

Chemometric Characterization and Classification of Five Venetian White Wines[†]

Ivo Moret, Giuseppe Scarponi,* and Paolo Cescon

Dipartimento di Scienze Ambientali, Università di Venezia, Dorsoduro 2137, I-30123 Venezia, Italy

A chemometric multiyear study of the characterization and classification of wines by chemical variables is presented. Five typical Venetian white wines (from the Veneto and Friuli regions of Italy) are considered: Soave Classico, Prosecco di Conegliano-Valdobbiadene, Verduzzo del Piave, Tocai di Lison, and Tocai delle Grave del Friuli. Data collected concern 273 samples originating from vintages between 1977 and 1984, and 27 chemical analytical variables (18 inorganic and classical determinations and 9 aroma compounds). The year-to-year variability is discussed. The multivariate statistical methods include the classification by Euclidean distances from group centroids, the *k*-nearest-neighbor rule, principal component analysis (and its derived method SIMCA), and canonical variate analysis. Statistical analysis reveals clear group structures in the data set and an acceptable wine separation in the multidimensional space. Best results are obtained by canonical variate analysis (overall correct classification about 90%), which shows that several chemical variables contribute to the classification, *cis*-3-hexen-1-ol being by far the most important, followed by 1-hexanol, potassium, nitrogen compounds, and total phenols.

INTRODUCTION

In food chemistry, the application of chemometric methods for characterizing or classifying products according to origin, quality, variety, type, or other features has already attracted considerable attention from researchers (Martens et al., 1983; Forina et al., 1987).

In particular, numerous applications of multivariate methods for wine classification are reported, in which several classes of chemical variables and sensory properties are considered and different statistical methods are used. An extensive review of the literature in this field is given by Forina et al. (1987).

More recently, pattern recognition methods have been used in the characterization and classification of the following wines: Catalan (Spain) white wines by classical oenological parameters and micro and trace metallic constituents (Larrecchi et al., 1988); Portuguese wines by free amino acid composition (Vasconcelos et al., 1989); Chardonnay wines of Trentino (Italy) by amino acids, volatile compounds, and metal ions (Seeber et al., 1991); French red wines by major acids (Etievant et al., 1989); Cabernet Sauvignon wines of California by headspace volatiles (Noble et al., 1993); German white wines by NMR spectroscopy (Vogel et al., 1993).

In previous papers we showed that multivariate statistical methods can be applied successfully to the chemometric characterization and classification of typical wines (Moret et al., 1980, 1983, 1984a-c, 1986, 1988; Scarponi et al., 1982, 1990). Five typical Venetian white wines (from the Veneto and Friuli regions of Italy, vintages from 1977 to 1983) were investigated; a few classical determinations, the concentration of several inorganic ions, and a series of aroma compounds were considered as chemical analytical variables. The wines considered in the study were the following: Soave Classico, Prosecco di Conegliano-Valdobbiadene, Verduzzo del Piave, Tocai di Lison Classico, and Tocai delle Grave del Friuli. The

statistical methods applied were the Fisher linear discriminant analysis, canonical variate analysis, the *k*-nearest-neighbor rule, principal component analysis, and the SIMCA method. Unfortunately, since the aroma compounds were not measured for the first few vintages, the complete data set for all of the variables was available only for samples of 1981, 1982, and 1983 vintages, with no more than two vintage years for each type of wine (Moret et al., 1986).

In the present work we analyze samples of the same five types of wine obtained from one more vintage year (1984). Chemical analytical results are merged with those already available, and the whole data set 1977-1984 is considered in studying temporal variability, while the homogeneous subset 1981-1984 (see below) is analyzed by univariate and multivariate statistical methods for characterization and classification purposes. The aim is to give greater account of the year-to-year variations of the chemical composition of wines in the chemometric investigation, particularly the classification ability of the procedure.

EXPERIMENTAL PROCEDURES

Wine Samples. The study concerns the following five Venetian white wines (the abbreviations used in the text below are given in parentheses): Soave Classico (Soave, S), Prosecco di Conegliano-Valdobbiadene (Prosecco, P), Verduzzo del Piave (Verduzzo, V), Tocai di Lison Classico (Tocai L., L), and Tocai delle Grave del Friuli (Tocai G., G). In this work measurements are extended to wines of the 1984 vintage; in particular, 59 samples were analyzed, i.e., 11 samples of Soave wine, 10 Prosecco, 13 Verduzzo, 7 Tocai L., and 18 Tocai G. (Table 1). As in previous investigations, samples were collected in the production zones, directly from wine producers, in bottles ready to market and stored in our laboratory at 3-4 °C until analysis. The wines were guaranteed to be typical by the Italian D.O.C. (Denominazione di Origine Controllata; Certified Brand of Origin) brand (Moret et al., 1986).

Analytical Determinations. Samples were analyzed to determine the following (abbreviations are given in parentheses): (1) sodium (Na), (2) potassium (K), (3) calcium (Ca), (4) magnesium (Mg), (5) chloride (Cl), (6) pH (pH), (7) titratable acidity (TA, given as tartaric acid), (8) phosphorus (P, given as P₂O₅), (9) ash content (AC), (10) alkalinity of the ash (ALA), (11) lithium (Li), (12) total nitrogen (TN), (13) theoretical sugar-free

[†] Part of the work was presented to the International Congress Euroanalysis VI (Paris, Sept 7-11, 1987), and it was awarded a Lavoisier Medal for quality and presentation.

Table 1. Number of Wine Samples by Type of Wine and Vintage Year

| vintage year | wine | | | | | total |
|---------------------------|-------|----------|----------|----------|----------|-------|
| | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. | |
| 1977 ^a | 14 | 14 | 14 | | | 42 |
| 1979 ^b | | 19 | | | | 19 |
| 1980 ^b | 8 | 14 | | 10 | | 32 |
| 1981 ^c | 10 | | 18 | 11 | | 39 |
| 1982 ^c | 11 | 12 | 16 | 10 | 18 | 67 |
| 1983 ^c | | 12 | | | 15 | 27 |
| 1984 ^c | 11 | 10 | 13 | 7 | 18 | 59 |
| total | 54 | 81 | 61 | 38 | 51 | 285 |
| total subset ^c | 32 | 34 | 47 | 28 | 51 | 192 |

^a Samples analyzed for the first 10 classical parameters only.

^b Samples analyzed for all of the classical parameters (first 18).

^c Samples analyzed for classical and aroma parameters.

dry extract (DE), (14) absorbance measured at 420 nm (A), (15) total phenols (TP, given as (+)-catechin), (16) 1-propanol (POL), (17) 2-methyl-1-propanol (MPOL), (18) 2-methyl-1-butanol plus 3-methyl-1-butanol (MBOL), (19) 2-methylpropanoic acid (MPA), (20) butanoic acid (BA), (21) 3-methylbutanoic acid (MBA), (22) hexanoic acid (HA), (23) octanoic acid (OA), (24) decanoic acid (DA), (25) 1-hexanol (HOL), (26) *trans*-3-hexen-1-ol (THOL), (27) *cis*-3-hexen-1-ol (CHOL). The measurement units of chemical variables are reported in Table 2; in some cases, a multiplicative factor is used for display purposes only. In the following, the first 18 determinations will be referred to as classical variables, while the latter 9 will be referred to as aroma variables.

Sodium and potassium (Moret et al., 1980) and lithium (Moret et al., 1983) were determined by flame emission spectrophotometry; calcium (Scarponi et al., 1982) and magnesium (Moret et al., 1980) were determined by atomic absorption spectrophotometry. Phosphorus (according to Schneyder), chloride (according to Volhard), total nitrogen (according to Kjeldahl), theoretical sugar-free dry extract, absorbance (measured at 420 nm), total phenols (according to Folin-Ciocalteu), pH, titratable acidity, ash content, and alkalinity of the ash were determined by classical procedures (Moret et al., 1980, 1983, 1986).

1-Propanol, 2-methyl-1-propanol, and 2-methyl-1-butanol plus 3-methyl-1-butanol were determined by gas chromatography after wine distillation using 2-pentanol as internal standard (Moret et al., 1983).

The volatile aroma components from 2-methylpropanoic acid to *cis*-3-hexen-1-ol were quantified by capillary gas chromatography after continuous liquid-liquid extraction with a pentane/dichloromethane mixture (2:1 by volume) and concentration to 1 mL in a Vigreux column (internal standard 1-heptanol) (Moret et al., 1984b). Identification of compounds by gas chromatography-mass spectrometry was obtained by coupling a Carlo Erba gas chromatograph (Model 5160) to a Hewlett-Packard mass spectrometer (Model 5970, mass selective detector).

