ELSEVIER

Journal of Chromatography A, 943 (2001) 123–137

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Classification of Nebbiolo-based wines from Piedmont (Italy) by means of solid-phase microextraction–gas chromatography–mass spectrometry of volatile compounds

Emilio Marengo[a], Maurizio Aceto[a,*], Valter Maurino[b]

[a]*Department of Sciences and Advanced Technologies, University of East Piedmont, corso Borsalino, 54, 15100 Alessandria, Italy*
[b]*Department of Analytical Chemistry, University of Turin, via Giuria, 5, 10125 Turin, Italy*

Received 16 July 2001; received in revised form 18 October 2001; accepted 19 October 2001

Abstract

Sixty-eight samples of wines from Piedmont (Italy) were analysed to determine their content of volatile compounds, using the solid-phase microextraction (SPME) technique coupled with gas chromatography–mass spectrometry (GC–MS). Samples were from five groups of wines: Barolo, Barbaresco, Nebbiolo d'Alba, Roero and Langhe Nebbiolo, all produced from the Nebbiolo grape in the Langhe and Roero areas (province of Cuneo, Piedmont) but differing in vintage (respectively, 3 years, 2 years, 1 year, 8 months and few months) and production zone. Thirty-five analytes were identified; peak area data, corrected for internal standard, were used for pattern recognition treatments. Principal components analysis, hierarchical cluster analysis, Kohonen self organising map, stepwise linear discriminant analysis and soft independent modelling of class analogy were applied to the data, revealing a good separation between the five groups. A main factor, strictly connected to wine vintage, was identified and found to be related to some analytes. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Wine; Statistical methods; Volatile compounds

1. Introduction

The possibility of recognizing a wine from its chemical parameters has received great attention in the last few years [1–3]. Many authors have proposed methods to discriminate wines, determining different compounds as markers. Classes of parameters utilised comprised elements [4–6], organic compounds [7–9], stable isotopes [10–12] between others. Frequently, authors suggested an integrated use of different chemical data [13–16].

As a general rule, it is noteworthy to say that every classification scheme is a story by itself, because variables used to obtain a good discrimination in a certain case may not be reliable in a different scheme. This is due to the complexity and heterogeneity of the wine matrix, which can reflect different peculiarities of soil, grape or production cycle. This is why it is rather difficult to separate variables for general use. Martin et al. [12] proposed stable isotopes analysis with NMR.

A classification study on wines from Piedmont (Italy) is particularly important because of the great value of some high-level products of this region, among them wines made from the Nebbiolo grape. Forina et al. [17] studied a classification scheme

*Corresponding author. Tel.: +39-0131-283809; fax: +39-0131-254410.

E-mail address: aceto@ch.unito.it (M. Aceto).

0021-9673/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0021-9673(01)01421-2

based on 35 samples of Barbaresco, Barolo, Gattinara and Nebbiolo d'Alba, utilising variables of different origin to yield a good classification model; they found that total condensed flavonoids, saturation (a colour measurement), *cis*-hex-3-enol and β-hydroxyethylbenzene were particularly useful as variables to discriminate between the four classes examined.

In our work, five classes of samples were considered: Barbaresco, Barolo, Langhe Nebbiolo, Nebbiolo d'Alba and Roero. These wines are produced in a relatively small area in the province of Cuneo, Piedmont. The main feature of this set of classes is that they differ in their vintage: 3 years for Barolo, 2 for Barbaresco, 1 for Nebbiolo d'Alba, 8 months for Roero and few months for Langhe Nebbiolo. This feature is reflected markedly on the value of these wines and this is why it is useful to develop a method for their correct attribution.

We chose to determine volatile compounds in these samples, because they are probably among those responsible for the so-called *bouquet*, that is the characteristic fragrance of wine, a feature that could be greatly influenced by vintage. These compounds are expected to be mainly esters, terpenes and lower alcohols. Analyses were performed with gas chromatography coupled with mass spectrometry (GC–MS) after collection of analytes by headspace solid-phase microextraction (SPME), a technique developed by Pawliszyn in 1989 [18] which presents many advantages: speed, ease of use, and selectivity towards different classes of compounds. Some works dealing with wine characterisation by SPME coupled with GC–MS and GC–FID are present in the literature [19–21]. However, none deal with the varieties of wine considered in this work.

2. Conditions

2.1. Chemicals

All reagents were of analytical grade or better. Volatile organic compounds, as well as the internal standard 4-chloro-2-methylphenol, were purchased from Sigma–Aldrich (St Louis, MO) and used as received.

Standard solutions were prepared by dilution of the proper amount of analyte in ethanol 95% (Merck, Darmstadt, Germany).

2.2. Equipment

Analyses were performed with an HP5890 gas chromatograph with mass spectrometric detection (HP5972, Agilent Technologies). The GC system was equipped with a 30 m HP-VOC capillary column (0.25 mm I.D., 1.5 μm film thickness). The split/splitless inlet was equipped with a 2 mm I.D. deactivated glass liner designed for SPME analysis (Supelco, Bellefonte, PA) and an electronic flow control. GC conditions were as follows: inlet at 250°C; splitless time: 3 min; oven initial temperature: 40°C (5 min) then to 250°C at 8°C/min; GC–MS transfer line at 250°C; carrier He (Extrapure, 99.9999%) at 1 ml/min. Peak areas were computed using extracted ion currents at m/z corresponding to the most abundant fragment in the mass spectrum of the analyte considered.

2.3. SPME conditions

Wine samples were extracted with headspace-SPME, the best SPME technique when target analytes are volatile organic compounds [18]. The following parameters were optimised: fiber type (either polydimethylsiloxane, PDMS, or polyacrylate, PA, from Supelco, Bellefonte, PA), contact time between fiber and headspace, salt addition (to increase volatile fraction of analytes via salting-out effect), sample temperature, sample volume and stirring. Optimal conditions of extraction were obtained using the following procedure: 20.0 ml of wine were transferred to a 25-ml vial (headspace volume was 5 cm^3), with 2 g of Na_2SO_4 added and 10 μl of an ethanolic solution (3000 mg/l) of 4-chloro-2-methylphenol as internal standard (1.5 mg/l concentration in the sample). The vial was sealed and headspace extraction was performed for 20 min at 25°C with a 100-μm PDMS fiber, keeping the sample under continuous stirring. The fiber was then transferred into the GC–MS inlet. The utilisation of the PA fiber resulted in moderately better recoveries for low molecular mass alcohols, but definitely poorer recoveries for esters and other low polarity compounds. Temperatures higher than 25°C resulted

in lower recoveries, due to increased partition of volatile analytes to the headspace. We did not check recoveries at lower temperatures.

