JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Classification and Class-Modeling of "Riviera Ligure" Extra-Virgin Olive Oil Using Chemical−Physical Parameters

RAFFAELLA BOGGIA,*,† PAOLA ZUNIN,† SILVIA LANTERI,† NICOLETTA ROSSI,‡ AND FILIPPO EVANGELISTI†

Dipartimento di Chimica e Tecnologie Farmaceutiche e Alimentari, Università di Genova, Via Brigata Salerno (ponte), I-16147 Genova, Italy, and Laboratorio Regionale di Analisi dei Terreni e Produzioni Vegetali, I-19038 Sarzana (SP), Italy

The Protected Designation of Origin (PDO) "Riviera Ligure" for extra-virgin olive oils from Liguria specifies three additional geographical mentions corresponding to three different geographical areas. To obtain a complete characterization of this typical Italian product, 217 samples of olive oils produced in this North Italian region during 1998/99 and 1999/2000 were analyzed. For each sample 31 variables were determined by chemical−physical analyses, and the data were subjected to a multivariate statistical analysis. For the 1998/99 crop, characterized by favorable climatic conditions, class-models of the three geographical areas were obtained with good predictive ability, also considering the influence of the month of olive harvesting. The oil samples from the 1999/2000 crop were similarly studied, but bad climatic conditions and a widespread *Dacus oleae* infestation leveled out the peculiar features of the oils produced in the three areas.

KEYWORDS: Extra-virgin olive oil; Riviera Ligure; PDO; chemical−physical parameters; classification; class-modeling; harvesting and agronomic conditions

INTRODUCTION

Extra-virgin olive oil has a highly variable chemical composition. This variability mostly depends on olive cultivars, climatic conditions, and agricultural practices.

Liguria is a North Italian region that stretches along a wide tract of the Mediterranean coast. Its eastern and western coasts (Rivieras) have considerably different orographic and climatic characteristics. Moreover, olive groves are present both along the whole coastline and in a large part of its inland, and their numerous cultivars are not homogeneously spread throughout the Region. For example, in La Spezia province (East Liguria), there are some cultivars, such as Razzola and Lantesca, which are quite absent in the other areas. Cv Taggiasca, on the other hand, is largely grown in Imperia province (West Liguria), whereas in the center of the region several cultivars, such as Pignola and the already mentioned Taggiasca, are grown together. Thus, extra-virgin olive oils from Liguria show a rather different chemical composition, even if more homogeneous characteristics can be outlined in oils coming from the three distinct areas formed respectively by the province of Imperia, the province of Savona, and the provinces of Genoa and La Spezia together.

For this reason extra-virgin olive oils from Liguria received the Protected Designation of Origin (PDO) "Riviera Ligure" (*1*) with three additional geographical mentions − "Riviera dei Fiori", "Riviera del Ponente Savonese", and "Riviera di Levante" − approximately corresponding to the geographical areas mentioned above (**Figure 1**). For each area the PDO indicates some special organoleptic features together with the minimum acceptable panel test score. Moreover, maximum acceptable acidity, peroxide number, and UV absorbance at 232 and 270 nm are reported, but only the values for acidity and peroxide number are slightly different. These few analytical parameters are certainly not sufficient to distinguish the three groups.

To obtain a more complete characterization, more than 200 samples of extra-virgin olive oils from Liguria were analyzed. The chemical-physical parameters reported in the PDO Regulation were determined, together with several further analytical variables related to oil composition. Then class-modeling methods were applied. This study aims at classifying oils on the basis of their geographical origin and at defining statistical models capable of determining the possible origin of unknown samples.

Olive oil samples produced from 1998/99 and 1999/2000 crops were studied. These years were particularly meaningful because they showed clearly different climatic conditions. In 1998−99 high summer temperatures and poor autumn rainfalls contributed to limiting *Dacus oleae* infestation and improving oil quality. On the contrary, in the following year lower summer temperatures favored the spreading of the infestation, and strong autumn winds and rainstorms further worsened olive oil quality.

MATERIALS AND METHODS

Samples. A total of 217 olive oils from the three geographical areas mentioned in the PDO regulation were analyzed. In particular, the olive

* Corresponding author. Phone: +39-010-353 2643/2631. Fax: +39-010-353 2684. E-mail: raffa@dictfa.unige.it.
† Università di Genova.
‡ Laboratorio Regionale di Analisi dei Terreni e Produzioni Vegetali.