The other laboratory equipment and reagents were the same as those mentioned in previous papers (Moret et al., 1980, 1981, 1984, 1986).

To assure consistency of data over the years, repeatability of measurements was checked, as usual, by reanalyzing some of the earlier samples, stored at 3–4 °C, before analysis of the new samples was begun.

Analyses of samples of the 1984 vintage were performed in the period October 1985–March 1986.

Data Set. Table 1 gives the number of samples analyzed for each type of wine and for each vintage year. It should be noted that at the beginning (vintage 1977) only the first 10 classical variables were measured; later measurements were extended to include initially the first 18 classical variables (vintages 1979 and 1980) and then all of the classical and aroma variables (vintages from 1981 to 1984). With regard to wines, the study started with a consideration of the Soave, Prosecco, and Verduzzo wines (vintage 1977); then the investigation was gradually extended to Tocai L. wine (since 1980 vintage) and to Tocai G. wine (since 1982). The number of samples analyzed for each type of wine is different depending on the size of the production area.

The whole data set (i.e., vintages from 1977 to 1984) was used only for the observation of temporal variability of chemical variables from year to year.

The data used in the context of multivariate analysis refer to the homogeneous subset of results related to the 1981–1984 vintages in which all of the variables are measured. It is noted that a 3-year complete set of data for each wine is present in this subset. In particular, for the 1982 and 1984 vintages samples from all five types of wines were collected, while for the vintages of 1981 and 1983 a complementary sampling was adopted (see Table 1).

Statistical Methods. The subdivision of samples over time and between wines (Table 1) suggested the choice of a descriptive-exploratory approach for the chemometric investigation of the data set.

First, we obtained univariate summary statistics of the analytical data. Univariate comparisons between samples of the same type of wine grouped according to the vintage year and between samples of different wines were made by considering the ratio of between-group variance to total variance; this ratio ranges from zero to one according to the increasing dissimilarity of the groups.

The multivariate statistical methods used for group structure evaluation of the wines are the classification of samples by their Euclidean distances from the group centroids (EDGC), the *k*-nearest-neighbor rule (KNN), principal component analysis (PCA), soft independent modeling of class analogy (SIMCA), and canonical variate analysis (CVA) (Varmuza, 1980; Massart et al., 1988). The methods EDGC and KNN are performed on standardized data (mean = 0, standard deviation = 1) obtained by considering the whole data set. In the SIMCA method standardization is made by grouping data by type of wine. In the PCA method standardization is performed on either grouped or ungrouped data according to the particular analysis.

Calculations were performed by using the following statistical packages: SIMCA (Wold and Sjostrom, 1977), SPSS (Hull and Nie, 1981), SAS (1985), STATGRAPHICS (1992), and PARVUS (1988).

RESULTS AND DISCUSSION

Chemical Results. The chemical results obtained in the present work for samples of 1984 vintage are collected, together with previous data, in an extensive table available as supplementary material. This table reports summary statistics of the complete data set (vintages from 1977 to 1984), i.e., mean, median, minimum–maximum values, and standard deviation for all of the chemical variables by type of wine and vintage year; the overall statistics and the ratio of between-group variance to total variance are also given. With regard to the 1981–1984 subset, the overall statistics are reported in Table 2.

Year-to-Year Variability. As has been pointed out previously (Moret et al., 1986), data variability due to different vintage years can be important for some variables and/or wines; it determines a reduction in the separation of different wines. To consider this aspect further, particularly to evaluate the effect with respect to aroma variables, we considered again the complete data set, its summary statistics since the 1977 vintage (data available as supplementary material), and graphical presentation of data by box plots displayed on the temporal scale. Table 3 and Figure 1, respectively, report annual statistics data and box plots for selected variables.

Data obtained on samples of the 1984 vintage confirm that annual variability, as appears from the ratio of between-group (year) variance to total variance (supplementary material), is generally limited. In particular, from analysis of the 135 ratios (27 variables by 5 wines) it can be seen that the contribution of the year-to-year variability to the total variance is not greater than 0.25 in 72% of the cases; it ranges between 0.25 and 0.40 in 18% of the cases, and it is more than 0.40 in 10% of the cases. Only in seven

Table 2. Mean, Relative Standard Deviation (RSD), Median, and Minimum-Maximum Values of Chemical Variables, by Type of Wine (1981-1984 Vintages)