2.4. Statistical methods

2.4.1. Principal components analysis (PCA)

In this well-known method [22,23], the principal components (PCs), i.e. new variables obtained as linear combinations of the original ones, orthogonal to each other, are calculated in such a way as to keep most of the information present in the original data set in the least possible number of new variables. The PCs can be plotted for visual inspections of the data to point out patterns hidden in the dataset.

2.4.2. Hierarchical cluster analysis (HCA)

This is another unsupervised method [24,25], in which samples are considered as lying in a p-dimensional hyperspace (with p the number of variables) and distances between samples are calculated, joining objects with an agglomerative procedure until all samples are joined to form a single cluster.

2.4.3. Kohonen self organising map (Kohonen SOM)

The Kohonen artificial neural network (K-ANN) or self organising map (SOM) [26–28] is based on a single layer of neurons, usually arranged in a square box, with its responses on the top of the box. Below each cell of the top layer there is a column of cells, one for each descriptor, which represent the weight of the network. The top layer is characterised by the number of cells on each edge. During the learning of the network, each sample is presented to the network. For each sample, the distance between the sample and every column of weights is calculated. The column with the minimum distance is considered the winning neuron. The weights of this neuron are modified so that at the following cycle the distance of the same sample from the winning neuron shall be smaller. A similar correction is applied to the neurons in the neighbourhood of the winner. This correction decreases with the distance. Usually the distance at which the correction takes place decreases during the learning phase. At the beginning the entire network is affected by every correction while in the last cycles only the winner neuron is corrected. Similarly at the beginning the learning rate, i.e. the amount of correction introduced is larger than in the last cycles. The final result is a map, the first layer, where the most similar samples are in the same cell or next to one another. The weights give an insight into the reason for the clustering of the objects. So the analysis of the first layer provides information on the similarity of the samples while the analysis of the weights provides information on the reason for their similarity.

2.4.4. Soft independent modelling of class analogy (SIMCA)

In this classification method [29,30] each class is described by an independent principal component analysis model. New samples are classified on the basis of their fit with the different PCA models. The optimal number of PCs for each model is chosen independently since the classes may exhibit different shapes and structures. For new samples the residuals and scores with respect to each PCA model are calculated. The residuals provide information on the ability of each model to describe the new data, a sort of object-to-model distance while the scores can be combined in a measure of the distance between the object and the model centre.

2.4.5. Stepwise linear discriminant analysis (SLDA)

In this supervised pattern recognition method [31,32], a classification model is built applying a forward step-wise variable selection algorithm. The most discriminating variables are selected on the basis of a Fisher's test.

3. Result and discussion

The SPME–GC–MS method developed was found to be fully suitable for the analysis of volatile compounds in wine, due to its selectivity and sensitivity. Detection limits are in the ng/l range; repeatability as calculated on five successive extraction cycles, is about 5% for all analytes considered.

Fig. 1 shows a typical chromatogram (total ion current) obtained from headspace-SPME–GC–MS analysis of a wine sample. Careful inspection of the

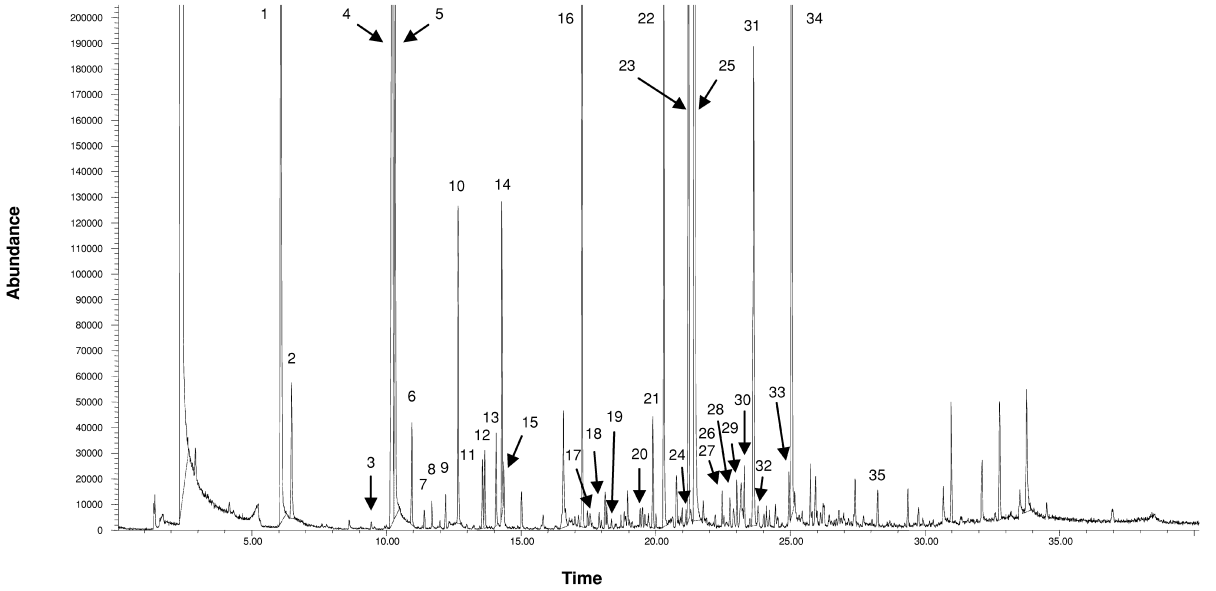

Fig. 1. Typical chromatogram (total ion current) obtained from headspace-SPME–GC–MS analysis of a ''Barolo'' sample. Identified analytes are listed in Table 1.

chromatogram allowed the assignment of a mass spectrum clearly different from background for more than 50 peaks. Thirty-five of them (listed in Table 1 in elution order) were unambiguously identified either by library search (match quality better than 90%) or by injection of authentic standards. These