10.1021/jf011289m CCC: $22.00
Published on Web 03/12/2002

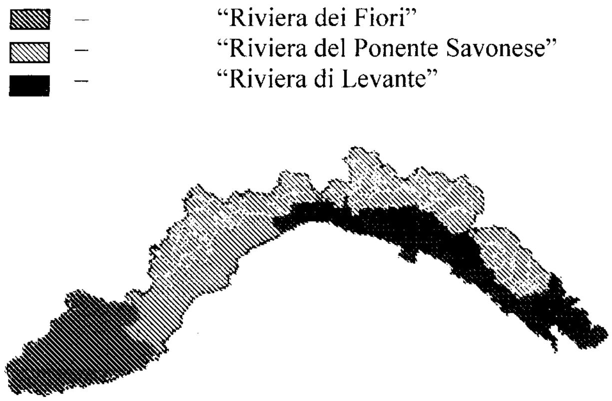

Figure 1. Geographical areas mentioned in the Protected Designation of Origin (PDO) Regulation for “Riviera Ligure”(*1*).

Table 1. Physical–Chemical Variables

no.	variable
1	acidity (% oleic acid)
2	peroxide value (meq O_2/kg)
	UV indexes
3	K_{232}
4	K_{270}
5	ΔK
	fatty acids (g/100 g)
6	palmitic
7	palmitoleic
8	eptadecanoic
9	eptadecenoic
10	stearic
11	oleic
12	linoleic
13	linolenic
14	arachic
15	eicosaenoic
	polar compounds[a]
16	3,4-DHPEA[b]
17	*p*-HPEA[c]
18	3,4-DHPEA-EDA[d]
19	total
20	S (*3*)
21	tocopherols (mg/kg oil)
22	% LLL[e]
	sterols (g/100 g)
23	cholesterol
24	24-methylencholest.
25	campesterol
26	campestanol
27	stigmasterol
28	sitosterol
29	Δ7-stigmastenol
30	Δ7-avenasterol
31	total sterols (mg/kg of oil)

[a] Polar compounds are expressed in mg/kg oil as HPEA. [b] 3,4-DHPEA= 3,4-dihydroxyphenylethanol. [c] *p*-HPEA = *p*- hydroxyphenylethanol. [d] 3,4-DHPEA-EDA = dialdehydic form of elenolic acid bonded with 3,4-dihydroxyphenylethanol. [e] LLL = trilinolein.

oil productions of 1998/99 and 1999/2000 were monitored by studying 107 samples from the first crop and 110 samples from the second crop.

Analytical Methods. Chemical−physical analyses were performed in order to determine the 31 variables reported in full in **Table 1**. Free acidity, peroxide value, UV absorbance, fatty acid composition, trilinolein content, and sterol fraction composition and content were determined according to European Community Regulation no. 2568/91 (*2*). Polar compounds were extracted from oils and analyzed by HPLC−DAD as previously reported by Evangelisti et al. (*3*). Tocopherols analysis was carried out by normal-phase HPLC according to American Oil Chemists’ Society methods (AOCS) (*4*). With regard to the determined variables, each sample can be defined as extra-virgin olive oil according to the European Community Regulation (*2*).

Multivariate Statistical Analysis. Several chemometric techniques were used to characterize this typical product on the basis of its chemical composition.

For each crop, a data matrix having as many rows as there were oil samples and 31 columns (the 31 chemical variables reported in **Table 1**) was built and studied by chemometric tools (set A and set B, respectively, for 1998/99 and 1999/2000). In both cases the samples were divided into three categories corresponding to the already-mentioned geographical areas according to their origin: Category 1, oils from “Riviera dei Fiori” (37 samples in set A, 26 samples in set B); Category 2, oils from “Riviera del Ponente Savonese” (23 samples in set A, 33 samples in set B); and Category 3, oils from “Riviera di Levante” (47 samples in set A, 51 samples in set B).

The univariate statistical parameters of the 31 variables for each crop and location are reported in **Table 2**.

At the beginning, an explanatory examination of the autoscaled data was performed using the principal component analysis (PCA) (*5*), a well-known technique to extract, rationalize, and visualize all useful information from the data set. It involves an orthogonal rotation that transforms the original variables into uncorrelated variables (new axes) called *principal components*, ordered according to their explained variance. The coefficients of the original variables defining each principal component are called “loadings”, and the projections of the objects on the new axes are called “scores”.

Afterward, classification techniques were applied in order to calculate classification rules allowing a discrimination among classes using chemical−physical variables only.