| no. | chemical variable | units | wine | | | | |
|-----|-------------------|-----------------------------|------------------|------------------|------------------|------------------|------------------|
| | | | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. |
| 1 | Na | mg/L × 10 ⁻¹ | 1.6 (0.46) | 1.5 (0.52) | 3.0 (0.52) | 2.1 (0.56) | 2.5 (0.44) |
| | | | 1.5 (0.3-2.9) | 1.3 (0.3-3.7) | 2.7 (0.8-8.7) | 1.8 (0.3-5.4) | 2.3 (1.0-5.5) |
| 2 | K | mg/L × 10 ⁻² | 4.7 (0.26) | 5.9 (0.21) | 7.8 (0.22) | 7.6 (0.17) | 7.7 (0.23) |
| | | | 4.6 (1.9-8.2) | 6.0 (2.8-8.1) | 7.7 (4.8-14.4) | 7.1 (4.3-10.4) | 7.7 (4.9-12.6) |
| 3 | Ca | mg/L × 10 ⁻¹ | 8.4 (0.21) | 9.8 (0.15) | 10.5 (0.24) | 9.1 (0.18) | 10.6 (0.21) |
| | | | 8.2 (5.1-13.0) | 9.6 (7.9-13.8) | 10.4 (5.7-17.9) | 9.4 (5.5-11.2) | 10.8 (5.7-15.6) |
| 4 | Mg | mg/L × 10 ⁻¹ | 7.8 (0.18) | 7.1 (0.12) | 7.5 (0.26) | 7.7 (0.15) | 7.2 (0.18) |
| | | | 7.6 (5.3-10.9) | 7.1 (5.4-9.1) | 7.2 (4.2-14.7) | 7.7 (5.5-9.9) | 6.9 (5.4-11.4) |
| 5 | Cl | mg/L × 10 ⁻¹ | 1.4 (0.36) | 1.8 (0.40) | 3.4 (0.55) | 2.8 (0.64) | 1.9 (0.94) |
| | | | 1.4 (0.7-2.4) | 2.1 (0.3-3.5) | 2.8 (0.8-9.2) | 2.1 (0.4-7.8) | 1.4 (0.4-8.5) |
| 6 | pH | | 3.23 (0.04) | 3.27 (0.05) | 3.38 (0.04) | 3.36 (0.04) | 3.38 (0.05) |
| | | | 3.22 (3.00-3.68) | 3.27 (2.94-3.59) | 3.46 (3.11-3.71) | 3.37 (2.98-3.62) | 3.35 (3.07-3.81) |
| 7 | TA ^a | g/L | 6.1 (0.13) | 5.8 (0.16) | 6.2 (0.13) | 5.8 (0.14) | 6.0 (0.14) |
| | | | 6.0 (4.8-8.9) | 5.6 (4.5-7.7) | 6.2 (4.4-8.6) | 5.9 (4.5-7.5) | 6.0 (4.0-9.0) |
| 8 | P ^b | mg/L × 10 ⁻² | 1.7 (0.33) | 1.5 (0.39) | 1.5 (0.34) | 2.1 (0.31) | 1.9 (0.42) |
| | | | 1.8 (0.6-2.7) | 1.6 (0.6-3.0) | 1.4 (0.7-2.7) | 2.3 (0.7-3.0) | 1.7 (0.4-3.8) |
| 9 | AC | g/L | 1.3 (0.24) | 1.6 (0.17) | 1.9 (0.19) | 1.9 (0.14) | 1.8 (0.31) |
| | | | 1.3 (0.7-2.0) | 1.6 (1.0-2.1) | 2.0 (0.8-2.8) | 2.0 (1.4-2.6) | 1.8 (0.7-3.1) |
| 10 | ALA | mequiv/L × 10 ⁻¹ | 1.5 (0.16) | 1.9 (0.14) | 2.1 (0.18) | 2.0 (0.13) | 1.9 (0.27) |
| | | | 1.5 (1.1-2.2) | 2.0 (1.2-2.3) | 2.1 (1.2-2.8) | 2.0 (1.2-2.5) | 1.9 (1.2-3.6) |
| 11 | Li | μg/L × 10 ⁻¹ | 1.0 (0.43) | 1.1 (0.18) | 1.4 (0.27) | 1.5 (0.29) | 1.1 (0.22) |
| | | | 0.8 (0.6-2.4) | 1.3 (0.8-1.6) | 1.3 (0.8-2.6) | 1.5 (0.8-2.7) | 1.1 (0.6-1.6) |
| 12 | TN | mg/L × 10 ⁻² | 2.5 (0.33) | 1.4 (0.34) | 1.6 (0.33) | 1.2 (0.50) | 1.6 (0.42) |
| | | | 2.5 (0.8-4.1) | 1.3 (0.6-3.2) | 1.6 (0.9-3.0) | 1.1 (0.5-3.2) | 1.6 (0.2-3.6) |
| 13 | DE | g/L × 10 ⁻¹ | 1.8 (0.13) | 1.8 (0.15) | 1.9 (0.12) | 2.0 (0.12) | 2.0 (0.18) |
| | | | 1.8 (1.2-2.2) | 1.7 (1.4-2.6) | 1.9 (1.5-2.8) | 2.0 (1.7-2.6) | 1.9 (1.4-3.5) |
| 14 | A | × 10 | 1.0 (0.39) | 1.3 (0.74) | 1.2 (0.43) | 1.4 (0.41) | 1.0 (0.38) |
| | | | 1.0 (0.5-2.1) | 1.0 (0.5-5.8) | 1.0 (0.4-2.7) | 1.4 (0.7-3.3) | 1.0 (0.3-2.1) |
| 15 | TP ^c | mg/L × 10 ⁻² | 2.4 (0.23) | 2.8 (0.67) | 3.4 (0.26) | 2.8 (0.21) | 2.2 (0.35) |
| | | | 2.3 (1.4-3.8) | 2.4 (1.4-13.0) | 3.3 (1.9-5.0) | 2.8 (1.8-4.3) | 2.0 (1.2-4.9) |
| 16 | POL | mg/L × 10 ⁻¹ | 2.6 (0.31) | 2.9 (0.25) | 3.1 (0.27) | 2.5 (0.30) | 3.0 (0.44) |
| | | | 2.4 (1.3-4.1) | 2.9 (1.7-4.5) | 3.0 (1.8-5.5) | 2.5 (1.0-4.2) | 2.9 (0.8-7.5) |
| 17 | MPOL | mg/L × 10 ⁻¹ | 5.1 (0.39) | 6.5 (0.33) | 7.8 (0.28) | 8.2 (0.24) | 8.4 (0.33) |
| | | | 5.1 (1.5-8.4) | 5.9 (3.0-11.5) | 7.2 (3.7-13.8) | 8.1 (4.4-11.9) | 8.1 (3.2-14.2) |
| 18 | MBOL | mg/L × 10 ⁻² | 2.1 (0.33) | 2.2 (0.25) | 2.4 (0.18) | 2.8 (0.20) | 2.7 (0.25) |
| | | | 2.0 (1.0-4.2) | 2.1 (0.9-3.4) | 2.4 (1.0-3.2) | 2.8 (1.6-4.3) | 2.8 (1.1-4.2) |
| 19 | MPA | mg/L | 1.2 (0.54) | 1.3 (0.30) | 1.7 (0.40) | 2.2 (0.39) | 2.0 (0.37) |
| | | | 1.1 (0.5-3.5) | 1.4 (0.5-2.3) | 1.6 (0.5-4.0) | 2.1 (1.1-4.5) | 1.8 (0.7-4.3) |
| 20 | BA | mg/L | 1.2 (0.43) | 1.1 (0.45) | 1.4 (0.62) | 1.1 (0.59) | 1.3 (0.47) |
| | | | 1.1 (0.5-2.6) | 1.0 (0.4-2.8) | 1.1 (0.2-4.7) | 1.0 (0.3-3.6) | 1.1 (0.6-3.6) |
| 21 | MBA | mg/L × 10 | 7.2 (0.52) | 7.1 (0.32) | 9.1 (0.38) | 13.0 (0.43) | 10.9 (0.33) |
| | | | 6.6 (3.3-20.7) | 7.3 (2.6-13.6) | 9.0 (3.3-19.9) | 11.9 (4.1-31.1) | 10.6 (4.8-20.2) |
| 22 | HA | mg/L | 3.0 (0.47) | 2.9 (0.33) | 2.0 (0.48) | 2.2 (0.38) | 2.7 (0.38) |
| | | | 2.7 (0.7-6.4) | 2.8 (1.3-5.3) | 1.7 (0.9-5.8) | 2.0 (1.1-4.9) | 2.5 (0.9-5.0) |
| 23 | OA | mg/L | 4.5 (0.50) | 4.8 (0.38) | 3.0 (0.51) | 3.4 (0.42) | 3.9 (0.42) |
| | | | 4.3 (1.0-8.8) | 4.9 (1.9-10.2) | 2.5 (1.2-8.2) | 3.2 (1.2-7.6) | 3.6 (1.3-8.3) |
| 24 | DA | mg/L × 10 | 9.8 (0.70) | 13.4 (0.54) | 7.8 (0.64) | 9.1 (0.49) | 10.2 (0.60) |
| | | | 8.7 (1.9-27.9) | 12.9 (3.3-38.3) | 7.0 (1.5-28.9) | 9.2 (0.7-18.9) | 9.8 (1.1-27.2) |
| 25 | HOL | mg/L | 1.2 (0.42) | 2.2 (0.37) | 2.4 (0.27) | 3.4 (0.31) | 2.5 (0.24) |
| | | | 1.1 (0.6-2.9) | 2.2 (0.9-3.7) | 2.3 (1.3-3.7) | 3.3 (1.9-5.7) | 2.5 (1.2-3.7) |
| 26 | THOL | mg/L × 10 ² | 7.2 (0.43) | 5.8 (0.52) | 8.4 (0.75) | 10.0 (0.50) | 7.9 (1.12) |
| | | | 6.4 (2.4-14.6) | 4.4 (1.7-12.2) | 6.2 (1.5-29.2) | 9.7 (3.6-21.0) | 6.0 (1.0-65.0) |
| 27 | CHOL | mg/L × 10 | 0.27 (0.53) | 1.1 (0.33) | 1.8 (0.37) | 4.2 (0.32) | 3.5 (0.48) |
| | | | 0.26 (0.07-0.84) | 1.1 (0.4-2.2) | 1.8 (0.2-4.1) | 4.2 (1.4-7.1) | 3.3 (1.2-8.9) |

^a Given as tartaric acid. ^b Given as P₂O₅. ^c Given as (+)-catechin.

Table 3. Median and First-Third Quartile Values for Selected Variables by Type of Wine and Vintage Year and Ratio of Between-Group (Year) Variance to the Total Variance (Variance Ratio)