Table 1
Compounds detected in wine samples

No.	Analyte	t_r		No.	Analyte	t_r	
1	Ethyl acetate	6.1	E	19	Ethyl 2-hexenoate	18.4	E
2	2-Methyl-1-propanol	6.5	A	20	Ethyl heptanoate	19.4	E
3	Ethyl propanoate	9.5	E	21	Methyl octanoate	19.9	E
4	3-Methyl-1-butanol	10.2	A	22	2-Phenylethanol	20.3	A
5	2-Methyl-1-butanol	10.3	A	23	4-Ethylphenol	21.1	A
6	Ethyl 2-methylpropanoate	10.9	E	24	Diethyl succinate	21.2	E
7	2-Methylpropyl acetate	11.4	E	25	Ethyl octanoate	21.5	E
8	2,3-Butanediol	11.7	A	26	3-Methyl-1-butyl hexanoate	22.5	E
9	Ethyl butanoate	12.2	E	27	2-Methyl-1-butyl hexanoate	22.6	E
10	Ethyl lactate	12.7	E	28	4-Ethylphenyl acetate	22.8	E
11	Ethyl 2-methylbutanoate	13.6	E	29	2-Phenylethyl acetate	23.0	E
12	Ethyl 3-methylbutanoate	13.7	E	30	Ethyl nonanoate	23.3	E
13	Hexanol	14.1	A	31	α-Ionone	23.7	T
14	3-Methyl-1-butyl acetate	14.3	E	32	Methyl decanoate	23.8	E
15	2-Methyl-1-butyl acetate	14.4	E	33	Ethyl 9-decenoate	25.0	E
16	Ethyl hexanoate	17.3	E	34	Ethyl decanoate	25.1	E
17	Hexyl acetate	17.6	E	35	Ethyl dodecanoate	28.3	E
18	Cymene	18.1	T				

E, ester; A, alcohol; T, terpene.

Table 2
Variance explained by the first PCs

PC	Eigenvalue	Explained var. (%)	Cumulative var. expl. (%)
1	9.17	26.2	26.2
2	6.20	17.7	43.9
3	4.14	11.8	55.8
4	1.78	5.1	60.9
5	1.70	4.9	65.8

analytes were utilised for statistical treatments, thus obtaining a data matrix of 68 samples described by 35 variables.

3.1. Chemometrical procedures

We used five pattern recognition methods: three unsupervised, that is principal components analysis (PCA), hierarchical cluster analysis (HCA) and Kohonen self organising map (SOM or K-ANN), and two supervised, that is stepwise linear discriminant analysis (SLDA) and soft independent modelling of class analogy (SIMCA). All statistical treatments were performed on peak area data, normalised by internal standard. The data were autoscaled for PCA, HCA and SLDA, autoscaled with respect to each class for SIMCA while they were range scaled for SOM. Due to the homogeneity of our data and to the use of an internal standard with all the samples, it was not necessary to perform quantitative determination of the analytes.

3.2. Principal components analysis (PCA)

Applied to our data set, PCA revealed that the first three principal components explain 55.8% of total variance (Table 2). A plot of the scores of PC1 versus PC2, i.e. the projections of the samples along the directions identified by the first two PCs, is reported in Fig. 2. It is apparent that samples are grouped in a way similar to our classification. Moreover, an interesting feature can be pointed out: such groups are arranged along the direction of PC1, in a way that reflects their vintage order (more precisely, in reverse vintage order), that is, samples of older wines (Barolo, Barbaresco) have lower scores and samples of younger wines (Nebbiolo d'Alba, Roero, Langhe Nebbiolo) have higher scores on PC1. It can be concluded that the first principal component describes a feature strictly related to vintage. The loadings, i.e. the coefficients that define

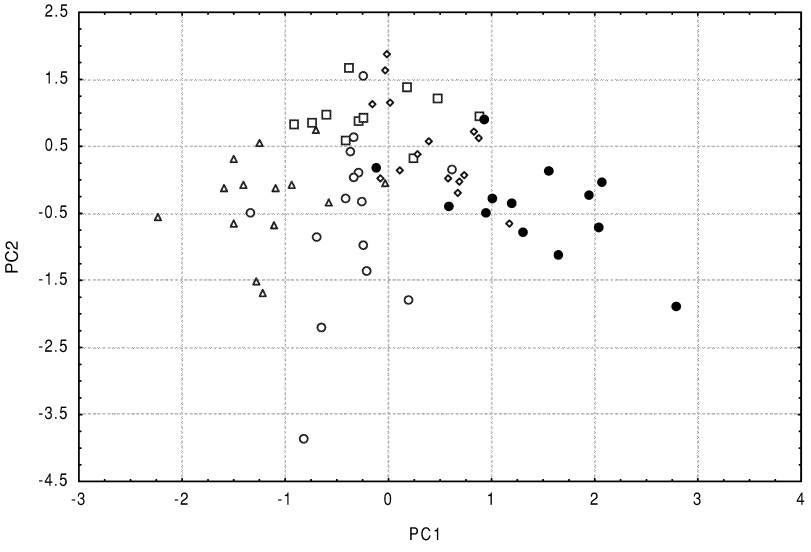

Fig. 2. Plot of Principal Component 1 versus Principal Component 2 scores. ○, Barbaresco; △, Barolo; ◇, Langhe Nebbiolo; □, Nebbiolo d'Alba; ●, Roero.

the weight of each original variable in the PCs, can then be investigated to understand which chemical compounds are responsible for the ranking of samples on the basis of the vintage on the first PC. A comparison of scores and loadings for PC1 allows the identification of the compounds having higher influence on this feature (Table 3); analytes with highly positive loadings on PC1 should be regarded as compounds whose concentration decreases with vintage, while analytes with highly negative loadings tend to increase. The most important compounds are all esters: ethyl 2-methylpropanoate, 3-methylbutyl hexanoate, ethyl decanoate, diethyl succinate, ethyl 3-methylbutanoate and ethyl octanoate. This assumption is confirmed by literature, if we consider the following:

(1) Esters formed by ethanol reaction with the major wine acids (tartaric, malic, lactic, succinic, etc.), also called *acid esters*, tend to increase with vintage, though esterification rates under wine conditions are very low, with lactic and succinic acids having the highest rates and tartaric acid the lowest [33].

(2) Esters formed through fermentation processes, or *neutral esters*, tend to decrease by hydrolysis because, in young wines, they are produced in concentrations higher than could be maintained in equilibrium conditions [33]. Non-ethyl esters, ethyl hexanoate, octanoate and decanoate behave this way.

Also the second principal component seems to contain information on the difference between the five wines. Barbaresco is at the leftmost end of this variable, moving towards the right direction we have Barolo and Roero, then Langhe Nebbiolo and Nebbiolo d'Alba. The most important variables in this case are 2-methylpropanol, 3-methylbutanol, 2-methylpropyl acetate and methyl octanoate.

No further indication can be drawn from the other principal components.