Linear discriminant analysis (LDA) (*6*) is a multivariate probabilistic classification method based on the use of multivariate probability distribution, under the hypothesis of normal distribution with the same variance−covariance matrix in all the considered classes.

Unequal class modeling (UNEQ) (*7, 8*) is the class modeling version of quadratic discriminant analysis (QDA), another probabilistic technique, which differs from LDA as it builds a category model (barycenter, a point in the multidimensional space) assuming a different multivariate normal distribution for each category; it uses Bayes theorem. In general, a class-modeling technique builds a model of the studied category: i.e., samples fitting the model are classified in the target category; the other objects are classified as rejected by the target category.

For each class, UNEQ defines the mathematical model and the category space around it as the confidence hyperellipsoid of the category, according to the Mahalanobis distance from the centroid. Confidence intervals can be built at different levels of significance; in this study a 95% level of confidence was considered. This method was applied to detect outliers and to assign samples with unknown geographical origin to their true category.

It must be underlined that LDA and UNEQ can be used only with well-defined data matrixes, i.e., those having a high ratio between number of objects and number of variables inside each category. To increase this ratio and to select the most discriminant variables among categories, stepwise linear discriminant analysis (SLDA) (*6*) was used.

To verify the models obtained by UNEQ, soft independent modeling of class analogy (SIMCA) (*9*) was applied to the original data sets. For each class of samples a PCA model was built; this model was based on the optimum number of components that best clusters the individual class, after a separate scaling for each category. The optimum number of components can vary from class to class; in this study this number is the one for which the explained variance of each class is 95%.

Nonerror-rate (the percentage of correctly classified objects), sensitivity (the nonerror-rate for a class), and specificity (the percentage of objects of other classes rejected by the class model under study) of the obtained models are the criteria used to measure the classification and the modeling performances (*10*).

All the classification and modeling methods build models that must be validated. In this study, the leave-one-out validation method was

Table 2. Univariate Statistical Parameters of the 31 Measured Variables for Each Crop (Set A and Set B) and Location

1998/99 crop

	set A 107 samples		Riviera dei Fiori 37 samples		Riviera del Ponente Savonese 23 samples		Riviera di Levante 47 samples	
V[a]	*m*[b]	*s*[c]	*m*	*s*	*m*	*s*	*m*	*s*
1	0.4	0.3	0.3	0.2	0.5	0.3	0.5	0.3
2	9	3	9	3	9	3	9	3
3	1.62	0.15	1.60	0.14	1.68	0.16	1.61	0.14
4	0.09	0.02	0.08	0.01	0.09	0.03	0.10	0.02
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	11.6	1.1	11.0	1.0	11.7	1.1	12.1	0.9
7	0.9	0.2	0.8	0.2	1.0	0.3	0.9	0.2
8	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0
9	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0
10	2.1	0.2	2.1	0.2	2.0	0.2	2.1	0.2
11	77.1	2.1	78.0	2.0	76.1	2.8	76.8	1.4
12	6.9	1.1	6.8	0.9	7.8	1.7	6.6	0.6
13	0.6	0.1	0.6	0.0	0.6	0.1	0.7	0.1
14	0.4	0.0	0.3	0.0	0.3	0.0	0.4	0.0
15	0.3	0.0	0.3	0.0	0.3	0.0	0.3	0.1
16	1.0	2.9	0.4	0.8	0.6	1.0	1.8	4.2
17	4.1	3.4	4.3	3.5	3.8	2.6	4.2	3.7
18	9.0	16.6	7.5	15.2	5.6	9.3	11.8	20.0
19	118.9	60.2	133.2	56.5	86.4	43.6	123.5	65.0
20	54.7	27.3	40.1	15.8	40.6	15.0	73.1	28.2
21	98.9	33.2	77.2	11.9	96.5	30.0	117.2	35.7
22	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.0
23	0.2	0.1	0.2	0.1	0.3	0.1	0.2	0.1
24	0.2	0.1	0.2	0.0	0.2	0.1	0.2	0.1
25	3.0	0.2	3.0	0.2	2.8	0.3	3.0	0.2
26	0.2	0.1	0.2	0.0	0.2	0.1	0.2	0.1
27	1.2	0.7	0.7	0.2	1.0	0.5	1.6	0.8
28	94.8	1.0	95.2	0.4	95.1	0.7	94.3	1.2
29	0.2	0.1	0.1	0.1	0.2	0.1	0.2	0.1
30	0.3	0.2	0.3	0.2	0.3	0.2	0.3	0.2
31	1215	227	1165	140	1311	338	1207	206