| chemical variable ^a | year | median (first-third quartile) | | | | |
|--|--|--|--|-------------------------|-----------------------|-------------------------|
| | | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. |
| K | 1977 | 6.6 (5.3-8.3) | 7.9 (6.8-8.2) | 9.7 (8.0-11.2) | | |
| | 1979 | | 7.2 (6.2-8.6) | | | |
| | 1980 | 3.7 (3.5-4.3) | 8.7 (7.2-10.2) | | 6.5 (6.0-7.5) | |
| | 1981 | 5.2 (4.8-5.9) | | 8.1 (7.7-9.6) | 9.0 (7.4-9.2) | |
| | 1982 | 4.1 (2.9-4.6) | 6.1 (5.7-7.5) | 6.4 (5.8-7.8) | 7.0 (6.8-7.1) | 7.9 (7.0-9.1) |
| | 1983 | | 6.0 (5.0-6.5) | | | 7.7 (6.0-8.1) |
| | 1984 | 4.3 (4.1-4.6) | 5.9 (5.0-6.5) | 7.8 (7.2-8.6) | 7.1 (6.5-8.5) | 7.2 (6.1-8.3) |
| overall ^b variance ratio | | 4.8 (4.1-5.8) 0.47 | 6.8 (5.9-8.1) 0.34 | 8.0 (6.8-9.6) 0.22 | 7.1 (6.7-8.5) 0.20 | 7.7 (6.3-8.3) 0.06 |
| Li | 1977 | | | | | |
| | 1979 | | 0.6 (0.6-0.8) | | | |
| | 1980 | 1.2 (0.9-7.3) | 0.5 (0.4-0.6) | | 0.6 (0.6-0.8) | |
| | 1981 | 0.9 (0.7-1.8) | | 1.2 (1.1-1.4) | 1.5 (1.1-1.6) | |
| | 1982 | 0.8 (0.7-1.2) | 1.0 (0.9-1.2) | 1.2 (1.0-1.6) | 1.4 (1.0-1.5) | 1.1 (0.9-1.3) |
| | 1983 | | 1.2 (1.0-1.3) | | | 1.0 (0.9-1.1) |
| | 1984 | 0.8 (0.7-1.1) | 1.2 (1.0-1.3) | 1.7 (1.6-1.8) | 2.0 (1.7-2.2) | 1.1 (1.0-1.3) |
| overall variance ratio | | 0.9 (0.7-1.3) 0.08 | 0.9 (0.6-1.2) 0.67 | 1.3 (1.1-1.7) 0.25 | 1.4 (0.9-1.6) 0.67 | 1.1 (0.9-1.3) 0.04 |
| MPA | 1981 | 1.0 (0.7-1.4) | | 1.2 (0.9-1.6) | 1.8 (1.3-2.2) | |
| | 1982 | 1.0 (0.8-1.2) | 1.4 (0.9-1.5) | 1.8 (1.5-2.0) | 2.3 (2.1-3.2) | 2.3 (1.9-3.0) |
| | 1983 | | 1.2 (1.0-1.5) | | | 1.5 (1.3-1.8) |
| | 1984 | 1.1 (0.9-1.4) | 1.4 (1.1-1.6) | 1.8 (1.3-2.2) | 1.8 (1.2-3.5) | 1.8 (1.5-2.1) |
| | overall ^b variance ratio | | 1.1 (0.8-1.4) 0.02 | 1.4 (1.0-1.6) 0.03 | 1.6 (1.2-2.0) 0.18 | 2.1 (1.6-2.7) 0.16 |
| DA | 1981 | 7.9 (2.8-12.8) | | 6.1 (3.6-9.2) | 9.2 (3.0-13.7) | |
| | 1982 | 4.3 (3.0-14.1) | 7.9 (5.9-14.4) | 6.8 (4.8-8.6) | 10.2 (8.4-10.9) | 6.8 (3.9-11.7) |
| | 1983 | | 14.2 (6.8-20.0) | | | 9.8 (6.9-11.9) |
| | 1984 | 12.6 (4.9-14.1) | 14.0 (10.0-16.2) | 10.7 (5.1-13.6) | 8.4 (7.6-10.0) | 11.2 (6.7-14.2) |
| | overall variance ratio | | 8.7 (3.8-13.7) 0.01 | 12.5 (6.4-16.5) 0.10 | 7.0 (4.0-9.7) 0.15 | 9.1 (6.6-11.1) <0.01 |
| HOL | 1981 | 0.8 (0.7-1.3) | | 2.2 (2.1-2.8) | 3.4 (3.0-5.3) | |
| | 1982 | 1.1 (1.0-1.4) | 2.3 (1.6-3.1) | 2.4 (2.1-3.1) | 3.2 (3.0-4.1) | 2.7 (2.0-3.1) |
| | 1983 | | 2.7 (1.8-3.4) | | | 2.5 (2.0-3.1) |
| | 1984 | 1.1 (0.9-1.7) | 1.8 (1.4-2.3) | 2.0 (1.5-2.7) | 2.8 (2.1-4.2) | 2.4 (2.1-2.7) |
| | overall ^b variance ratio | | 1.1 (0.8-1.3) 0.07 | 2.2 (1.6-2.9) 0.09 | 2.3 (2.0-3.0) 0.04 | 3.3 (2.8-4.1) 0.07 |
| CHOL | 1981 | 0.1 ₈ (0.1 ₂ -0.1 ₈) | | 1.6 (1.3-1.9) | 4.0 (3.6-5.4) | |
| | 1982 | 0.3 ₀ (0.1 ₅ -0.3 ₆) | 0.9 (0.8-1.2) | 1.6 (1.4-2.2) | 4.0 (3.7-4.8) | 3.6 (2.9-3.7) |
| | 1983 | | 1.2 (1.0-1.3) | | | 2.5 (2.1-2.8) |
| | 1984 | 0.2 ₉ (0.2 ₆ -0.3 ₃) | 1.2 (0.8-1.3) | 2.3 (1.8-2.4) | 4.6 (2.8-5.2) | 3.8 (2.9-5.6) |
| | overall variance ratio | | 0.2 ₆ (0.1 ₈ -0.3 ₃) 0.03 | 1.1 (0.8-1.3) 0.02 | 1.8 (1.4-2.2) 0.21 | 4.2 (3.6-5.0) 0.04 |

^a Measurement units and multiplicative factors as in Table 2. ^b Overall: results for all of the available vintages.

cases is the ratio greater than 0.60 (0.75 being the maximum value), four of which concern phosphorus, two lithium, and the last magnesium.

Considering the annual variability, some interesting situations are noted. With respect to classical variables, major variations are observed for K, TN, AC, ALA, and P, but the tendency to separate the wines is confirmed, except for the last variable (Table 3; Figure 1). It is interesting to note that for AC, P, and, to a lesser degree, K and ALA, variation is due to a clear decreasing trend, for practically all of the wines. This trend can be tentatively associated to changes in winemaking technology (e.g., crushing technology, decrease of the duration of skin-juice contact time). In the case of Li (Table 3) there is remarkable stability with respect to Soave and Tocai G., but high variability for the other wines, which drastically reduces the tendency of this variable to separate the wines even though differences are observed in each vintage year. Other classical variables show less year-to-year variation but also fewer differences among wines.

The annual variability of aroma compounds is, in general, comparable to that of classical variables. Acids BA, HA, and OA show major variations. In particular,

BA (but also HA and OA to a lesser degree) appears stable for the years 1981-1983, while in 1984 it presents values that are considerably higher than those observed in previous years (see Figure 1). Acid DA (Table 3) appears not to be affected by the vintage factor, but it does not show any change between wines. The same is valid for THOL. Variables that show noticeable stability coupled with moderate-high tendency to separate wines are represented by MPA, HOL, and CHOL (Table 3 and Figure 1).

Characterization and Classification of Wines. For the sake of clarity it is perhaps useful to remember that the set of data for the statistical analysis below refers to the 1981-1984 subset.

Univariate Statistical Analysis. The chemical analytical results related to the 1981-1984 subset are summarized in Table 2. For each type of wine, data of the three vintages are grouped together and the following statistics are reported: the mean, the relative standard deviation (RSD), the median, and the minimum and maximum values. It can be observed that generally the RSD is not too high and the values of mean and median are in good agreement.