Table 3
Analytes with higher loadings on PC1, either positive or negative

Compounds	Loadings on PC1
Increasing with vintage	
Ethyl acetate	−0.77
Ethyl 2-methylpropanoate	−0.85
Ethyl lactate	−0.74
Ethyl 2-methylbutanoate	−0.70
Ethyl 3-methylbutanoate	−0.82
Ethyl succinate	−0.83
α-Ionone	−0.75
Decreasing with vintage	
Ethyl hexanoate	0.74
Ethyl octanoate	0.80
Ethyl decanoate	0.83
Hexyl acetate	0.78
3-Methylbutyl hexanoate	0.84

3.3. Hierarchical cluster analysis (HCA)

To perform cluster analysis we used Ward's method of agglomeration and Euclidean distances to evaluate similarity between samples. Applying HCA to our data set, we obtained the dendrogram reported in Fig. 3. Three main clusters can be identified: from top to bottom, a group containing nearly all Roero and most of Langhe Nebbiolo samples, a second group with nearly all samples of Nebbiolo d'Alba and samples from other groups, and a final group with nearly all Barolo and Barbaresco. Again, samples seem to cluster on the basis of vintage; this confirms the indications from PCA.

By using a transposed data matrix, it is possible to cluster variables, to see whether correlation occurs. Applying HCA to the variables, we found a pattern confirming the PCA results: variables previously described as positively correlated to PC1 tend to cluster together, far from those negatively correlated to PC1. This suggests that the data set can be simplified by eliminating redundant information; a few variables could be determined to point out differences between groups. This affirmation should be confirmed by supervised methods.

3.4. Kohonen self organising map (Kohonen SOM)

In the present case, several architectures of the network were investigated. The one providing the best results was a 7×7 network. A total of 500 iterations were used, with the learning rate decreasing linearly from 0.5 to 0.01. In the same time the range of the weights correction decreased from 7 to 1. The final result is reported in Fig. 4 and Table 4. The result shows a rather good separation of the different wines. Barbaresco and Barolo samples are

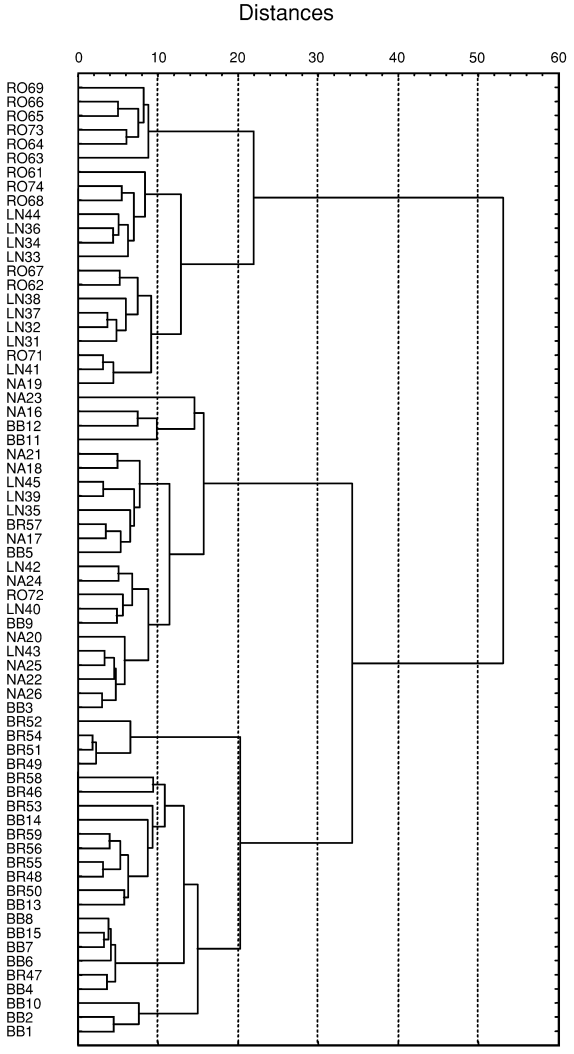

Fig. 3. Dendrogram of cases obtained with hierarchical cluster analysis on the 68 samples.

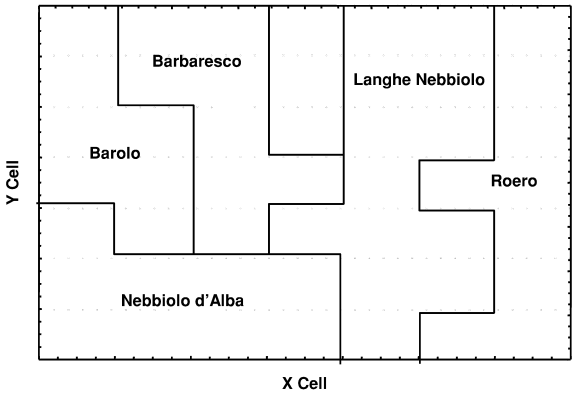

Fig. 4. Kohonen network first layer.

next to each other, as expected on the basis of the previous results. The analysis of the surface plots showing the weights (Fig. 5) confirms that these two wines are characterised by a larger amount of the acid esters. The characterisation of the other wines is also interesting, which is based on several chemical variables. There are some typical behaviours of the surfaces representing the weights:

(1) Ethyl acetate, ethyl 2-methylpropanoate, ethyl lactate, ethyl 2-methylbutanoate, ethyl 3-methylbutanoate, ethyl heptanoate, diethyl succinate, α-ionone (high for Barbaresco and Barolo, low for Roero, intermediate for the others).

(2) 2-Methylpropanol, 2-methylbutanol, 3-methylbutanol, 2-methylpropyl acetate, ethyl propanoate, 1-hexanol, 2-methylbutyl acetate, phenylethanol, 2-methylbutyl hexanoate (high for Barbaresco, Barolo and Roero, low for Langhe Nebbiolo and Nebbiolo d'Alba).

(3) Ethyl butanoate, ethyl hexanoate, hexyl acetate, ethyl 2-esenoate, methyl octanoate, ethylphenol, ethyl octanoate, 3-methylbutyl hexanoate, 2-ethylphenyl acetate, methyl decanoate, ethyl 9-decenoate, ethyl decanoate, ethyl dodecanoate (high for Roero and Langhe Nebbiolo, low for Barbaresco, Barolo and Nebbiolo d'Alba).

The power of the Kohonen network to group the objects shows its efficiency in this example.