1999/2000 crop

	set B 110 samples		Riviera dei Fiori 26 samples		Riviera del Ponente Savonese 33 samples		Riviera di Levante 51 samples	
V	*m*	*s*	*m*	*s*	*m*	*s*	*m*	*s*
1	0.6	0.3	0.4	0.3	0.6	0.3	0.7	0.3
2	13	4	13	4	15	4	11	4
3	1.75	0.25	1.75	0.23	1.91	0.29	1.63	0.16
4	0.11	0.03	0.10	0.02	0.11	0.03	0.12	0.03
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	12.6	0.9	12.0	0.9	12.6	0.9	12.9	0.8
7	1.0	0.2	0.8	0.1	1.1	0.2	0.9	0.1
8	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0
9	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0
10	1.9	0.2	2.0	0.2	2.0	0.2	1.9	0.2
11	76.0	1.8	76.9	1.6	75.4	2.3	75.9	1.3
12	7.0	1.0	6.9	0.7	7.5	1.3	6.7	0.6
13	0.7	0.1	0.7	0.1	0.7	0.1	0.7	0.1
14	0.3	0.0	0.3	0.0	0.3	0.0	0.4	0.0
15	0.3	0.0	0.3	0.0	0.3	0.0	0.3	0.0
16	0.5	1.7	0.5	1.7	0.8	2.2	0.4	1.3
17	3.7	4.5	5.9	7.3	2.9	3.4	3.1	2.7
18	9.9	18.6	6.3	16.3	10.0	13.8	11.7	22.1
19	107.7	57.7	118.6	47.9	97.6	60.3	108.6	60.5
20	82.4	44.8	66.1	37.0	61.9	32.5	104.0	46.0
21	58.2	16.7	49.4	11.8	57.3	14.5	63.2	18.4
22	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0
23	0.2	0.1	0.2	0.0	0.2	0.1	0.2	0.1
24	0.2	0.0	0.2	0.0	0.2	0.1	0.2	0.0
25	2.9	0.2	2.9	0.3	2.8	0.3	2.9	0.2
26	0.1	0.0	0.2	0.0	0.1	0.0	0.1	0.0
27	1.6	0.8	1.2	0.6	1.4	0.7	1.9	0.7
28	94.3	0.9	94.7	0.9	94.5	1.0	94.0	0.8
29	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1
30	0.5	0.2	0.4	0.2	0.5	0.2	0.4	0.1
31	1350	231	1255	146	1438	324	1342	168

[a] *V* = variable. [b] *m* = mean. [c] *s* = standard deviation.

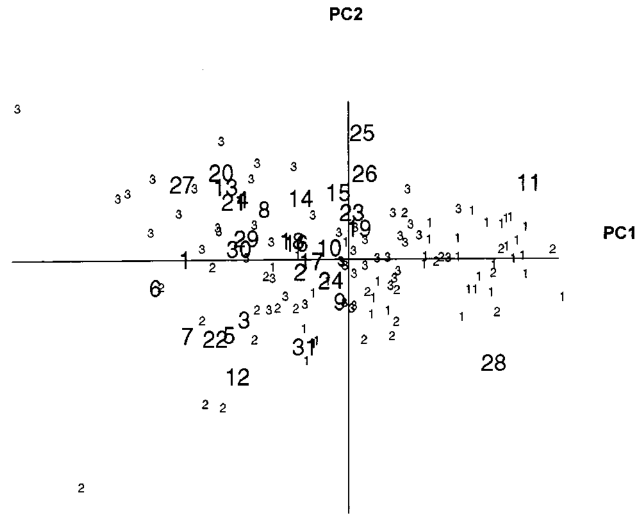

Figure 2. Biplot of the first two principal components (Set A).

Table 3. Feature Selection by SLDA

	set						
V[a]	A1	A2	A3	Anov	Adec	Ajan	B1
1	•		•			•	•
2	•		•			•	•
3	•		•				•
4	•		•	•	•	•	•
5	•	•			•		•
6	•	•		•	•		•
7	•	•		•	•	•	•
8					•		
9	•	•		•	•		
10				•			•
11				•	•		•
12	•	•			•		
13				•			•
14	•	•		•		•	
15					•	•	•
16							•
17				•	•		•
18							
19	•	•					
20				•	•		
21	•	•		•		•	•
22					•		•
23	•	•					•
24				•		•	
25	•	•		•	•	•	•
26	•	•			•		•
27	•	•					
28							
29							•
30	•	•		•		•	•
31	•	•					•

[a] *V* = variable.

used, dividing each data set into a training set (used to calculate classification rules and class-models) and an evaluation set (used to evaluate the prediction ability of rules and models).

