Biplot Analysis Applied to Enological Parameters in the Geographical Classification of Young Red Wines

J. C. RIVAS-GONZALO[1], Y. GUTIERREZ[2], A. M. POLANCO[3], E. HEBRERO[4], J. L. VICENTE[5], P. GALINDO[6], and C. SANTOS-BUELGA[7]*

Fifteen variables including conventional enological parameters, phenolics, and color-related parameters were analyzed in 45 young red wines belonging to the Spanish zones Ribera de Duero and Toro with Certified Brand of Origin (CBO). The data obtained were subjected to HJ biplot analysis to determine whether, with the variables employed, it would be possible to differentiate the wines according to their CBO. Alcohol, fixed acidity, pH, degree of ionization of anthocyanins, total phenolics, and procyanidins were the variables with the best relative contributions to discrimination between the two CBO's. A logistic model for estimating the probability of a given sample belonging to a certain CBO is proposed. According this prediction model, 100% of the observations for Ribera de Duero and 81.81% for those of Toro are well classified.

KEY WORDS: biplot analysis, geographical classification of wines

Over the years, there has been a constant search to characterize food products in general and wines in particular. There have been different underlying aims, such as avoidance of fraudulent practices and deeper insight into the peculiarities of such products. This characterization became especially important with the advent of Certified Brands of Origin (CBO), whose commercial products must conform to strictly defined characteristics.

For the characterization of wines, it is common to use sensory analysis which, unfortunately, is mostly subjective. However, there is an increasing trend towards the use of objective parameters. In wines, the choice of parameters to be used for such purposes is difficult, since it is necessary to take into account not only the variations in composition typical of the grapes themselves, but also those related to the processes of production, storage and aging. In addition, when dealing with secondary components, the analysis may become further complicated because these are present in low concentrations and can also have complex structures.

With a view to classifying wines according to their CBOs, several authors have used different techniques of statistical analysis. Kwan *et al.* (11) classified Moselle and German Rhine white wines, according to the composition of certain metals and volatile compounds. They found the best results using least squares multilinear regression (LEAST). Tapias *et al.* (27) used enological parameters to classify Spanish red wines belonging to four CBOs. Three supervised methods were used, obtaining coincident classifications results: linear discriminant analysis (LDA), linear learning machine (LLM), and K-nearest neighbor (KNN). The LDA was applied by Santamaría *et al.* (22) to differentiate red and rosé wines from pale wines according the phenolic composition. A discriminant analysis method and an unsupervised method (clustering) was used for the chemometric characterization of red wines from Majorca (Spain) by Mulet *et al.* (14); they employed 20 variables, including conventional enological parameters, aromatic substances, and metals. Callao *et al.* (3) applied four pattern recognition methods of analysis (LDA, LLM, KNN, and LEAST) to characterize red wines produced in three zones of Catalonia (Spain) using volatile constituents and enological parameters. The SIMCA method (soft independent modelling class analogy) was applied by Larrechi *et al.* (13) to 12 enological parameters and seven metal concentrations to characterize white wines from two adjacent Spanish CBOs. Vasconcelos and Chaves das Neves (28) characterized wines by their amino acid contents, using a hierarchical clustering, principal component analysis (PCA), and discriminant analysis (DA). DA and PCA methods were also used to classify varieties of grape on the basis of their anthocyanin composition (12,21).

Less used for characterization has been the biplot methods of multivariate analysis (4,5). Recently, this technique was applied to different parameters to determine the maturity status of apricots (2) and a variant of the same, proposed by Galindo (6), was used by the authors of the present work (23) for the classification of young red wines, according to their anthocyanin profiles.

[1,2,3,4,7]Departamento de Química Analítica, Nutrición y Bromatología. Universidad de Salamanca. Ave. Campo Charro s/n. 37007 Salamanca. Spain; [5,6]Departamento de Estadística y Matemática Aplicadas. Universidad de Salamanca. Ave. Campo Charro s/n. 37007 Salamanca. Spain.

*Author to whom correspondence should be addressed.

This research was conducted at Laboratorio de Nutrición y Bromatología of Universidad de Salamanca, Spain.

The authors thank C.I.C.Y.T. (ref. no. PB86-0201) and Junta de Castilla y León for financial support given and Nick Skinner for translation of this paper.

Manuscript submitted for publication 6 August 1991.

The aim of this work was to classify wines belonging to two Spanish CBOs by applying the HJ-biplot method to a series of variables, among them enological parameters (ethanol, pH, volatile acidity, fixed acidity, and total titratable acidity), phenolic compounds (anthocyanins, procyanidins, total phenolics), and color-related parameters (wine color density, hue, and degree of ionization of anthocyanins). An additional aim was to determine which of these variables was the most important to perform a correct classification. Finally, we calculated the probability of a given sample belonging to a certain CBO, establishing a model of statistical prediction from the variables selected.