3.5. Soft independent modelling of class analogy (SIMCA)

In the present case, a PCA model with two PCs for each class was selected by cross-validation of the classification model. The final Non Error Rate (NER%) is 88.24 while the leave-one-out (LOO) NER% is 66.18. No better LOO NER% can be obtained by increasing the number of PCs in any class model. The corresponding classification matrices (fitting and cross-validated) are reported in Table 5. The main problems of misclassification in the cross-validated models take place between Langhe Nebbiolo assigned to Roero, Nebbiolo d'Alba to Roero, Barolo to Barbaresco and Barbaresco to

Table 4
Sample location in the first layer of the self organising Kohonen map

Sample	X-Cell	Y-Cell	Sample	X-Cell	Y-Cell	Sample	X-Cell	Y-Cell
BB-01	2	7	NA-24	5	6	BR-47	1	4
BB-02	3	7	NA-25	1	1	BR-48	1	5
BB-03	1	1	NA-26	1	1	BR-49	1	6
BB-04	3	4	LN-27	4	3	BR-50	1	4
BB-05	3	3	LN-28	5	2	BR-51	2	5
BB-06	3	4	LN-29	6	2	BR-52	3	6
BB-07	3	5	LN-30	5	5	BR-53	2	3
BB-08	4	4	LN-31	3	2	BR-54	1	6
BB-09	1	1	LN-32	5	5	BR-55	2	5
BB-10	1	7	LN-33	5	1	RO-56	7	1
BB-11	3	7	LN-34	5	2	RO-57	6	1
BB-12	5	7	LN-35	3	1	RO-58	7	3
BB-13	2	5	LN-36	5	6	RO-59	7	7
BB-14	3	6	LN-37	4	1	RO-60	7	5
BB-15	3	4	LN-38	5	7	RO-61	7	4
NA-16	4	2	LN-39	1	2	RO-62	6	1
NA-17	3	2	LN-40	5	3	RO-63	6	4
NA18	2	2	LN-41	2	1	RO-64	7	7
NA-19	4	1	BR-42	1	7	RO-65	4	1
NA-20	2	1	BR-43	4	5	RO-66	4	7
NA-21	1	2	BR-44	1	5	RO-67	7	2
NA-22	1	1	BR-45	1	4	RO-68	6	4
NA-23	1	3	BR-46	1	7			

BB, Barbaresco; BR, Barolo; LN, Langhe Nebbiolo; NA, Nebbiolo d'Alba; RO, Roero.

Barolo. These misclassifications do not confound expensive wines with the cheaper ones. The sensitivity, i.e. the non-error rate per class is satisfactory for all classes except Langhe Nebbiolo, which is heavily misclassified as Roero. The cross-validated result is very poor for Langhe Nebbiolo, but also Nebbiolo d'Alba shows a low sensitivity, again confounded with Roero. As a consequence, Roero is characterised by a low specificity, i.e. purity of the class both in fitting and cross-validation. In cross-validation also Nebbiolo d'Alba, Barolo and Barbaresco have lower specificities, the latter because of the exchange of samples between the two classes.

All variables exhibit a modelling power greater than 80%. The variables with the largest discriminating power are: ethyl 9-decenoate, ethyl 2-methylpropanoate, 3-ethylphenyl acetate, 3-methylbutyl acetate, ethyl 2-esenoate, ethyl propanoate, 2-methylbutyl acetate.

3.6. Stepwise linear discriminant analysis (SLDA)

Two calculations were performed with $F_{\text{to-enter}} = 2.0$ and $F_{\text{to-enter}} = 4.0$. The results of the calculations are summarised in Table 6. The overall Non Error Rate (NER%) for the first model is 88.2, the Leave-One-Out cross-validated one is 72.1. From the classification matrix, it can be noted that most of the errors correspond to assignments of Nebbiolo d'Alba samples to Barbaresco and Barbaresco samples to Nebbiolo d'Alba. The classification of a younger wine in the classes of Barolo and Barbaresco takes place only once (in the class of Barbaresco). This is a good result, since Nebbiolo d'Alba and Barbaresco are much more expensive wines and it appears easy to find possible adulteration. The linear classification model which provides these results contains 11 original variables (ethyl 2-methylpropanoate, 2-methylbutanol, ethyl lactate, ethyl 2-esenoate, ethyl decanoate, 3-methylbutyl acetate, ethyl 3-methylbutanoate, 2,3-butanediol, 2-methylbutyl acetate, 2-methylpropyl acetate, hexyl acetate).

A simpler discriminant model is obtained using the greater value of the $F_{\text{to-enter}}$. In this calculation only the most discriminant variables are maintained. The best model in this case contains only five