The prediction rate of each model is reported with its confidence interval computed using binomial distribution (*11*).

Multivariate data evaluation was carried out by QPARVUS (*12*).

RESULTS AND DISCUSSION

The mean and the standard deviation of the 31 measured variables for each crop and location are reported in **Table 2**. The univariate analysis of this table does not show any relevant

Table 4. LDA Results for 1998/99 Crop

class name	classification rate (%) of set A1	Anov	Adec	Ajan	prediction rate (%) of set A1	Anov	Adec	Ajan
Riviera dei Fiori	97.6	100.0	100.0	100.0	94.6	100.0	100.0	75.0
Riviera del Ponente Savonese	78.5	100.0	100.0	100.0	65.2	100.0	100.0	88.9
Riviera di Levante	94.0	100.0	100.0	100.0	91.5	100.0	91.7	83.3
					80.4<86.9<91.9 [a]	92.1≤100.0 [a]	87.2<97.1<99.8 [a]	64.6<82.6<93.8 [a]

[a] Mean prediction rate (%) with its 90% confidence interval.

Table 5. UNEQ Results for 1998/99 Crop

class name	classification rate (%) of set A1	A2	A3	prediction rate (%) of set A1	A2	A3
Riviera dei Fiori	100.0	100.0	81.1	64.9	70.3	75.7
Riviera del Ponente Savonese	95.7	87.0	56.5	30.4	47.8	39.1
Riviera di Levante	97.9	91.5	63.8	95.7	83.0	55.3
				63.0<71.0<78.1 [a]	63.0<71.0<78.1 [a]	50.5<58.9<66.9 [a]

	sensitivity (%)		specificity [b] (%)	sensitivity (%)		specificity [b] (%)	sensitivity (%)		specificity [b] (%)
Riviera dei Fiori	83.8	vs 2	87.0	86.5	vs 2	78.3	89.2	vs 2	26.1
		vs 3	74.5		vs 3	68.1		vs 3	42.6
Riviera del Ponente Savonese	87.0	vs 1	2.7	87.0	vs 1	0.0	95.7	vs 1	0.0
		vs 3	14.9		vs 3	17.0		vs 3	10.6
Riviera di Levante	85.1	vs 1	35.1	83.0	vs 1	16.2	91.5	vs 1	2.7
		vs 2	73.9		vs 2	65.2		vs 2	8.7

[a] Mean prediction rate (%) with its 90% confidence interval. [b] Specificity is reported as a value for each class versus the other classes.

differences among the three geographical origins of the studied olive oils, therefore a multivariate study is performed.

The PCA results for set A are shown in **Figure 2**. This figure represents the biplot (projection of the object scores and of the variable loadings) on the first two principal components, which explain about 35% of the data matrix information (the first 10 components concentrate 80% of the total variance).

The analysis of the loading plot emphasizes a partial separation between the objects belonging to the category Riviera dei Fiori with respect to the other objects on the first principal component, which is mainly built by palmitic acid, oleic acid, stigmasterol, and acidity. Riviera dei Fiori oils are characterized by a relatively high oleic acid content and relatively low levels of palmitic acid, stigmasterol, and acidity. The other two categories seem to be more separated on the second principal component, which is mainly composed of linoleic acid and campesterol.

To classify and build the models of the classes, class-modeling methods were used. The predictive ability of each model was computed keeping in mind that the prediction rate must be larger than the random assignments, whose value corresponds to 41% in the case of three classes and 107 predicted objects.

LDA was applied to a reduced data matrix (18 variables), obtained using SLDA (set A1 in **Table 3**). The classification results are shown in **Table 4** (set A1), and the class-separation is presented visually in **Figure 3**, which shows the scores of each object on the plane of the two canonical variables. The leave-one-out validation method was used to test the prediction classification ability.

By analyzing the results, it is possible to detect an objective difference among the three different geographical areas mentioned in the PDO.