Materials and Methods

Samples: The samples used in the work correspond to Spanish young red wines from 1986 and 1987 crops belonging to the CBOs Toro (11 samples) and Ribera de Duero (34 samples). The wines were obtained directly from the cellars belonging to the respective Regulating Councils of both CBOs, instead of being bought from stores, in order to guarantee their suitability and representativeness.

Methods: pH, alcohol content, volatile acidity, total titratable acidity, and fixed acidity were determined using the reference methods of OIV (15). Total phenolics were determined using the Folin-Ciocalteau reagent (18) and by measuring absorbance at 280 nm of the wine at pH = 0, according Somers and Evans (24). As an index of the catechin content, the substances reactive to vanillin were determined (26). The procyanidins were analyzed by conversion to anthocyanins on heating in acid medium (26). The ratio between substances reactive to vanillin and procyanidins was also obtained.

Anthocyanins were determined by HPLC (10) and by the photometric methods of Ribéreau-Gayon and Stonestreet (18) and of Somers and Evans (24). The wine color density was obtained by the sum of the absorbances of the wine, after filtering, at 420 and 520 nm (25) and also by the sum of the absorbances at 420, 520, and 620 nm (8). Wine color hue was considered as the relationship between absorbances at 420 and 520 nm (8). The chemical age of the wines was determined by the relationship between the absorbances at 520 nm of the wine bleached with SO_2 and after addition of acetaldehyde. The degree of ionization was the percentage of anthocyanins observed to be ionized after abolishing the SO_2 effect upon wine color (24).

All analyses were always carried out in triplicate; Table 1 shows the means of three determinations.

Statistical analysis: Statistical analysis was performed using a combination of two techniques: a modification of Biplot methods known as the HJ-biplot (6) and a logistic discriminant analysis. The former was used as the method for choosing the variables before the discriminant analysis; the second was employed in order to obtain a model that would estimate the probability that a given observation would, in fact, correspond to a particular CBO.

Biplot methods permit the joint plotting, in a reduced dimension, of the rows (samples) and columns (variables) of a data matrix. Gabriel (4) described two types of biplot representations; these are known as the GH'-biplot and the JK-biplot, each of them with different properties. In the first, one obtains a high quality of representation for the variables, but the method is poor for the observations. In the second, one obtains a high quality of representation for the observations, but the variables are poorly represented. Thus, in the strict sense, one cannot speak in terms of a simultaneous representation.

Galindo (6) proposed the technique known as the HJ-biplot, which achieves an optimum quality of representation both for the rows and the columns; additionally, it plots both of them on the same reference system. This representation is intimately related to principal component analysis in the sense that the matrices of variance and covariance are plotted on the two planes that absorb the greatest part of the variability. With corresponding factorial analysis, the interpretation of the results in both techniques is similar.

The distance that separates two row points (observations) on the plot is interpreted in terms of similarity; the angle formed by the vectors, representing the parameters, is interpreted in terms of correlation and the proximity between row points and columns, in terms of preponderance (since each observation is situated at the baricenter of the values taken by each of the parameters). If the data are not standardized, the length of the vectors is interpreted in terms of variability.

Together with the classic interpretation of biplots, Galindo and Cuadras (7) defined a set of measures that are essential for a correct interpretation, relative contributions, and qualities of representation for each element (row or column).

Results and Discussion

The first purpose of this paper was the search for a model to predict the CBO of a particular wine. With the kind of data employed (a categorical response and several continuous predictors), a discriminant technique should be used. From preliminary inspection of the data, it was clear that dispersion was not the same in both CBOs, so the use of classical linear discriminant analysis (LDA) was not warranted. Logistic discriminant (LD) was more appropriate because it is more robust than LDA when the previous assumptions for the application of this are not true (17). On the other hand, LD provides the expected probability that a wine belongs to a CBO, and the model is easier to interpret. Cluster methods were not suitable in this case, and SIMCA is only a modification of LDA not usually available in standard statistical packages.

The second purpose, and perhaps the most important, was structural: to select the most important variables to classify the wines and remove the superfluous ones. The most common approach to this problem is to use stepwise methods in order to establish a suitable

Table 1. Results obtained for the enological variables from Ribera de Duero and Toro wine samples.