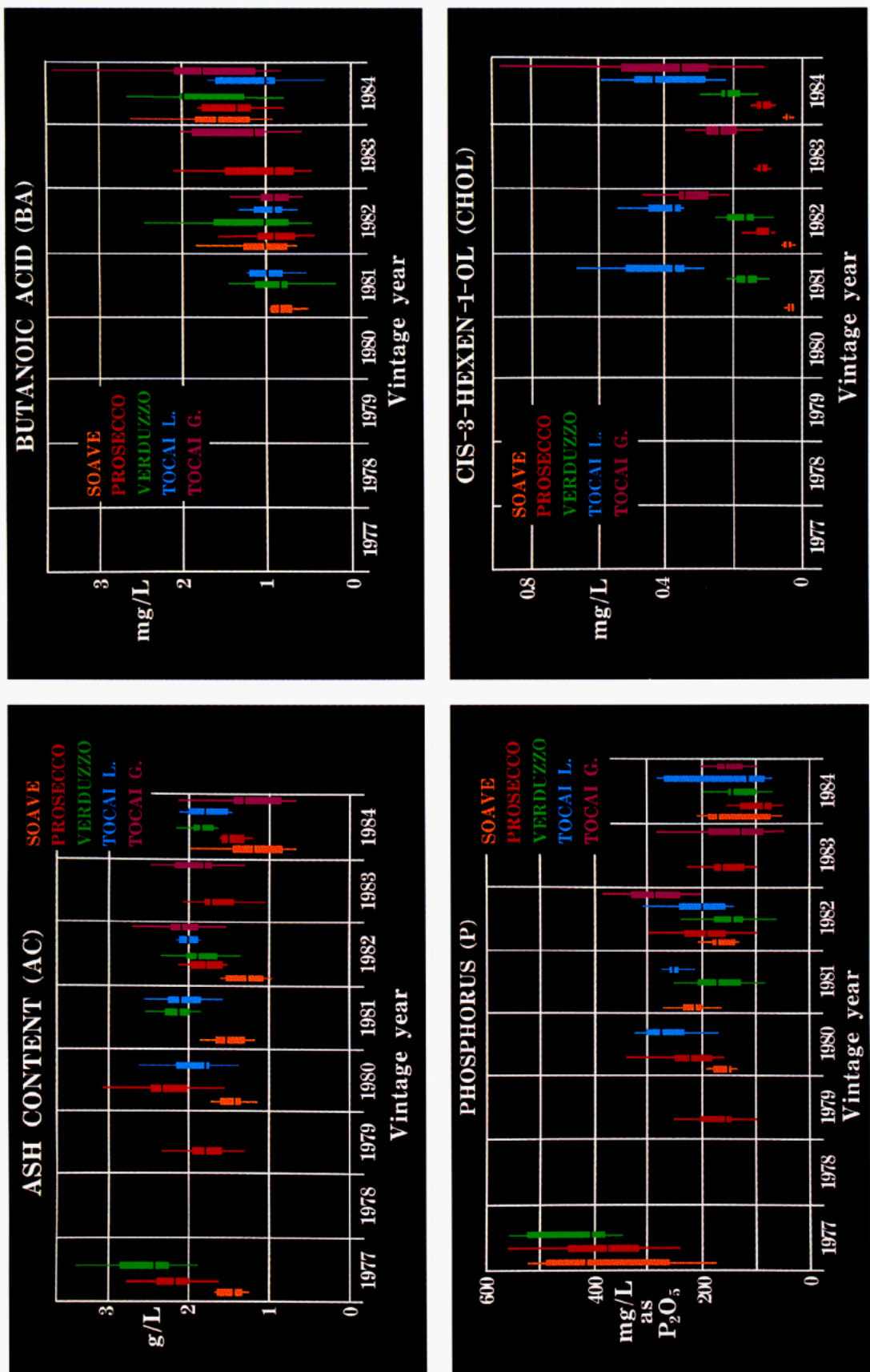


Figure 1. Box plots for selected variables by type of wine and vintage year.

Table 4. Ratios of Between-Group Variance to Total Variance for 1981–1984 Vintages

| no. | chemical variable | between years ^a | | | | | between wines ^b |
|-----|-------------------|----------------------------|----------|----------|----------|----------|----------------------------|
| | | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. | |
| 1 | Na | 0.01 | 0.47 | 0.06 | 0.22 | 0.13 | 0.20 |
| 2 | K | 0.23 | 0.05 | 0.15 | 0.31 | 0.06 | 0.38 |
| 3 | Ca | 0.03 | 0.03 | 0.35 | 0.05 | 0.24 | 0.15 |
| 4 | Mg | 0.26 | 0.33 | 0.03 | 0.06 | 0.07 | 0.03 |
| 5 | Cl | 0.14 | 0.06 | 0.03 | <0.01 | 0.09 | 0.19 |
| 6 | pH | 0.11 | 0.07 | 0.17 | 0.03 | 0.20 | 0.15 |
| 7 | TA | 0.28 | 0.39 | 0.22 | 0.05 | 0.02 | 0.04 |
| 8 | P | 0.30 | 0.43 | 0.12 | 0.32 | 0.62 | 0.12 |
| 9 | AC | 0.18 | 0.18 | 0.19 | 0.21 | 0.50 | 0.24 |
| 10 | ALA | 0.15 | 0.02 | 0.01 | 0.17 | 0.20 | 0.20 |
| 11 | Li | 0.08 | 0.13 | 0.25 | 0.43 | 0.04 | 0.21 |
| 12 | TN | 0.30 | 0.24 | 0.20 | <0.01 | 0.05 | 0.28 |
| 13 | DE | 0.10 | 0.19 | 0.22 | 0.38 | 0.03 | 0.07 |
| 14 | A | 0.32 | 0.01 | 0.13 | 0.02 | 0.06 | 0.06 |
| 15 | TP | 0.33 | 0.06 | 0.02 | 0.09 | 0.05 | 0.16 |
| 16 | POL | 0.36 | 0.24 | 0.05 | 0.03 | 0.05 | 0.06 |
| 17 | MPOL | 0.23 | 0.17 | 0.06 | 0.01 | 0.25 | 0.21 |
| 18 | MBOL | 0.16 | 0.19 | 0.18 | <0.01 | 0.15 | 0.17 |
| 19 | MPA | 0.02 | 0.03 | 0.18 | 0.16 | 0.30 | 0.24 |
| 20 | BA | 0.40 | 0.21 | 0.20 | <0.01 | 0.29 | 0.02 |
| 21 | MBA | 0.07 | 0.13 | 0.18 | 0.13 | 0.15 | 0.23 |
| 22 | HA | 0.32 | 0.28 | 0.36 | 0.25 | 0.23 | 0.13 |
| 23 | OA | 0.19 | 0.17 | 0.28 | 0.23 | 0.17 | 0.13 |
| 24 | DA | 0.01 | 0.10 | 0.15 | <0.01 | 0.09 | 0.09 |
| 25 | HOL | 0.07 | 0.09 | 0.04 | 0.07 | 0.01 | 0.44 |
| 26 | THOL | 0.23 | 0.25 | 0.05 | 0.03 | 0.02 | 0.05 |
| 27 | CHOL | 0.03 | 0.02 | 0.21 | 0.04 | 0.22 | 0.63 |

^a Data grouped by vintage year. ^b Data grouped by type of wine.

Table 5. Correlations Greater than or Equal to |0.70| for 1981–1984 by Type of Wine

| chemical variables | wine | | | | | overall |
|--------------------|-------|----------|----------|----------|----------|---------|
| | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. | |
| HA-OA | 0.92 | 0.92 | 0.92 | 0.96 | 0.90 | 0.93 |
| OA-DA | 0.73 | 0.89 | 0.89 | | 0.88 | 0.81 |
| MPA-MBA | 0.86 | | 0.86 | 0.86 | | 0.81 |
| HA-DA | | | 0.76 | | 0.81 | |
| BA-HA | 0.81 | | | | | |
| BA-OA | 0.76 | | | | | |
| MBOL-MBA | 0.83 | 0.74 | | | 0.75 | |
| MPOL-HA | -0.77 | | | | | |
| MBOL-HA | -0.70 | | | | | |
| MBOL-OA | -0.70 | | | | | |
| K-AC | | 0.85 | | 0.75 | 0.71 | 0.72 |
| K-ALA | 0.71 | 0.83 | | | 0.84 | 0.75 |
| K-pH | | 0.73 | | | 0.76 | 0.70 |
| AC-ALA | | 0.78 | | | | |
| pH-TA | | | -0.70 | -0.79 | | |
| pH-ALA | | | | 0.71 | 0.76 | 0.70 |
| P-AC | | 0.70 | | | | |
| Ca-Li | 0.74 | | | | | |

From an enological viewpoint only three samples (two Prosecco wines and one Tocai G.) show unusually high values, i.e., total phenols 1.30×10^3 mg/L, absorbance 0.58, and *trans*-3-hexen-1-ol 0.650 mg/L, respectively.

The influence of different vintages, as apparent from the 1981–1984 subset, was evaluated by considering the ratio of between-group variance to total variance of data grouped according to vintage year. The results, given in Table 4, show that variations of mean values in the three years are generally not important (in only a few cases, about 20%, the ratio is higher than 0.25). The maximum value for this ratio, i.e. 0.63, is observed for the phosphorus variable in Tocai G. wine; however, high values for this variable are observed for the other types of wine too, except the Verduzzo. Details on the annual variability are given above (Year-to-Year Variability).

Table 4 also gives (last column) the variance ratio computed on data grouped by type of wine. The most differentiating variables in the univariate comparison of wines are the *cis*-3-hexen-1-ol (CHOL, variance ratio 0.63),

the 1-hexanol (HOL, ratio 0.44), the potassium (K, ratio 0.38), and the total nitrogen (TN, ratio 0.28).

Selected correlations between variables (coefficients greater than or equal to |0.70|) are reported in Table 5 for each type of wine and the whole data set. The only pair of variables having a high correlation in all cases is that of hexanoic acid and octanoic acid. Other correlations of almost general validity are those between octanoic acid and decanoic acid and between 2-methylpropanoic acid and 3-methylbutanoic acid. It is noted that in no cases are high correlations observed between aroma components and the other variables. The correlations observed may be considered normal from a chemical enological point of view. The wine with the highest correlations is Soave, followed by Prosecco and Tocai G.

Multivariate Statistical Analysis. a. KNN and EDGC Methods. Euclidean and Mahalanobis distances between group centroids are given in Table 6. Considering Euclidean distances, it can be observed that the nearest wine

Table 6. Euclidean and (in Parentheses) Mahalanobis Distances between Group Centroids (1981–1984 Vintages)

| | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. |
|----------|-------|----------|----------|----------|----------|
| Soave | | (3.6) | (4.5) | (6.1) | (4.8) |
| Prosecco | 9.5 | | (3.3) | (4.5) | (4.1) |
| Verduzzo | 15.7 | 3.3 | | (3.8) | (3.0) |
| Tocai L. | 34.5 | 17.5 | 4.6 | | (3.2) |
| Tocai G. | 26.7 | 15.5 | 5.9 | 5.2 | |

to Soave in the multidimensional space is Prosecco, while the other three wines are more distant, with the Tocai L. the most distant. The centroid of Prosecco wine has as its nearest neighbor that of Verduzzo, and this distance is the lowest one. Moreover, the centroid of Verduzzo is about equally separated (with low distances) from Prosecco, Tocai L., and Tocai G. A quite similar situation is observed when the Mahalanobis distances are considered. It is useful to note here that the overall relative position of group centroids in multidimensional space seems to be similar to that observed later using the canonical variates analysis (see Figure 2).