Samples belonging to categories Riviera dei Fiori and Riviera di Levante are always well predicted (>90%). On the contrary,

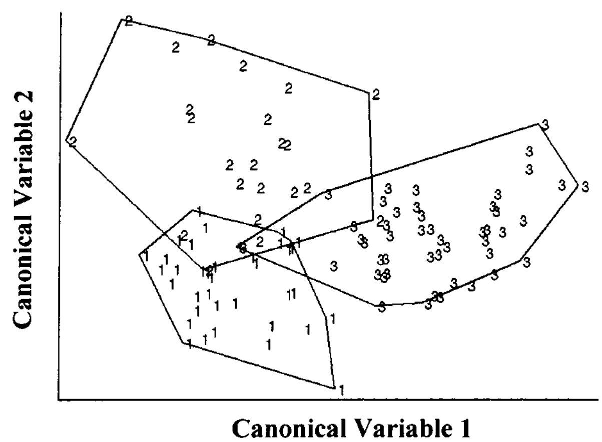

Figure 3. Plot of the first versus the second canonical variable (Set A1).

the prediction ability for Riviera del Ponente Savonese is less satisfactory, probably because of the central geographical position of this area and the heterogeneity of the grown olive cultivars which strongly affects the fatty acid composition of the oils.

To estimate the influence of the harvest time on the geographical classification of the oils, the samples were divided into three further groups: November (36 oil samples), December (35 samples), and January (23 samples) according to the harvesting period. It is important to underline that oils not produced during these months were not taken into account at this stage.

The same statistical analysis was repeated on each group and LDA was applied on reduced data matrixes (**Table 3**) obtained using SLDA: set Anov (14 variables selected); set Adec (14 variables selected); and set Ajan (10 variables selected). The results are reported in **Table 4**. The geographical classification ability seems to be independent of the harvesting period, which

Table 6. SIMCA Results (original data: set A) for 1998/99 Crop

class name	classification rate (%)	prediction rate (%)
Riviera dei Fiori	94.6	62.2
Riviera del Ponente Savonese	95.7	60.9
Riviera di Levante	91.5	87.2
		65.0<73.0<79.8[a]

	sensitivity (%)	specificity (%)[b]	
Riviera dei Fiori	100.0	vs 2	100.0
		vs 3	100.0
Riviera del Ponente Savonese	86.4	vs 1	75.7
		vs 3	87.2
Riviera di Levante	88.4	vs 1	67.6
		vs 2	91.3

[a] Mean prediction rate (%) with its 90% confidence interval. [b] Specificity is reported as a value for each class versus the other classes.

is usually delayed in the western areas of Liguria with respect to the eastern part. This fact confirms that the objective difference among the oils is primarily due to agronomic causes. The prediction results are really excellent for sets Anov and Adec, but are less satisfactory for set Ajan. The lower predictive ability for January samples may be related to the leveling effect of over-ripening on the peculiar features of the oils produced in Liguria.

In addition to these classification results, the class-modeling technique UNEQ was applied to the data set A1. Moreover, to verify the characterizing importance of the first 4 variables reported in **Table 1**, which are indicated in the PDO Regulation (*2*), we decided to use UNEQ on two other data sets, A2 and A3 (**Table 3**). In the A2 data set, each object is described by a vector of 14 variables, the 18 variables selected by SLDA minus the 4 PDO variables, whereas in the A3 data set each object is described by the 4 PDO variables.

The classification results are shown in **Table 5**; by taking into account the prediction ability, it is possible to point out that the class models obtained using set A2 globally yield the best results for all classes. Moreover, the best model for Riviera di Levante was obtained by using all 18 features selected by SLDA; on the contrary, the best model for Riviera dei Fiori was built by using only the 4 PDO variables. For the 3 classes and for the 3 sets, the model sensitivities are quite similar, but differences become evident considering the model specificity.

Taking into account the specificity of data set A2, whose global prediction results are the best ones, it is evident that class 1 model is specific with respect to the other two class models (**Table 5**). It is a quite compact model, able to accept only few samples belonging to the other classes. On the contrary, this does not apply to the Riviera del Ponente Savonese class, whose model is rather scattered and not specific; this fact explains the results obtained with LDA.

To verify the model validity, independently of the feature selection used, SIMCA was performed on the whole data set (original data, set A) and the results are reported in **Table 6** and in **Figure 4**. This figure represents the Coomans plot (*12*) for the first and the third class (Riviera dei Fiori versus Riviera di Levante).

SIMCA results obtained using the reduced data sets of **Table 3** (A1, A2, A3) do not particularly improve the prediction ability obtained using the whole data set A.