Sample	A (%)	VA g acetic acid/L	TA g tartaric acid/L	FA g tartaric acid/L	pH	TPR g gallic acid/L	TPS	V mg catechin/L	PC mg cyanidin/L	ACR Total anthocyanins mg/L	ACS Total anthocyanins mg/L	ACC malvidin-3-glucoside mg/L	CI	CI'	H	I (%)	CA	V/PC
1-RD86	12.8	1.20	6.7	5.2	3.7	2827	50.8	811	3794	386	287	181	7.81	8.95	0.720	18.4	0.489	0.21
2-RD86	12.8	0.75	6.9	6.0	3.5	1818	37.8	968	1736	144	141	69	4.88	5.55	0.755	23.6	0.480	0.56
3-RD86	12.5	1.00	7.2	6.0	3.6	1459	35.1	866	2306	225	132	78	5.52	6.35	0.456	36.8	0.598	0.38
4-RD86	11.9	0.70	7.7	6.8	3.3	2054	32.1	978	3420	204	110	84	4.64	5.15	0.675	36.4	0.420	0.29
5-RD86	12.5	0.95	7.7	6.3	3.6	2930	49.6	1128	3158	214	148	75	6.99	7.87	0.672	34.2	0.450	0.36
6-RD86	12.1	0.50	5.8	5.2	3.2	1906	30.6	875	2931	167	95	74	3.98	4.36	0.716	38.1	0.434	0.30
7-RD86	12.2	0.80	5.9	4.9	3.4	2071	35.6	754	3101	252	160	101	7.60	8.84	0.716	28.5	0.501	0.24
8-RD86	12.6	0.40	5.4	4.9	3.3	1996	30.6	1202	3001	315	124	101	6.15	7.11	0.740	27.7	0.566	0.40
9-RD86	13.0	0.40	4.6	4.1	3.6	2687	41.7	1354	4785	293	170	137	6.60	7.85	0.930	21.6	0.557	0.28
10-RD86	12.4	0.35	5.5	5.0	3.3	2468	30.0	739	2800	152	67	56	5.49	6.23	0.750	30.3	0.689	0.26
11-RD86	12.6	0.35	5.6	5.2	3.3	1965	30.4	839	2878	144	78	61	5.25	5.96	0.756	33.6	0.584	0.29
12-RD86	12.2	0.40	4.8	4.4	3.5	1731	28.4	770	2649	193	107	75	5.86	6.65	0.770	25.6	0.614	0.29
13-RD86	12.8	0.40	5.2	4.7	3.5	1827	29.0	960	2746	179	101	65	5.77	6.65	0.748	29.2	0.607	0.35
14-RD86	13.0	0.60	5.3	4.5	3.6	2110	38.9	489	2587	358	238	172	6.29	6.97	0.682	22.8	0.411	0.19
15-RD87	12.3	0.40	4.5	4.1	3.7	925	20.8	615	1792	231	177	168	1.95	2.11	0.893	22.4	0.227	0.34
16-RD87	11.3	0.30	4.7	4.3	3.7	1637	30.0	881	2386	254	189	154	3.68	4.19	0.859	21.3	0.274	0.37
17-RD87	10.8	0.30	4.8	4.4	3.7	1103	21.9	658	1906	158	120	89	2.20	2.44	0.880	20.7	0.319	0.35
18-RD87	11.2	0.40	4.4	3.9	3.7	1528	28.2	735	2188	275	203	146	4.05	4.57	0.674	17.8	0.382	0.34
19-RD87	11.9	0.35	4.3	3.9	3.6	1468	24.8	666	1889	222	169	132	3.31	3.74	0.706	16.9	0.376	0.35
20-RD87	11.5	0.40	4.4	3.9	3.6	1462	25.7	612	1887	236	178	130	3.98	4.46	0.665	19.2	0.371	0.33
21-RD87	11.7	0.25	5.7	5.4	3.7	1887	33.7	830	2634	290	223	186	3.17	3.53	0.695	22.8	0.257	0.32
22-RD87	11.6	0.65	4.5	3.7	3.8	2071	32.1	906	2598	242	187	146	2.63	2.99	0.742	20.6	0.280	0.35
23-RD87	12.4	0.40	4.8	4.3	3.7	2074	35.9	770	2702	220	167	101	5.34	6.01	0.633	26.1	0.435	0.29
24-RD87	11.4	0.55	4.6	3.9	3.7	1481	27.5	726	2046	273	186	154	2.09	2.32	0.685	19.7	0.247	0.35
25-RD87	12.0	0.50	4.8	4.2	3.6	1359	22.0	543	1771	184	111	87	3.46	3.96	0.640	24.7	0.383	0.31
26-RD87	12.7	0.35	4.3	3.9	3.6	1981	28.6	853	2506	270	212	156	3.09	3.42	0.717	19.5	0.274	0.34
27-RD87	11.9	0.55	4.5	3.8	3.6	1890	29.5	748	2570	309	230	179	2.80	3.04	0.640	22.2	0.265	0.29
28-RD87	11.9	0.55	4.6	3.9	3.6	1706	32.2	659	2282	248	189	149	2.12	2.23	0.631	21.2	0.222	0.29
29-RD87	12.7	0.80	5.6	4.6	3.8	1503	24.2	637	2036	384	200	159	3.83	4.31	0.680	18.0	0.404	0.31
30-RD87	12.2	0.40	4.0	3.5	3.8	1778	25.2	670	2871	263	189	139	4.20	4.81	0.687	17.1	0.470	0.23
31-RD87	12.4	0.60	5.2	4.5	3.5	2000	29.2	756	2298	289	199	160	5.13	5.74	0.608	24.9	0.386	0.33
32-RD87	12.7	0.50	5.1	4.5	3.6	1444	23.9	606	2086	232	166	115	4.45	4.93	0.606	24.1	0.410	0.29
33-RD87	12.6	0.35	5.1	4.7	3.6	1706	27.8	652	2509	300	205	163	4.23	4.62	0.584	22.8	0.380	0.26
34-RD87	11.2	0.35	4.8	4.3	3.6	1515	20.0	528	1910	258	168	136	3.98	4.34	0.555	22.1	0.429	0.28
35-T86	13.0	1.60	6.8	4.8	3.8	2153	38.4	716	2547	166	122	44	5.34	6.06	0.774	14.7	0.670	0.28
36-T86	14.0	0.40	4.4	3.9	3.4	2481	44.3	1313	3392	288	197	146	5.76	6.64	0.930	16.7	0.559	0.39
37-T86	13.2	0.55	4.7	4.0	3.6	1665	38.0	834	2956	206	204	82	5.15	5.85	0.790	14.2	0.554	0.28
38-T86	13.4	1.10	4.9	3.5	3.6	3152	56.2	1561	5082	244	174	90	7.91	9.12	0.746	21.2	0.629	0.31
39-T86	13.2	0.65	4.9	4.1	3.6	3109	54.2	1355	4536	280	189	115	7.88	9.30	0.828	16.1	0.624	0.30
40-T86	13.9	0.80	4.6	3.6	3.7	1975	33.9	544	2390	254	182	113	4.42	5.02	0.819	18.6	0.444	0.23
41-T87	14.0	0.60	4.1	3.4	3.4	2334	47.3	1210	3422	354	253	183	6.58	7.74	0.790	17.2	0.501	0.35
42-T87	13.9	0.30	4.7	4.3	3.6	1915	39.8	1128	2970	229	188	89	5.00	5.67	0.720	18.3	0.445	0.38
43-T87	12.1	0.70	5.5	4.6	3.4	2171	35.9	1177	3313	280	202	130	6.04	6.81	0.659	26.4	0.397	0.36
44-T87	13.9	0.40	3.8	3.3	3.9	2668	47.1	1270	2922	345	268	148	7.07	8.37	0.861	12.9	0.525	0.43
45-T87	12.3	0.65	4.8	4.0	3.6	3071	51.6	1496	4600	388	267	180	7.01	8.26	0.854	23.0	0.432	0.33