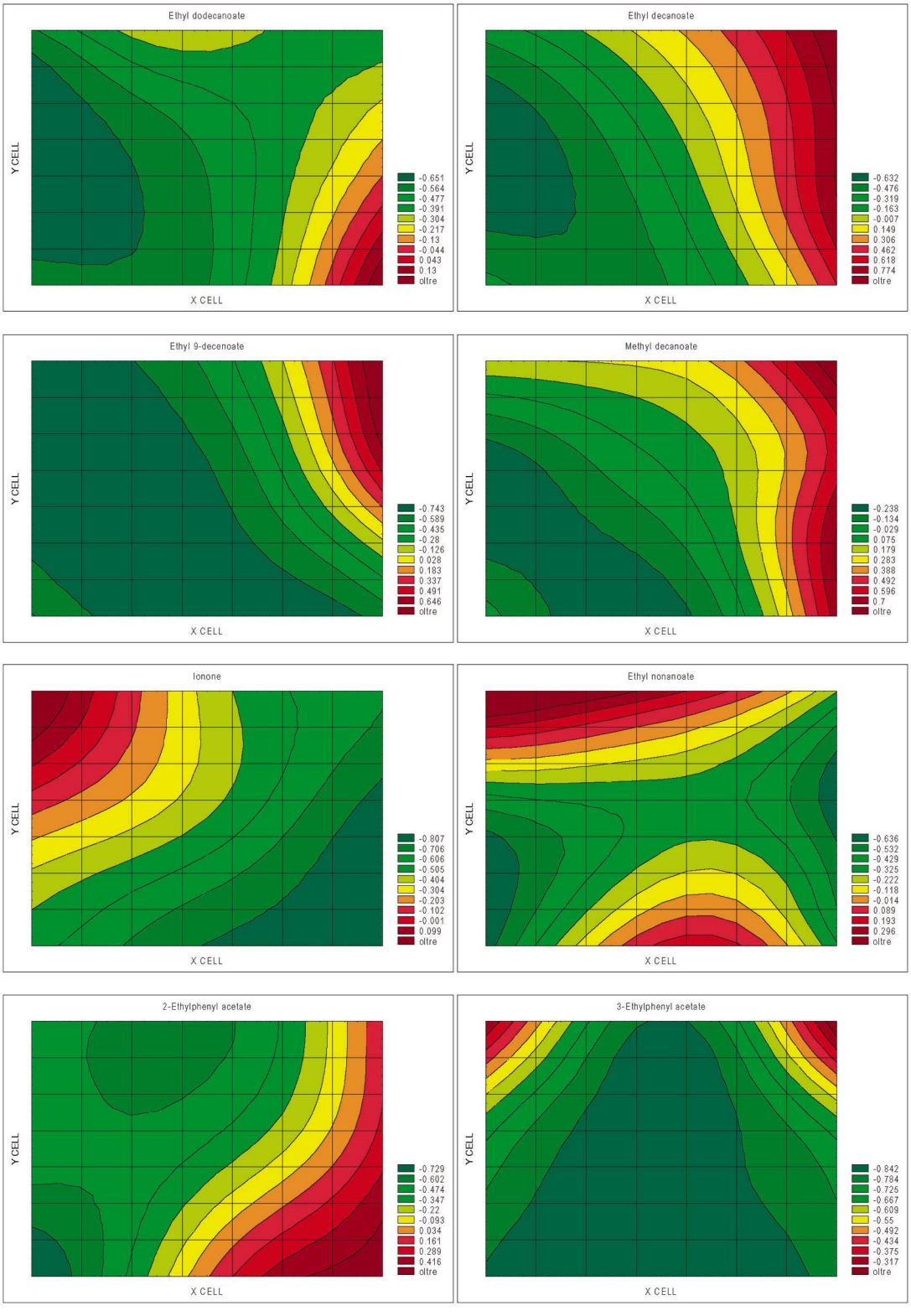

Fig. 5. Surface plots of Kohonen weights.

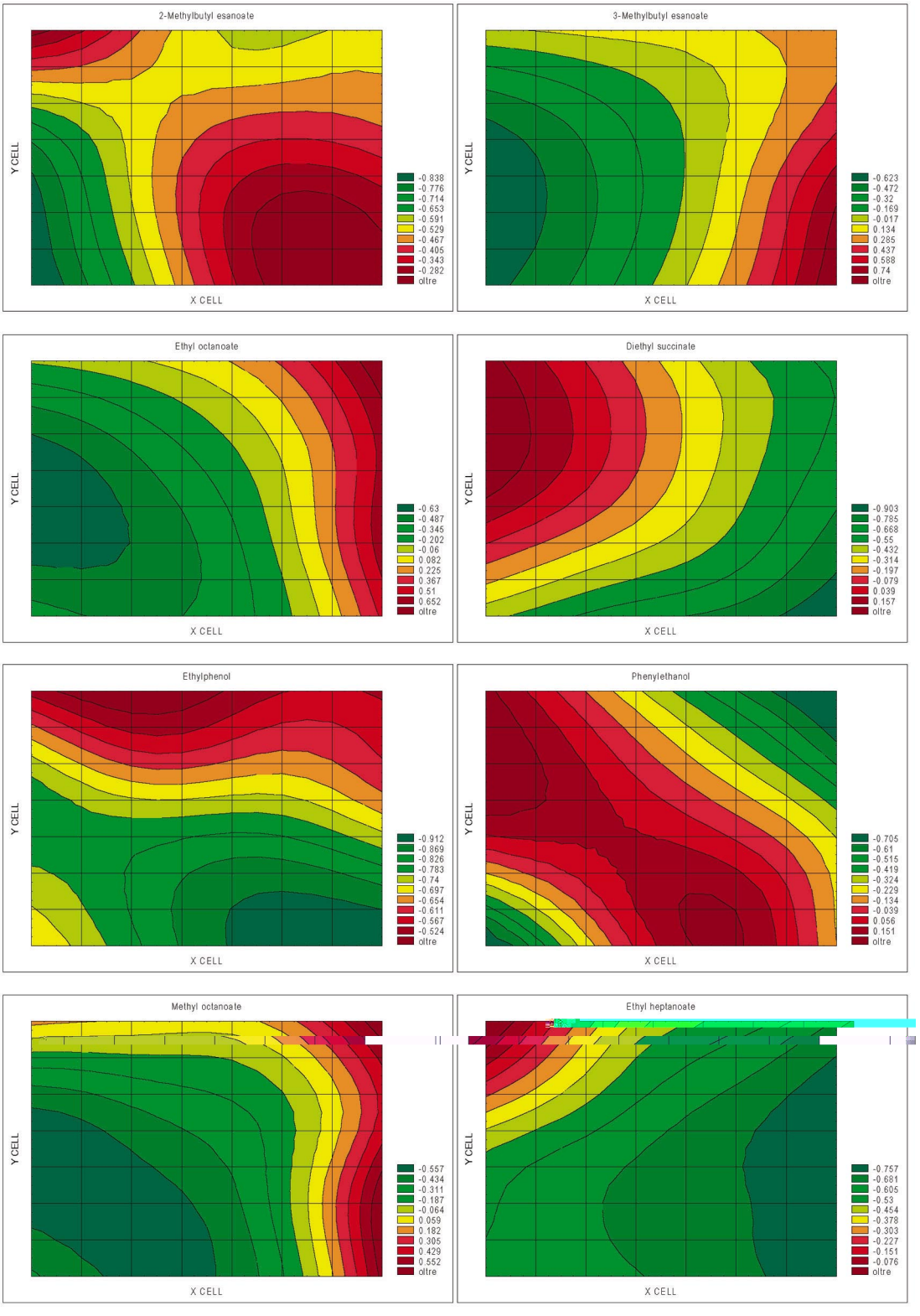

Fig. 5. (*continued*)