The application of chemometric methods allowed building of useful models for extra-virgin olive oils from Riviera dei Fiori, Riviera del Ponente Savonese, and Riviera di Levante produced from the 1998−99 olive crop. In that year favorable

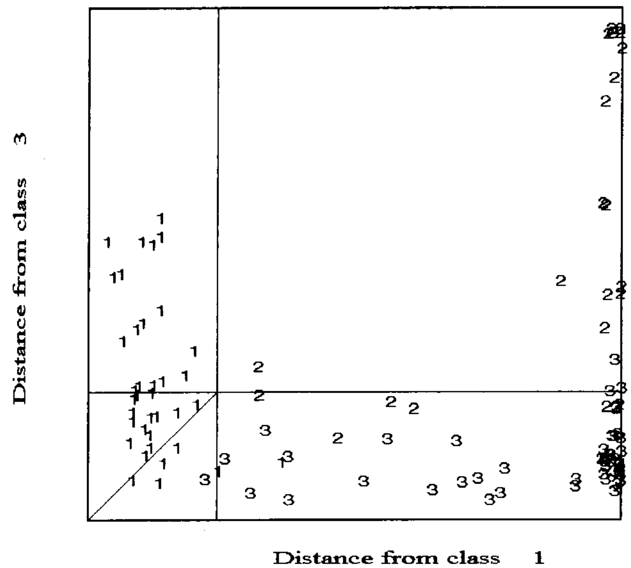

Figure 4. Coomans Plot for the first and the third class (Set A).

Table 7. Class-Modeling of the 1999–2000 Oils Using 1998–99 Mathematical Model

class name	UNEQ[a] of set A1	UNEQ[a] of set A2	SIMCA[a] of set A
Riviera dei Fiori	1/26	3/26	0/26
Riviera del Ponente Savonese	1/33	10/33	13/33
Riviera di Levante	44/51	50/51	40/51

[a] Results are reported as the number of true assignments versus the class object number.

climatic conditions and limited *Dacus oleae* infestation allowed the development of particular features of the oils produced in the three areas mentioned in the PDO Regulation. The good results for Riviera dei Fiori oils are certainly to be ascribed to the cultivar homogeneity, which is reflected by the PDO Regulation (> 90% Taggiasca) for the Imperia area. As far as Riviera di Levante is concerned, the obtained results show that the different olive cultivars grown in Genoa and La Spezia provinces yield oils with homogeneous analytical characters. Finally, oils from Riviera del Ponente Savonese are affected by the irregular presence of different cultivars. In this area, the oils obtained from some cultivars, such as Colombaia and Olivotto, have chemical compositions different from (i.e., lower oleic and higher linoleic acid content) those of Taggiasca, which is still the predominant cultivar. Thus, if the amounts of Colombaia and Olivotto olives approach the highest allowed levels, the fatty acid composition is considerably influenced.

To verify the performances of the mathematical model built with the 1998/99 oil samples, this model was applied to the oils of the following year. Unfortunately, UNEQ (both on set A1 and set A2) and SIMCA results (on the original data) are acceptable only for the Riviera di Levante class, but they are not useful for the prediction of the other two categories (**Table 7**). Therefore, we can underline that, as far as the studied chemical−physical variables are concerned, the oil samples from all western Liguria (Riviera dei Fiori and Riviera del Ponente Savonese) obtained from the 1999/2000 crop are different from similar samples obtained in the previous year.

Thus, we decided to study the new crop independently of the previous one (data set B). **Table 8** summarizes the LDA results obtained after the feature selection of 21 relevant

Table 8. LDA Results for 1999/2000 Crop

class name	classification rate (%) of set B1	prediction rate (%) of set B1
Riviera dei Fiori	82.8	73.1
Riviera del Ponente Savonese	81.9	75.8
Riviera di Levante	92.1	86.3
		72.7<80.0<86.0[a]

[a] Mean prediction rate (%) with its 90% confidence interval.

Table 9. UNEQ Results for 1999/2000 Crop

class name	classification rate (%) of set B1	prediction rate (%) of set B1
Riviera dei Fiori	100.0	15.4
Riviera del Ponente Savonese	100.0	58.1
Riviera di Levante	100.0	92.2
		54.5<62.7<70.4[a]

[a] Mean prediction rate (%) with its 90% confidence interval.