Abbreviations: A = alcohol; VA = Volatile acidity; TA = Total titratable acidity; FA = Fixed acidity; TPR = Total phenolics (Folin-Ciocalteau reagent); TPS = Total phenolics (by Somers and Evans method); V = Substances reactive to vanillin; PC = Procyanidins; ACR = Anthocyanins (by Ribéreau-Gayon and Stonestreet method); ACS = Anthocyanins (by Somers and Evans method); ACC = Anthocyanins (by HPLC method); CI = Color intensity (as Sudraud, 1958); CI' = Color intensity (as Glories, 1984), H = Hue; CA = Chemical age; I = Degree of ionization of anthocyanins.

subset of variables. However, stepwise methods do not usually achieve the optimal subset. Moreover, if the predictors are closely related, certain important variables may be excluded from the final model and their classification power can not be assessed. The HJ-biplot method was used as a selection technique. The HJ-biplot has essentially the same interpretation as the classical biplot methods (4) and combines the principal advantages of biplots and correspondence analysis (9).

HJ-biplot analysis: Table 1 was used as INPUT for the statistical analysis. This is a matrix of 45 rows (samples) by 18 columns (variables). The origin of the wines is identified by the CBO notation as RD (Ribera de Duero) or T (Toro) and the year of harvest as 86 (1986) or 87 (1987).

The HJ-biplot representation, on the plane of maximum inertia, is shown in Figure 1. To facilitate the reading of the graph, the cloud of data points of the rows is offered separately on an expanded scale.

The weight of inertia on the plane surpasses 58%, the first eigenvalue is 15.42 and the second 10.98. Accordingly, both are well differentiated, and it is possible to interpret both axes separately. The projections onto the plane of maximum inertia shows that the first axis discriminates the samples essentially by the year of harvest. However, the most important variables for differentiating the samples as a function of the CBO are the characteristics of axis 2. The greatest relative contributions of the first factorial axis are attained for the variables shown in Table 2. These variables are the most important for the differentiation of the samples as a function of the year of harvest, especially in the Ribera de Duero wines, since the number of observations for the Toro CBO is small and hence less reliable.

It was expected that the anthocyanins would constitute the variable that most contributes to the differentiation of samples according to the year of harvest. In red wines, these compounds undergo considerable modifications over time, since they are involved in reactions of degradation and polymerization. Accordingly, the contribution is greater for the anthocyanins determined by HPLC, since with this technique only the free anthocyanins are determined, whereas by spectrophotometric methods (18,24) part of the polymerized anthocyanins are also quantified. (1,16). Additionally, the anthocyanins and their degrees of polymerization are closely related to the wine color density and the chemical age of the wines, such that these variables also participate in the differentiation of the samples according to the year of harvest.