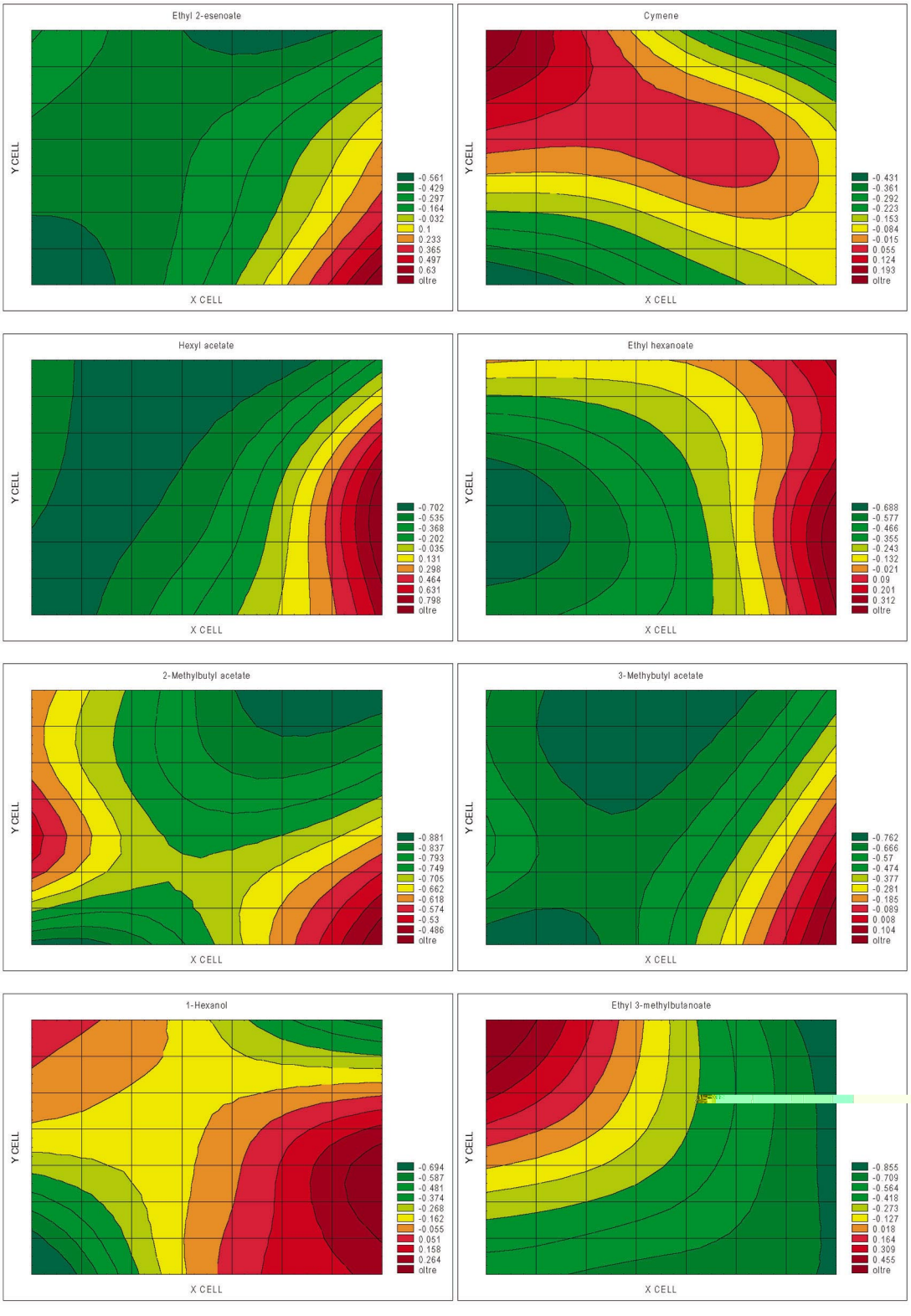

Fig. 5. (*continued*)

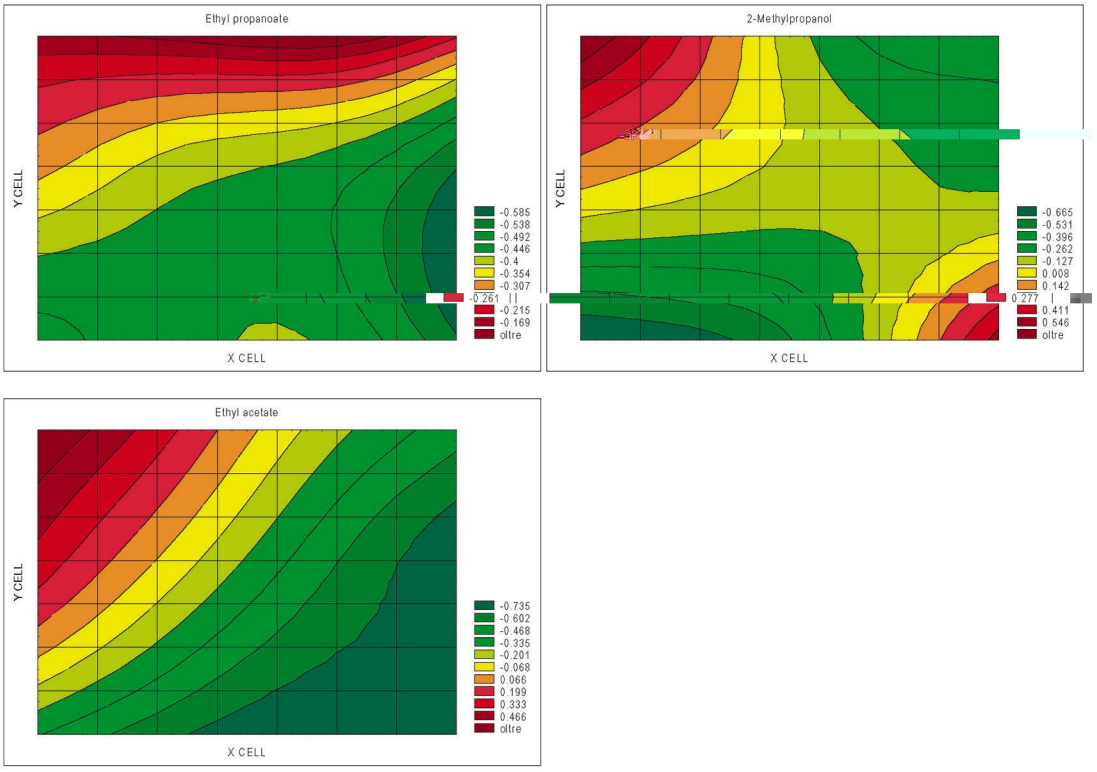

Fig. 5. (*continued*)

original variables, which, with the exception of ethyl 2-esenoate, are the same as those most relevant in PC1 (ethyl 2-methylpropanoate, ethyl 2-esenoate, ethyl decanoate, ethyl 3-methylbutanoate, hexyl acetate). This confirms the ability of PCA to identify the systematic information connected with the differences between the wine classes. Of course the quality of the latter discriminant model is slightly worse than the previous one. The overall NER% is 77.9, the Leave-One-Out cross-validated one is 73.5. The cross-validated NER% of the reduced model is slightly better than the ones for the expanded classification model, which means that the predictive ability of the former one is better. Then a slight overfitting must be present in the more complex model. Again most errors correspond to misclassifications between Barolo and Barbaresco samples, reflected in their values of sensitivity and specificity.

Table 5
Fitting and cross-validated SIMCA classification matrices

	BB	NA	LN	BR	RO	Sens.
Fitting						
BB	13	1	0	1	0	86.7
NA	0	10	0	0	1	90.9
LN	0	1	10	0	4	66.7
BR	0	0	0	14	0	100
RO	0	0	0	0	13	100
Spec.	100	83.3	100	93.3	72.2	
Cross-validated						
BB	12	1	0	2	0	80
NA	1	6	0	0	4	54.4
LN	0	2	4	0	9	26.7
BR	3	0	0	11	0	78.5
RO	1	0	0	0	12	92.3
Spec.	70.6	66.7	100	84.6	48	

The reduced model shows satisfactory values of sensitivity and specificity for all classes. The difference between the fitting and the cross-validated results is not very large, which demonstrates that only true information has been retained.

