* **2000**, *67*, 227−233.

(35) Bettinelli, M.; Baroni, U.; Spezia, S.; Terni, C. Determination of trace elements in honey using ETV-ICP-MS. *At. Spectrosc.* **2000**, *21*, 195−204.

(36) Yilmaz, H.; Yavuz, Ö. Content of some trace elements in honey from South-eastern Anatolia. *Food Chem.* **1999**, *65*, 475−476.

(37) Merin, U.; Bernstein, S.; Rosenthal, I. A parameter for quality of honey. *Food Chem.* **1998**, *63*, 241−242.

(38) Jolliffe, I .T. *Principal Component Analysis*; Springer: New York, 2002.

(39) Massart, D. L.; Kaufman, L. *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*; Wiley: New York, 1983.

(40) Brereton, R. G. *Chemometrics, Applications of Mathematics and Statistics to Laboratory Systems*; Ellis Horwood: New York, 1990; p 263.

(41) Vandeginste, B. G.; Massart, L.; Buydens, L. M.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B*; Elsevier: Amsterdam, The Netherlands, 1998; p 201.

(42) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. Multivariate data analysis in chemistry. In *Chemometrics, Mathematics and Statistics in Chemistry*; Kowalski, B. R., Ed.; Riedel Publishing: Dordrecht, The Netherlands, 1984; pp 17−96.

(43) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH: New York, 1993; pp 119−148.

(44) Kohonen, T. *Self-Organizing Maps*; Springer: Berlin, Germany, 1997; p 48.

(45) Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*; Academic Press: New York, 1998; p 90.

(46) Peña, R. M.; García, S.; Iglesias, R.; Barro, S.; Herrero, C. Authentication of Galician (NW Spain) quality brand patatoes using metal analysis. Classical pattern recognition techniques vs. a new vector quantization-based procedure. *Analyst* **2001**, *126*, 2186−2193.

(47) Kohonen, T. Learning vector quantization and the self-organizing map. In *Theory and Applications of Neural Networks*; Mannion, C. L. T., Taylor J. G., Eds.; Springer-Verlag: New York, 1992; pp 235−242.

(48) Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*; Springer-Verlag: New York, 1985.

Received for review March 3, 2006. Revised manuscript received July 18, 2006. Accepted July 21, 2006. We express thanks to the Instituto Lucense de Desarrollo (INLUDES) for financial support.

JF060823T