Table 10. SIMCA Results for 1999/2000 Crop

	classification rate (%) of set		prediction rate (%) of set	
class name	B	B1	B	B1
Riviera dei Fiori	65.4	65.4	19.2	23.1
Riviera del Ponente Savonese	90.9	90.9	75.8	69.7
Riviera di Levante	90.2	82.3	74.5	72.6
			53.6<61.8<69.5[a]	51.8<60.6<67.8[a]

[a] Mean prediction rate (%) with its 90% confidence interval.

variables by SLDA (data set B1). Even if LDA prediction results are satisfactory for all classes, the use of class-modeling techniques such as UNEQ (**Table 9**) and SIMCA (**Table 10**) confirms good predictive ability only for Riviera di Levante oils. It is evident that bad climatic conditions and widespread infestation during 1999−2000 leveled out the special characteristics of the oils produced in the three areas. This effect was also increased by the fact that farmers often add fallen olives to already unhealthy olives in order to increase a poor harvest. These facts justify both the inability of the 1998−99 statistical model to single out the possible origin of these oils and the difficulty in building new class models with satisfactory predictive ability for 1999−2000.

ABBREVIATIONS USED

PDO, Protected designation of origin; HPLC−DAD, high-performance liquid chromatography with diode array detector; AOCS, American Oil Chemists' Society; PCA, principal component analysis; LDA, linear discriminant analysis; UNEQ, unequal class modeling; QDA, quadratic discriminant analysis; SLDA, stepwise linear discriminant analysis; SIMCA, soft independent modeling of class analogy; 3,4-DHPEA, 3,4-dihydroxyphenylethanol; *p*-HPEA, *p*-hydroxyphenylethanol; 3,4-DHPEA-EDA, dialdehydic form of elenolic acid bonded with 3,4-dihydroxyphenylethanol.

LITERATURE CITED

(1) EC Regulation 1997. European Community Regulation No. 123/97. *Off. J. Eur. Communities: Legislation* **1997** *40* (1/23/1997).
(2) EC Regulation 1991. European Community Regulation No. 2568/91. *Off. J. Eur. Communities: Legislation* **1991** *34* (9/5/1991).
(3) Evangelisti, F.; Zunin, P.; Tiscornia, E.; Petacchi, R.; Drava, G.; Lanteri S. Stability to oxidation of virgin olive oils as related to olive conditions: study of polar compounds by chemometric methods. *J. Am. Oil. Chem. Soc*. **1997**, *74*, 1017−1021.
(4) *Official Methods of Analysis of AOCS*, 5th ed. AOCS: Champaign, IL, 1997; Method **Ce 8-89** reapproved in 1997.
(5) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.
(6) Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. Supervised Pattern Recognition. In *Handbook of Chemometrics and Qualimetrics: Part B*; Vandeginste, B. G. M., Rutan, S. C., Eds,; Elsevier: Amsterdam, The Netherlands, 1998; pp 207−241.
(7) Derde, M. P.; Massart, D. L. UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution. *Anal. Chim. Acta* **1986**, *184*, 33−51.
(8) Forina, M.; Lanteri, S.; Sarabia, L. Distance and class space in the UNEQ class-modeling technique. *J. Chemometrics* **1995**, *9*, 69−89.
(9) Wold, S.; Sjostrom, M. SIMCA: A method for analysing chemical data in terms of similarity and analogy. In *Chemometrics, Theory and Application;* Kowalski, B. R., Ed; ACS Symposium Series No. 52; American Chemical Society: Washington, DC, 1977; pp 243−282.
(10) Frank, I. E.; Todeschini, R. In *The Data Analysis Handbook*; Vandeginste, B. G. M., Rutan, S. C., Eds.; Elsevier: Amsterdam, The Netherlands; 1994.
(11) Forina, M.; Lanteri, S.; Rosso, S. Confidence intervals of the prediction ability and performance scores of classifications methods. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 121−132.
(12) Forina, M.; Lanteri, S.; Armanino, C. QPARVUS 3.0: An extendable package of programs for data explorative analysis, classification and regression analysis. Department of Chimica e Tecnologie Farmaceutiche e Alimentari, University of Genova: Genova, Italy, 2000. (Free download at http://PARVUS.unige.it).

Received for review October 1, 2001. Revised manuscript received January 18, 2002. Accepted January 22, 2002. This research was carried out for the Regione Liguria project "Miglioramento qualitativo della produzione di olio di oliva ligure", supported by the European Union (UE Reg. no. 528/99).

JF011289M