Table 3 shows the variables with the greatest relative contribution to axis 2, which are the most important for differentiating the samples by their geographic origin. The changes undergone by these variables over time are less pronounced, but not insignificant; thus, it might be possible to achieve better differentiation if samples from only one year were being studied.

Among the parameters with the greatest relative contribution to axis 2, fixed acidity and total titratable acidity play an important role. Fixed acidity is related to the organic acids present in the grapes, which are a function of the variety, the state of maturity, and the climatic conditions. In the wines of both CBOs, the principal variety is Tempranillo; however, there are differences in climate and in the state of maturity at which the grapes are harvested. In the Toro region, the aim is to obtain wines with a high alcohol content (above 12.5°), so the grapes are harvested when their sugar content is maximized, and

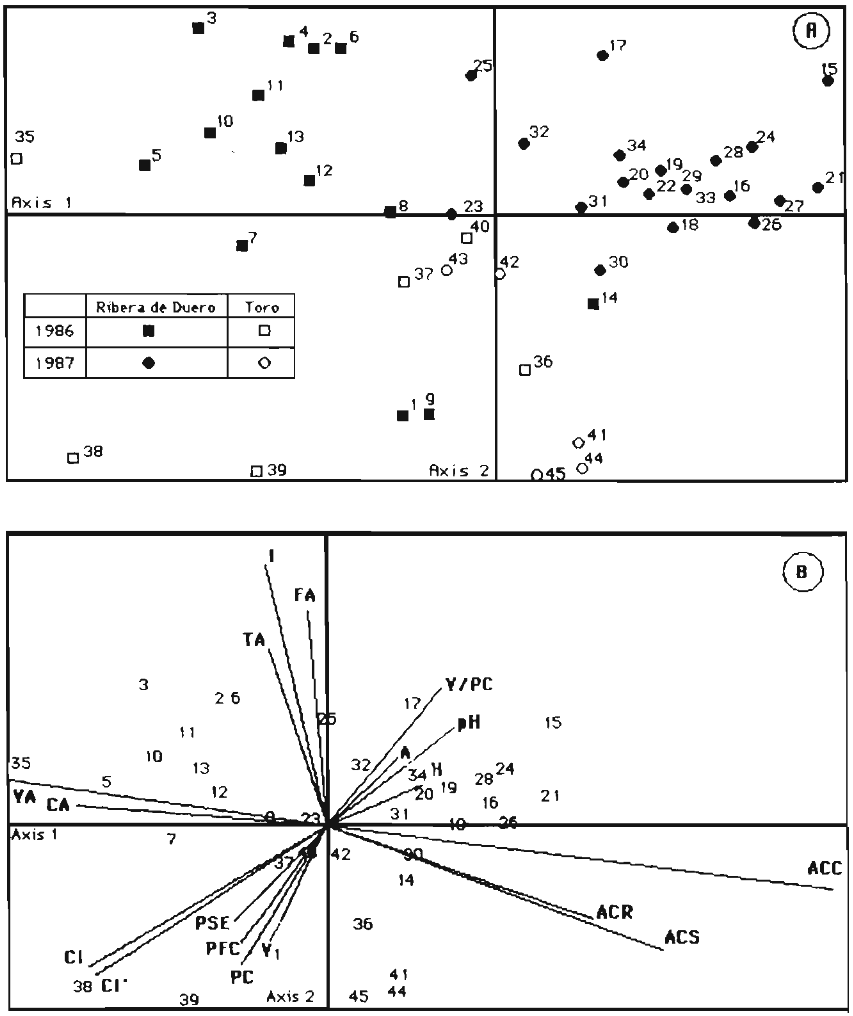

Fig. 1. Projection of samples (A) and variables (B) onto the maximum plane of inertia.

Table 2. Variables with the greatest contribution to axis 1.

Variable	Relative contribution
Anthocyanins (determined by HPLC)	900
Anthocyanins (by Somers and Evans method)	670
Anthocyanins (by Ribéreau-Gayon and Stronestreet method)	610
Chemical age	490
Wine color density (as Sudraud)	470
Wine color density (as Glories)	460

therefore, the amount of certain fixed acids (*e.g.*, malic acid) has been reduced. The total titratable acidity is influenced by volatile acidity, which is dependent upon the alcohol content and the treatment of the wines.

The high relative contribution to axis 2 by the total phenolics and the procyanidins can be explained in two ways. On one hand, Tempranillo grapes grown in the Toro region are richer in phenolic compounds than those cultivated in the Ribera de Duero region, a result of their different adaptation to the environment. On the other, the traditional winemaking practices used in the Toro CBO cellars tend to favor the presence of high phenolic contents in the wine, for which long maceration processes are used.

Figure 1b shows that substances reactive to vanillin (V), alcohol (A), pH, wine hue color (H), and the V/PC ratio do not have a significant contribution to axis 2. However, they do seem to be included in the direction

Table 3. Variables with the greatest relative contribution to axis 2.