Table 7 reports the classification results obtained using KNN and EDGC methods; in the first case data are referred to $k = 5$, for which the overall correct classifications were higher (71%) than those obtained using $k = 1$ (63%) and 3 (69%). In general, the classification results can be considered quite good. Two extreme situations are represented by the Soave wine, with a good result obtained with both methods (correct classifications of 84% and 91%, respectively), and the Tocai G., for which the correct classifications are not good (61% and 57%). A special case is that of the Tocai L. wine, for which the correct classification is good (86%) when the method EDGC is used and not good (61%) if the method KNN is considered. An opposite situation, albeit less evident, can be observed for the Prosecco wine. In any case, these simple methods of classification underline the tendency of wine samples to occupy zones of multidimensional space which are separated according to the types of wine.

b. Principal Component Analysis (PCA) and the SIMCA Method. Table 8 reports the cumulative percentage of the total variance explained by the first 10 principal components obtained from the whole data set (first column) and for each type of wine separately (disjoint principal components).

With regard to the overall PCA it can be noted that the first few components (e.g., three) explain only a low fraction of the total variance. This finding can partially explain the almost complete overlap of different wines on the scatter plot (not given here) of samples on the first three principal components; this overlap is in contrast with the classification results observed above. The only wine that shows a tendency to occupy a separate region of the plot is the Soave.

Recently, a new method for ranking principal components to reflect group structure has been proposed (Krzanowski, 1992). This method, based on a canonical variate criterion, is designed to overcome the problem arising from the fact that valuable information on group separation in a population can reside in components coming after the first few. Application of this method to the present data set gives the following ranking for the first seven principal components (significance of F value lower than 0.001): 1st, 13th, 6th, 3rd, 5th, 21st, 4th. However, results improve only slightly.

Considering the disjoint principal components, it can be noted that, besides the general low proportion of the total variance explained by the first few PC, the wine that features the highest values for this is Soave. This fact

could be due to the higher degree of correlation between variables for Soave than for the other wines (see Table 5).

Results of the application of the SIMCA method for modeling the wine groups can be considered a direct consequence of the characteristics of the data set as highlighted by principal component analysis. In particular, only for the three wines Soave, Prosecco, and Tocai G. has it been possible, after exclusion of the outliers according to the SIMCA criterion, to obtain a one-component model statistically significant according to the cross-validation test. It is considered that these three wines have the highest number of correlations greater than $|0.7|$ and for which the first principal component explains the highest fractions of the total variance. The SIMCA method shows the presence of four outliers (13% of the total) in the case of Soave wine, five (15%) for Prosecco, and seven (14%) for Tocai G.

Table 9 gives the loadings of the first principal component for the three significant models. It is noted that all of the aroma compounds (except THOL and CHOL) appear to have considerable importance in the principal components and that similar values of the loadings for the three wines are generally observed for these variables. Conversely, contrasting values are often observed for the other variables.

Of particular importance is the fact that variable CHOL, which features the highest univariate differentiation between the wines (see the ratios of between-group variance to total variance given in Table 4), has a very low correlation with the first principal component of each group. This characteristic, possibly due to the general low correlation of CHOL with all of the other variables, may limit the success of sample classifications based on principal components (Krzanowski, 1992).

The selectivity of models found for the three stated wines was evaluated by fitting to each model all of the samples belonging to the other types of wine. Results showed that only 3 samples (2 P and 1 V) of 160 fitted the model of Soave wine, 10 samples (2 S, 2 G, and 6 V) of 158 fitted the model of Prosecco, and 82 samples (12 S, 12 L, 29 P, and 29 V) of 141 fitted the model of Tocai G. In conclusion, the SIMCA models of Soave and Prosecco wines can be considered selective and sufficiently selective, respectively, and they can be proposed as a basis for the assessment of the membership of an unknown sample to the considered type of wine.

c. Canonical Variate Analysis (CVA). The eigenvalues associated with the four canonical variates, together with the corresponding canonical correlation coefficients, are given in Table 10. It can be seen that the first variate is by far the most important one (highest eigenvalue); the second and third are of almost equal relevance, while the fourth can be considered less important.

According to the correlation coefficients of variables with the first canonical variate, the variables that contribute primarily to the discrimination are in the order CHOL (corr. 0.87), HOL (0.70), K (0.64), MPA (0.54), MBA (0.51), MPOL (0.50), AC (0.48), MBOL (0.46), TN (-0.45), ALA (0.40). These variables, except MPOL, are the same that feature the first nine higher ratios of between-group variance to total variance (see Table 4). Variables contributing most to the second and third canonical variates are TP (corr. 0.45), TN (-0.42), Li (0.37), and Cl (0.30) and Na (-0.49), K (-0.35), Cl (-0.34), and Ca (-0.32), respectively. It can be observed that the first canonical variate is mainly correlated to aroma components, being dominated by *cis*-3-hexen-1-ol, whose importance as a wine-separating variable has been already stressed above

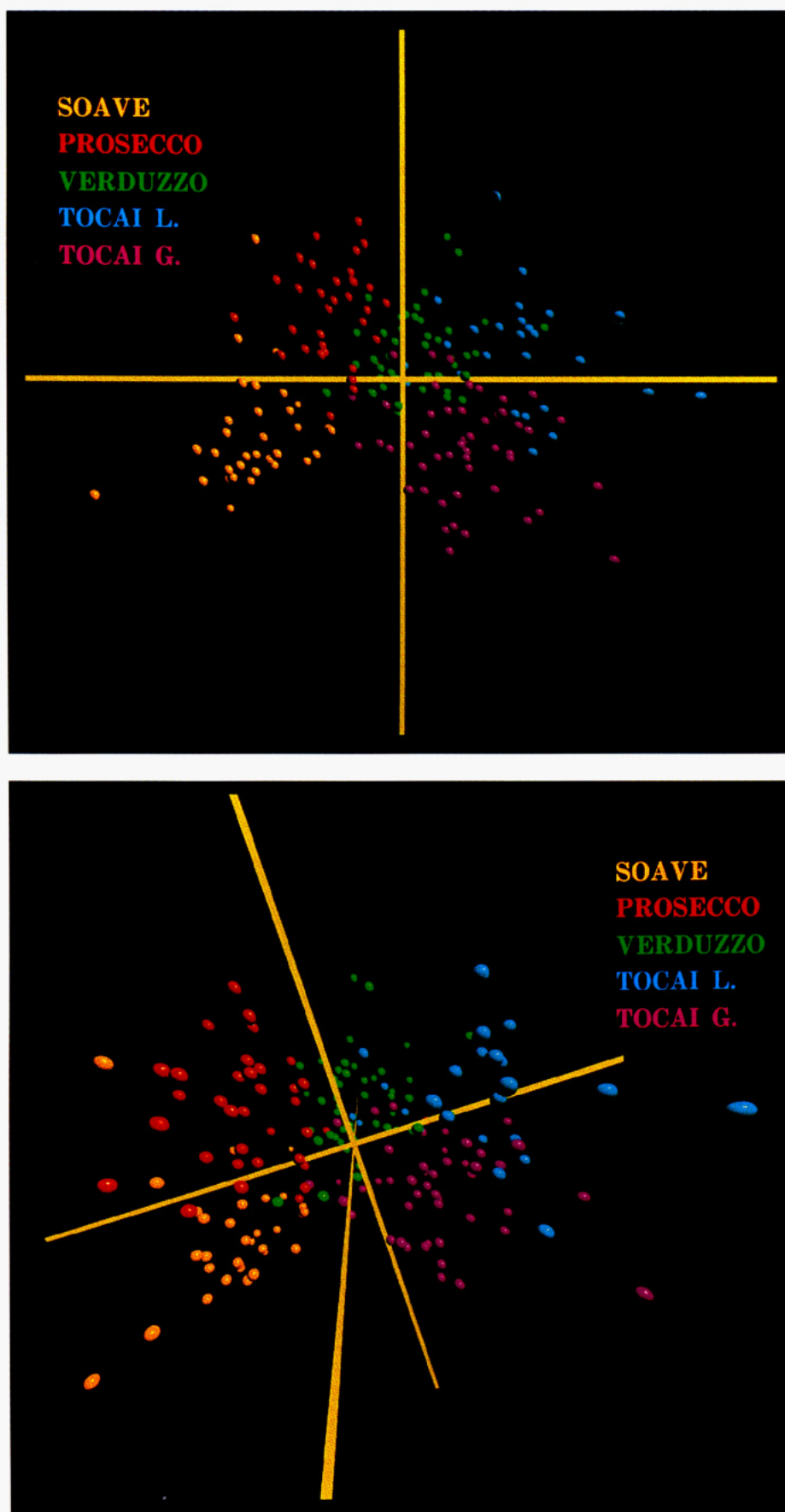


Figure 2. Scatter plot of wine samples on the first two (a, top) and on the first three (b, bottom) canonical variates (1981–1984 vintages). Horizontal axis, first CV; vertical axis, second CV.