The SLDA model with five variables performs

Table 6
Fitting and cross-validated SLDA classification matrices ($F_{\text{to enter}}$ = 2.0 and 4.0)

	Fitting						Cross-validated					
	BB	NA	LN	BR	RO	Sens.	BB	NA	LN	BR	RO	Sens.
F = 2.0												
BB	14	1	0	1	0	93.3	11	2	0	2	0	73.3
NA	1	9	1	0	0	81.8	2	7	1	0	1	63.6
LN	0	1	14	0	0	93.3	0	1	13	0	1	86.7
BR	2	0	0	12	0	85.7	4	1	0	9	0	64.3
RO	0	1	1	0	11	84.6	0	1	3	0	9	69.2
Spec.	82.3	75	87.5	92.3	100		64.7	58.3	76.5	81.8	81.8	
F = 4.0												
BB	10	1	2	2	0	66.7	10	1	2	2	0	66.7
NA	1	9	1	0	0	81.8	1	7	1	1	1	63.6
LN	0	1	14	0	0	93.3	0	1	14	0	0	93.3
BR	3	1	0	10	0	71.4	4	1	0	9	0	64.3
RO	1	0	2	0	10	76.9	1	0	2	0	10	76.9
Spec.	66.7	75	73.7	83.3	100		62.5	70	73.7	75	90.9	

better than the SIMCA model and should be preferred for classification purposes.

4. Conclusions

Wines considered in this classification scheme can be easily discriminated by SPME/GC–MS determination of volatile compounds. Samples subjected to longer vintage show typical distributions different from younger wines; this makes it possible to recognise possible adulteration. The goal of our work, i.e. the chemical characterization of the different wines, was obtained both by supervised and unsupervised chemometrical methods. Some of the wines could not be completely discriminated, due to the real similarity of their features.

From the point of view of the chemometrical methods employed, the following statements can be made:

(1) PCA allows a very good characterization of the different wines and the interpretation of the chemical aspects connected with the characterization itself.

(2) SOM appears very powerful in creating classes of samples and its results point out different aspects of the dataset from those emerging from PCA. The two chemometrical techniques appear to be complementary.

(3) SLDA seems to be the best classification method and the interpretability of its models improves the results of the statistical analysis.

Acknowledgements

The authors are gratefully indebted to the Chamber of Commerce of Cuneo, Sede Distaccata of Alba, in the person of Geom. Ravera, for providing us with the samples.

References

[1] A. Kaufmann, Mitt. Gebiete Lebensm. Hyg. 89 (1998) 97.
[2] A. Kaufmann, J. AOAC Int. 80 (1997) 665.
[3] I.S. Arvanitoyannis et al., Trends Food Sci. Technol. 10 (1999) 321.
[4] M.J. Baxter, H.M. Crews, M.J. Dennis, I. Goodall, D. Anderson, Food Chem. 60 (1997) 443.
[5] J.D. Greenouch, H.P. Longerich, S.E. Jackson, Aust. J. Grape Wine Res. 3 (1997) 75.
[6] L. Sun, K. Danzer, G. Thiel, Fresenius J. Anal. Chem. 359 (1997) 143.
[7] J. Aires de Sousa, Am. J. Enol. Vitic. 47 (1996) 410.
[8] D. De La Calle-Garcia, M. Reichenbächer, K. Danzer, Vitis 37 (1998) 181.
[9] I. Arozarena, A. Casp, R. Marin, M. Navarro, Eur. Food Res. Technol. 212 (2000) 108.
[10] P. Horn, P. Schaaf, B. Holbach, S. Hölz, H. Eschnauer, Z. Lebensm. Unters. Forsch. 196 (1993) 407.

[11] B. Holbach, H. Förstel, H. Otteneder, H. Hützen, Z. Lebensm. Unters. Forsch. 198 (1994) 223.
[12] G.J. Martin, M. Martin, in: H.F. Linskens, J.F. Jackson (Eds.), Wine Analysis, Springer, Berlin, 1988, p. 258.
[13] M.P. Day, B. Zhang, G.J. Martin, Am. J. Enol. Vitic. 45 (1994) 79.
[14] I. Moret, G. Scarponi, P. Cescon, J. Agric. Food Chem. 42 (1994) 1143.
[15] S. Rebolo, R.M. Peña, M.J. Latorre, S. Garcia, A.M. Botana, C. Herrero, Anal. Chim. Acta 417 (2000) 211.
[16] E. Gomez-Plaza, R. Gil-Munoz, A. Martinez-Cutillas, J. Sci. Food Agric. 80 (2000) 497.
[17] C. Armanino, M. Forina, M. Castino, A. Piracci, M. Ubigli, Analyst 115 (1990) 907.
[18] C.L. Arthur, J. Pawliszyn, Anal. Chem. 62 (1990) 2145.
[19] D. De la Calle Garcia, S. Magnaghi, M. Reichenbächer, K. Danzer, J. High Resolut. Chromatogr. 19 (1996) 257.
[20] D. De la Calle Garcia, M. Reichenbächer, K. Danzer, C. Hurlbeck, C. Bartzsch, K.H. Feller, J. High Resolut. Chromatogr. 20 (1997) 665.
[21] J. Weber, M. Beeg, C. Bartzsch, K.H. Feller, D. De la Calle Garcia, M. Reichenbächer, K. Danzer, J. High Resolut. Chromatogr. 22 (1999) 322.
[22] W.O. Kwam, B.R. Kowalski, J. Food Sci. 43 (1978) 1320.
[23] E. Marengo, C. Baiocchi, M.C. Gennaro, P.L. Bertolo, S. Lanteri, W. Garrone, Chemometr. Intell. Lab. Syst. 11 (1991) 75.
[24] J. Davis, Statistics and Data Analysis in Geology, Wiley, New York, 1986.
[25] B.S. Everitt, Cluster Analysis, Heineman, London, 1980.
[26] T. Kohonen, Self-Organisation and Associative Memory, 3rd ed., Springer, Berlin, 1987.
[27] J. Zupan, M. Novic, I. Ruisanchez, Chemometr. Intell. Lab. Syst. 38 (1997) 1.
[28] J. Zupan, J. Gasteiger, Neural Network for Chemist: an Introduction, VCH, Weinheim, 1993.
[29] S. Wold, Kem. Kemi 9 (1982) 401.
[30] I.E. Frank, S. Lanteri, Chemometr. Intell. Lab. Syst. 5 (1989) 247.
[31] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, London, 1979.
[32] M. James, Classification Algorithms, Collins, London, 1985.
[33] Y. Margalit, Concepts in Wine Chemistry, The Wine Appreciation Guild, San Francisco, CA, 1997.