Variable	Relative contribution
Fixed acidity	840
Total titratable acidity	780
Degree of ionization of anthocyanins	640
Total phenolics (determined by Folin-Ciocalteu reagent)	480
Total phenolics (by Somers and Evans method)	400
Procyanidins	390

that defines the separation between CBOs. Of these, alcohol and pH have been included as being contributors for the separation since they have a significant representation on the first principal plane (47% and 69%, respectively). pH is a parameter closely related to acidity, while high alcohol is demanded by the regulations of the Toro CBO. The remaining variables mentioned have been excluded since their position on the plane is not significant.

The total phenolics determined with the Folin-Ciocalteau reagent or with the Somers and Evans method measure the same parameter by different procedures. Since they correlate highly, inclusion of both results may lead to problems in a regression model. Accordingly, the measurements performed with the Folin-Ciocalteau reagent were used, since they made a larger contribution. The Somers and Evans index was discarded.

It should be noted that some of the variables chosen have already been considered by other authors as important for the characterization and classification of wines: ethanol, total titratable acidity (3,27), pH (3,14), and total phenolics (14,22).

Using only the variables chosen, the HJ-biplot analysis was repeated to check the stability of the separation obtained. Figure 2 shows the simultaneous plot of the samples and the variables on the maximum plane of inertia for this second analysis. In the figure, the ellipses of confidence for the samples of each of the CBOs have been represented. As may be seen, it is the first principal axis that separates both CBOs, and the quality of representation is much better (98.4%). It is also possible to obtain important information about the behavior of the variables; for example, the alcohol content and fixed acidity are variables that are closely related, as shown by the fact that the angle formed by the vectors representing them is very small. The same is the case with pH and the procyanidins. The first axis absorbs the greatest part of the information with an inertia weight of 85% and the percentage of variance absorbed by axis 1 of 65%.

Logistic discriminant: With the variables selected in the previous step, a model of logistic regression was fitted, where the dependent variable was defined as follows: Y = 1, if the sample belonged to the Ribera de Duero CBO; Y = 0, if the samples belonged to the Toro CBO.

The model would be of the type:

$$P(Y = 1) = 1/[1+ \exp (b_0 + b_{1x1} + b_{2x2} + ...)].$$

That is, it estimates the probability of a given sample belonging to the Ribera de Duero CBO. The probability

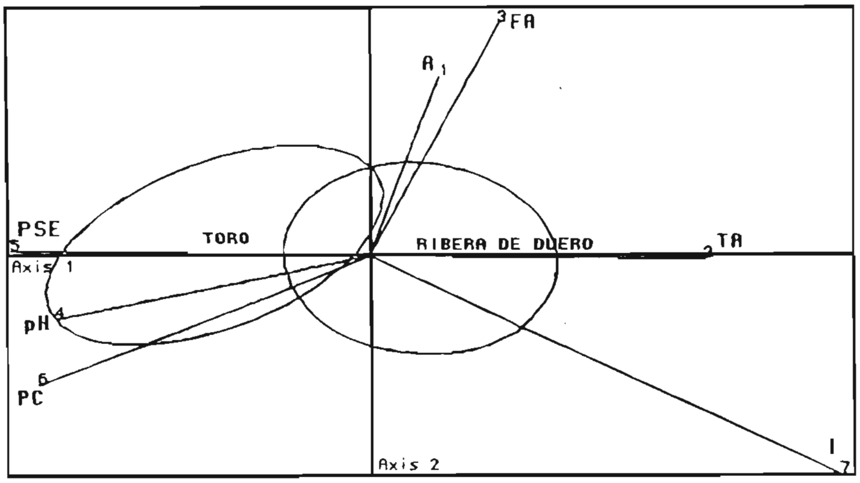

Fig. 2. Simultaneous plot of the samples and variables selected onto the maximum plane of inertia showing the ellipses of confidence for each CBO.

Table 4. Parameters of the logistic regression model.

Variable	Coefficient	Std Error	t-value
Intercept	- 42.2966	35.9762	- 1.1757
X_1 Alcohol content	- 0.7675	0.8825	- 0.8696
X_2 Total titratable acidity	1.4063	1.1766	1.1952
X_3 Fixed acidity	12.4131	8.0319	1.5455
X_4 pH	- 0.0001	0.0034	- 0.0152
X_5 Total phenolics (Somers and Evans method)	- 0.1956	0.1662	- 1.1769
X_6 Procyanidins	0.0008	0.0014	0.5428
X_7 Degree of ionozation of anthocyanins	0.3200	0.2463	1.2989

of it belonging to the Toro CBO would be P(Y= 0) = 1 - (P(Y = 1). Table 4 shows the coefficients estimated with their corresponding standard errors.