Table 7. Classification of Wine Samples Obtained by Methods KNN and (in Parentheses) EDGC (1981–1984 Vintages)

| actual group | no. of samples | predicted group | | | | | correct rate, % |
|--------------|----------------|-----------------|----------|----------|----------|----------|-----------------|
| | | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. | |
| Soave | 32 | 27 (29) | 2 (1) | 1 (1) | 0 (1) | 2 (0) | 84 (91) |
| Prosecco | 34 | 2 (4) | 30 (26) | 1 (2) | 0 (0) | 1 (2) | 88 (76) |
| Verduzzo | 47 | 1 (1) | 6 (6) | 32 (34) | 1 (2) | 7 (4) | 68 (72) |
| Tocai L. | 28 | 0 (0) | 1 (0) | 4 (4) | 17 (24) | 6 (0) | 61 (86) |
| Tocai G. | 51 | 1 (3) | 7 (5) | 10 (9) | 2 (5) | 31 (29) | 61 (57) |
| overall | | | | | | | 71 (74) |

Table 8. Principal Component Analysis (1981–1984 Vintages)^a

| com- ponent | overall | cumulative proportion of total variation, % | | | | |
|----------------|---------|---|----------|----------|----------|----------|
| | | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. |
| 1 | 23 | 26 | 24 | 18 | 19 | 22 |
| 2 | 34 | 43 | 37 | 34 | 36 | 38 |
| 3 | 43 | 56 | 47 | 44 | 47 | 48 |
| 4 | 51 | 64 | 57 | 53 | 56 | 56 |
| 5 | 57 | 71 | 64 | 64 | 64 | 62 |
| 6 | 61 | 76 | 71 | 67 | 71 | 68 |
| 7 | 66 | 80 | 77 | 72 | 77 | 72 |
| 8 | 70 | 84 | 82 | 77 | 81 | 76 |
| 9 | 74 | 87 | 85 | 81 | 85 | 80 |
| 10 | 77 | 89 | 88 | 84 | 89 | 83 |

^a Cumulative proportion of total variation explained by the first ten components, by type of wine. The first column refers to the overall data set.

Table 9. Loadings on the First Principal Component, Obtained by the SIMCA Method (1981–1984 Vintages)

| chemical variable | wine | | |
|-------------------|--------|----------|----------|
| | Soave | Prosecco | Tocai G. |
| Na | 0.243 | -0.179 | 0.026 |
| K | 0.059 | 0.212 | 0.264 |
| Ca | 0.306 | -0.063 | 0.098 |
| Mg | 0.231 | 0.210 | 0.213 |
| Cl | -0.004 | 0.025 | 0.185 |
| pH | 0.025 | 0.261 | 0.225 |
| TA | 0.068 | -0.239 | -0.035 |
| P | 0.009 | 0.259 | 0.210 |
| AC | 0.006 | 0.270 | 0.312 |
| ALA | -0.004 | 0.197 | 0.289 |
| Li | 0.292 | -0.185 | 0.105 |
| TN | 0.028 | 0.035 | -0.049 |
| DE | 0.203 | -0.028 | 0.269 |
| A | -0.018 | 0.105 | 0.178 |
| TP | 0.126 | 0.184 | 0.199 |
| POL | 0.215 | 0.046 | 0.040 |
| MPOL | 0.289 | -0.051 | 0.217 |
| MBOL | 0.262 | 0.241 | 0.196 |
| MPA | 0.228 | -0.048 | 0.228 |
| BA | -0.218 | -0.188 | -0.194 |
| MBA | 0.237 | 0.224 | 0.202 |
| HA | -0.288 | -0.296 | -0.269 |
| OA | -0.293 | -0.306 | -0.245 |
| DA | -0.222 | -0.269 | -0.205 |
| HOL | 0.248 | 0.276 | 0.103 |
| THOL | 0.059 | 0.094 | -0.022 |
| CHOL | 0.062 | 0.014 | 0.055 |

considering the univariate distributions (see Table 3, Figure 1, and Table 4, last column). The second canonical variate is related principally to phenols and nitrogen compounds, while inorganic components mainly contribute to the third variate.

Figure 2 shows the scatter plots of samples on the first two canonical variates and on the first three variates, respectively. A good separation of groups can be observed, and this evidence is confirmed by the classification results obtained by considering the Euclidean distances of samples from the group centroids in the space defined by the four canonical variates (Table 11). The overall value of 90% for the correct classification can be considered quite

Table 10. Eigenvalues Associated with the Canonical Variates (λ) and Canonical Correlation Coefficients (ρ) (1981–1984 Vintages)

| canonical variate | λ | ρ |
|-------------------|-----------|--------|
| 1 | 3.51 | 0.88 |
| 2 | 1.31 | 0.75 |
| 3 | 1.06 | 0.72 |
| 4 | 0.56 | 0.60 |

satisfactory; best results (97%) are observed for Soave wine, while the Tocai G. presents the lowest number of samples classified correctly (84%).

An unbiased value of the actual percentage of correct classifications based on canonical variates was estimated by using the well-known method of the training set/test set partition of data (Lachenbruch, 1975). In this method the data set is split in two parts, defined as the training set and the test set; then the samples of the test set are classified on the basis of the canonical variates obtained with the training set. Twenty repetitions of random partitioning data with 90% of the cases in the training set have been carried out using the SELECT procedure of SPSS (Hull and Nie, 1981); the average correct classification rate on the test set is 79.8% (RSD 8.8%), as opposed to the value of 89.6% obtained by the procedure of classifying the same cases used for the computation of the canonical variates (apparent rate). These results confirm that the apparent value for the correct classification rate is a little optimistic with respect to the more realistic estimate obtained by the training set/test set partition method.

The peculiar behavior of the four canonical variates in discriminating the wine groups can be traced by comparing scatter plots and classification results. The first canonical variate (high correlation with CHOL, HOL, and K) efficiently separates wines Soave and Prosecco both from each other and from the remaining group of V-L-G. The second canonical variate (mainly related to TP, TN, and Li) helps to discriminate the wine Tocai G. from Verduzzo and Tocai L. (correct classification for G passes from 39% to 76%). The third (principally related to Na and, to a lesser degree, K, Cl, and Ca) and fourth (correlation with Li, Ca, and DA) variates improve the general classification, providing, in particular, better separation between Verduzzo and Tocai L. (third variate). Finally, the correct classifications for wines L and G remain lower (84–86%) than those of wines S, P, and V (91–94%).

However, it should be noted that, in spite of low classification, the two wines Tocai L. and Tocai G., which are produced from the same grape variety in different, but adjacent, geographical areas, are separated from each other. The problem of low classification of L and G is mainly due to the location of Verduzzo samples which occupy in the canonical space a neighboring position with respect to both Tocai L. and Tocai G. In fact, the samples of Tocai L. and Tocai G. which are erroneously classified are mainly included in the Verduzzo group.