The prediction model serves two purposes: to estimate the above-described probabilities and to classify the observation in one of the two groups. With this aim, the level of error or proportion of poorly classified samples was evaluated according to the available data. It should be noted that the proportion will be higher when samples extracted from a new group of wines are used than when those of the group used to calculate the model are used. The jackknife method was chosen as a procedure of cross-validation. This consists of eliminating a sample, re-estimating the model, and using the new model to classify the sample withdrawn. This process was repeated with all the samples and the end the result obtained was evaluated.

To evaluate the classification, the following procedure was used: a cut-off point was chosen in such a way that if the estimated probability P(Y = 1) was greater than this cut-off point, the sample was classified as Ribera de Duero and, if smaller, as Toro. Figure 3 shows the percentages of samples that were well classified, for both CBOs and for the whole set, considering different cut-off points.

In order to correctly classify the observations from Ribera de Duero, it was necessary to take a cut off point of 0.47, whereas to correctly classify all the samples of the Toro wine, the cut-off point should be set above 0.75. With no additional information, for a cut-off level of between 0.35 and 0.47, 100% of the observations for Ribera de Duero and 81.81% for those of the Toro CBO were well classified (95.56% for the overall set), such that a suitable cut-off point would be approximately 0.40.

In summary, to classify an unknown sample, it would be necessary to perform an analysis of the variables chosen in the first step (alcohol content, fixed acidity, pH, total phenolics, procyanidins, and degree of ionization). These results must then be substituted in the prediction model to estimate the P(Y = 1) probability, classifying it as Ribera de Duero if this probability proves to be greater than 0.40 (with an approximate reliability of 100%) and as Toro in the opposite case (with a reliability of 81.82%).

Conclusions

Use of the HJ-biplot method has allowed us to select the most important variables for the differentiation of wines from two regions of Spain with certified brands of origin and to obtain a logistic model which, from the variables chosen, permit one to classify these wines as a function of their CBO. The suitability of the statistical method employed is apparent in the fact that the wines had quite similar characteristics, they were made with the same variety of grapes, and were produced in adjacent geographic zones. It is, therefore, to be expected that the method could be successfully applied for the classification of wines produced in other regions as long as suitable variables are included in the study to perform the differentiation.

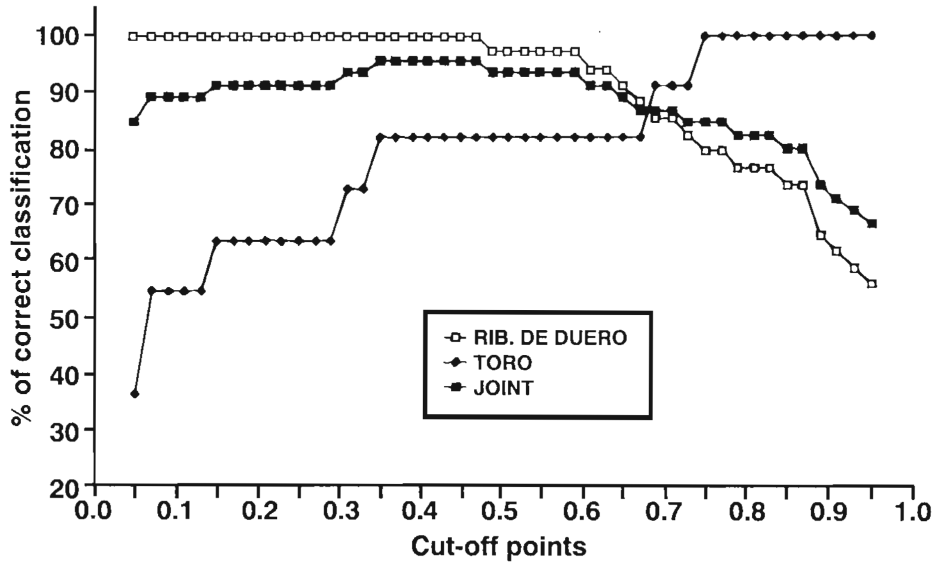

Fig. 3. Plot of percentages of correct classification of samples according to different cut-off points on applying the proposed logistic prediction model.

Literature Cited

1. Bakker, J., N. W. Preston, and C. F. Timberlake. The determination of anthocyanins in aging red wines: comparison of HPLC and spectral methods. Am. J. Enol. Vitic. 37: 121-6 (1986).

2. Brown, G. S., and T. D. Walker. Indicators of maturity in apricots using biplot multivariate analysis. J. Sci. Food Agric. 53:321-31 (1990).

3. Callao, M. P., J. Guasch, M. S. Larrechi, and F. X. Rius. Aroma components and chemometric analysis to characterize catalan wine producing zones. *In:* Proceedings of II World Congress of Food Technology. E. Primo and P. Fito (Eds.). Univ. Politécnica, Valencia, Spain (1987).

4. Gabriel, K. R. The biplot graphic display of matrices with application to principal component analysis. Biometrika, 58: 453-67 (1971).

5. Gabriel, K. R. Biplot. *In:* Encyclopædia of Statistical Sciences. S. Kotz and N. L. Johnson (Eds.). John Wiley and Sons, New York (1981).