Table 11. Classification of Samples by Their Euclidean Distances from the Group Centroids in the Canonical Space (1981–1984 Vintages)

| actual group | no. of samples | predicted group | | | | | correct rate, % |
|--------------------------------|----------------|-----------------|----------|----------|----------|----------|-----------------|
| | | Soave | Prosecco | Verduzzo | Tocai L. | Tocai G. | |
| First Canonical Variate | | | | | | | |
| S | 32 | 29 | 3 | | | | 90.6 |
| P | 34 | 2 | 30 | 2 | | | 88.2 |
| V | 47 | | 7 | 27 | 2 | 11 | 57.4 |
| L | 28 | | | 3 | 19 | 6 | 67.9 |
| G | 51 | | 2 | 14 | 15 | 20 | 39.2 |
| overall % | | | | | | | 65.1 |
| First Two Canonical Variates | | | | | | | |
| S | 32 | 29 | 3 | | | | 90.6 |
| P | 34 | 1 | 29 | 4 | | | 85.3 |
| V | 47 | | 5 | 35 | 4 | 3 | 74.5 |
| L | 28 | | | 7 | 19 | 2 | 67.9 |
| G | 51 | 1 | | 6 | 5 | 39 | 76.5 |
| overall % | | | | | | | 78.6 |
| First Three Canonical Variates | | | | | | | |
| S | 32 | 30 | 2 | | | | 93.8 |
| P | 34 | 1 | 32 | 1 | | | 94.1 |
| V | 47 | | 2 | 43 | | 2 | 91.5 |
| L | 28 | | | 3 | 22 | 3 | 78.6 |
| G | 51 | | 2 | 6 | 2 | 41 | 80.4 |
| overall % | | | | | | | 87.5 |
| First Four Canonical Variates | | | | | | | |
| S | 32 | 30 | 2 | | | | 93.8 |
| P | 34 | 1 | 32 | 1 | | | 94.1 |
| V | 47 | | 2 | 43 | | 2 | 91.5 |
| L | 28 | | | 4 | 24 | | 85.7 |
| G | 51 | | 3 | 5 | | 43 | 84.3 |
| overall % | | | | | | | 89.6 |

Conclusion. The chemical variables (inorganic components, classical determinations, and a number of aroma components) used to provide a quantitative and scientific characterization of the wines proved to be sufficiently stable over time (i.e., from vintage to vintage). The variables were well suited to detect and to single out good multivariate group structures for the five wines (which appear well enough separated in the multidimensional space of the chemical variables) and make it possible to characterize and to classify samples according to the origin, from a multiyear data base.

Among the chemical compounds, *cis*-3-hexen-1-ol shows a particularly high discriminating power (especially for Soave and Prosecco wines from others), coupled with good stability in time. Similar results highlighting the discrimination ability of this compound have also been reported in a chemometric investigation of four red wines of the Piemonte region of Italy (Armanino et al., 1990). In this respect it is noted that a recent investigation showed that *cis*-3-hexen-1-ol does not undergo appreciable chemical transformations during alcoholic fermentation; this fact led researchers to suggest the use of this compound in characterizing grape varieties (Herraiz et al., 1990).

With regard to the chemometric procedures adopted to classify wines, best results are obtained using canonical variate analysis, while principal component analysis and its derived method, SIMCA, show poorer efficiency in highlighting the group structure.

ACKNOWLEDGMENT

This research was supported by the Italian Ministry of Education, M.P.I. 40% and 60% grant "Chemiometria", and by the National Research Council of Italy (CNR) grant "Progetto Finalizzato Chimica Fine".

Supplementary Material Available: Summary statistics of the complete data set (vintages from 1977 to 1984), i.e., mean, median, minimum–maximum values, and standard deviation for all of the chemical variables by type of wine and vintage year, the overall statistics, and the ratio of between-group variance to total variance (27 pages). Ordering information is given on any current masthead page.

LITERATURE CITED

- Armanino, C.; Forina, M.; Castino, M.; Piracci, A.; Ubigli, M. Chemometrical investigation on four red wines from a single cultivar grown in the Piedmont region. *Analyst* **1990**, *115*, 907–910.
- Etievant, P.; Schlich, P. Varietal and geographical classification of French red wines in terms of major acids. *J. Sci. Food Agric.* **1989**, *46*, 421–438.
- Forina, M.; Lanteri, S.; Armanino, C. Chemometrics in food chemistry. In *Topics in current chemistry*; Springer-Verlag: Berlin, 1987; pp 93–143.
- Herraiz, T.; Herraiz, M.; Reglero, G.; Martin-Alvarez, P. J.; Cabezudo, M. D. Changes in the composition of alcohols and aldehydes of C₆ chain length during the alcoholic fermentation of grape must. *J. Agric. Food Chem.* **1990**, *38*, 969–972.
- Hull, C. H.; Nie, N. H. *SPSS Update 7-9*; McGraw-Hill: New York, 1981; Chapter 15, pp 292–300.
- Krzanowski, W. J. Ranking principal components to reflect group structure. *J. Chemom.* **1992**, *6*, 97–102.
- Lachenbruch, P. A. *Discriminant Analysis*; Hofner: New York, 1975; Chapter 2, p 33.
- Larrechi, M. S.; Franques, M. R.; Ferre, M.; Rius, F. X. Structure modelling and discrimination of Catalan white wines. *J. Chemom.* **1988**, *3*, 261–274.
- Martens, H.; Russwurm, H., Jr., Eds. *Food research and data analysis*; Applied Science: London, 1983.
- Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: a textbook*; Elsevier: Amsterdam, 1988.

- Moret, I.; Scarponi, G.; Capodaglio, G.; Zanin, S.; Camaiani, G.; Toniolo, A. Analytical parameters in the characterization of three Venetian wines. Application of the statistical linear discriminant analysis. *Am. J. Enol. Vitic.* **1980**, *31*, 245-249.
- Moret, I.; Scarponi, G.; Cescon, P. Multiple discriminant analysis in the analytical differentiation of Venetian white wines. 5. Influence of lithium and total nitrogen concentration in improving the classification. *Rass. Chim.* **1983**, *35*, 319-328.
- Moret, I.; Di Leo, F.; Giromini, V.; Scarponi, G. Multiple discriminant analysis in the analytical differentiation of Venetian white wines. 4. Application to several vintage years and comparison with the k-nearest neighbour classification rule. *J. Agric. Food Chem.* **1984a**, *32*, 329-333.
- Moret, I.; Scarponi, G.; Cescon, P. Aroma components as discriminating parameters in the chemometric classification of Venetian white wines. *J. Sci. Food Agric.* **1984b**, *35*, 1004-1011.
- Moret, I.; Scarponi, G.; Capodaglio, G.; Giromini, V.; Cescon, P. Multiple discriminant analysis in the analytical differentiation of Venetian white wines. 6. Simultaneous differentiation of four typical wines. *Ann. Chim. (Rome)* **1984c**, *74*, 73-85.
- Moret, I.; Capodaglio, G.; Scarponi, G.; Romanazzi, M. Statistical evaluation of the group structures of five Venetian wines from chemical measurements. *Anal. Chim. Acta* **1986**, *191*, 331-350.
- Moret, I.; Scarponi, G.; Capodaglio, G. Characterization of a typical wine by a SIMCA analysis of aroma composition data. *Ann. Chim. (Rome)* **1988**, *78*, 285-297.
- PARVUS. Elsevier: Amsterdam, The Netherlands, 1988.
- SAS. SAS Institute, Cary, NC, 1985.
- Scarponi, G.; Moret, I.; Capodaglio, G.; Cescon, P. Multiple discriminant analysis in the analytical differentiation of Venetian wines. 3. A reelaboration with addition of data from samples of 1979 vintage Prosecco wine. *J. Agric. Food Chem.* **1982**, *30*, 1135-1140.
- Scarponi, G.; Moret, I.; Capodaglio, G.; Romanazzi, M. Cross-validation, influential observations and selection of variables in chemometric studies of wines by principal components analysis. *J. Chemom.* **1990**, *4*, 217-240.
- Seeber, R.; Sferlazzo, G.; Leardi, R.; Dalla Serra, A.; Versini, G. Multivariate data analysis in classification of musts and wines of the same variety according to vintage year. *J. Agric. Food Chem.* **1991**, *39*, 1764-1769.
- Shimoda, M.; Shibamoto, T.; Noble, A. C. Evaluation of headspace volatiles of Cabernet Sauvignon wines sampled by an on-column method. *J. Agric. Food Chem.* **1993**, *41*, 1664-1668.
- STATGRAPHICS. Manugistics, Inc., Rockville, MD, 1992.
- Varmuza, K. *Pattern recognition in chemistry*; Springer-Verlag: Berlin, 1980.
- Vasconcelos, A. M. P.; Chaves des Neves, H. J. Characterization of elementary wines of vitis vinifera varieties by pattern recognition of free amino acid profiles. *J. Agric. Food Chem.* **1989**, *37*, 931-937.
- Vogel, J. T. W. E.; Tas, A. C.; van den Berg, F.; van der Greef, J. A new method for classification of wines based on proton and carbon-13 NMR spectroscopy in combination with pattern recognition techniques. *Chemom. Intell. Lab. Syst.* **1993**, *21*, 249-258.
- Wold, S.; Sjostrom, M. SIMCA: a method for analyzing chemical data in terms of similarity and analogy. In *Chemometrics: theory and application*; Kowalski, B. R., Ed.; American Chemical Society: Washington, DC, 1977; pp 243-282.

Received for review August 9, 1993. Revised manuscript received January 4, 1994. Accepted February 7, 1994.*

* Abstract published in *Advance ACS Abstracts*, March 1, 1994.