6. Galindo, P. Una alternativa de representación simultánea: HJ-biplot. Questió, 10:13-23 (1986).

7. Galindo, P., and C. M. Cuadras. Una extensión

del método biplot y su relacin con otras técnicas. 91 pp. Publicaciones de Bioestadística y Biomatemática de la Universidad de Barcelona, Barcelona, Spain (1986).

8. Glories, Y. La couleur des vins rouges. II. Mesure, origine et interpretation. Connaiss. Vigne Vin, 18:253-71 (1984).

9. Greenacre, M. J. Theory and Application of Correspondence Analysis. Academic Press, London (1984).

10. Hebrero, E., C. Santos-Buelga, and J. C. Rivas-Gonzalo. High Performance Liquid Chromatography-Diode Array Spectroscopy identification of anthocyanins of *Vitis vinifera* variety Tempranillo. Am. J. Enol. Vitic. 39:227-33 (1988).

11. Kwan, W. O., and B. R. Kowalski. Classification of wines by applying pattern recognition to chemical composition data. J. Food Sci. 43:1320-3 (1978).

12. Larice, J. L., P. Archier, C. Rocheville-Divorne, S. Coen, and J. P. Roggero. Composition anthocyanique des cépages. II. Essai de classification sur trois ans par analyse en composantes principales et étude des variations annuelles de cépages de même provenance. Rev. Française d'Oenol. 121:7-12 (1989).

13. Larrechi, M. S., P. Puig, and F. X. Rius. Objective grounds for the existence of the viticultural D. O. Conca de Barbera. *In:* Proceedings of II World Congress of Food Technology. E. Primo and P. Fito (Eds.). Univ. Politécnica, Valencia, Spain (1987).

14. Mulet, A., A. Berna, M. Forcen, and J. Mora. Vitivinicultural Producing Zones in Majorca. *In:* Proceedings of II World Congress of Food Technology. E. Primo and P. Fito (Eds.). Univ. Politécnica, Valencia, Spain (1987).

15. Office International de la Vigne et du Vin. Recueil des méthodes internationales d'analyse des vins et des moûts. 368 pp. O.I.V., Paris (1990).

16. Peri, C., and C. Pompei. An estimate of different phenolic fractions in wines. Am J. Enol. Vitic. 22:59-64 (1971).

17. Press, S. J., and S. Wilson. Choosing between logistic regression and discriminant analysis. J. Am. Statist. Assoc. 73 (364):699-705 (1978).

18. Ribíreau-Gayon, J., E. Peynaud, P. Sudraud, and P. Ribéreau-Gayon. Sciences et techniques du vin. Tome I. Analyse et contrôle des vins. 671 pp. Dunod, Paris (1972).

19. Ribéreau-Gayon, P., and E. Stonestreet. Le dosage des anthocyanes dans le vin rouge. Bull. Soc. Chim. Fr. 9:2649-52 (1965).

20. Rivas-Gonzalo, J. C., Y. Gutiérrez, E. Hebrero, and C. Santos-Buelga. Comparison of methods for the determination of anthocyanins in red wines. Am. J. Enol. Vitic., 43:210-14 (1992).

21. Roggero, J. P., J. L. Larice, C. Rocheville-Divorne, P. Archier, and S. Coen. Composition anthocyanique des cépages. I. Essai de classification par analyse en composantes principales et par analyse factorielle discriminante. Rev. Française d' Oenol. 112:41-8 (1988).

22. Santamaria, G., J. L. Garrido, and C. Diez. The use of phenol compounds as parameters for distinguished red and rosé wines from pale wine in multivariate analysis. Z. Lebensm. Unters. Forsch. 182:112-14 (1986).

23. Santos, C., S. S. Muñoz, Y. Gutiérrez, E. Hebrero, J. L. Vicente, P. Galindo, and J. C. Rivas. Characterization of young red wines by application of HJ biplot analysis to anthocyanin profiles. J. Agric. Food Chem. 39:1086-90 (1991).

24. Somers, T. C., and M. E. Evans. Spectral evaluation of young red wines: anthocyanins equilibria, total phenolics, free and molecular SO_2, "chemical age". J. Sci. Food Agric. 28:279-87 (1977).

25. Sudraud, P. Interpretation des courbes d'absorption des vins rouges. Ann. Technol. Agric. 7:203-8 (1958).

26. Swain, T., and E. Hillis. The phenolic constituents of *Prunus domestica*. J. Sci. Food Agric. 10:63 (1959).

27. Tapias, R. M., M. S. Larrechi, J. Guasch, J. Rubio, and F. X. Rius. Enological parameters and pattern recognition methods in the geographical differentiation of Spanish red wines. Am. J. Enol. Vitic. 37:195-201 (1986).

28. Vasconcelos, A. M. P., and H. J. Chaves das Neves. Characterization of elementary wines of *Vitis vinifera* varieties by pattern recognition of free amino acid profiles. J. Agric. Food Chem., 37:931-7 (1989).