Sugar Adulterations Control in Concentrated Rectified Grape Musts by Finite Mixture Distribution Analysis of the *myo*- and *scyllo*-Inositol Content and the D/H Methyl Ratio of Fermentative Ethanol

Ambrogio Monetti,* Giuseppe Versini, Giorgio Dalpiaz, and Fabiano Reniero†

Istituto Agrario di S. Michele all'Adige, 38010 S. Michele all'Adige-TN, Italy

Adulterations detection in concentrated rectified musts (CRM) could be strengthened by either making acceptance criteria capable of better reflecting the features of genuine samples or employing additional tracers suitable to ascertain the addition of exogenous sugars. In fact, thanks to their purity, CRM are an ideal substrate for adulterations with sugars of plants other than grape, in particular beet sucrose, able to emulate genuine samples. The present work shows that, moving from well-established standard methods like isotopic analysis of D/H and $^{13}C/^{12}C$ ratios and presence determination of *myo*-inositol, a better definition of genuineness characteristics could be achieved through a multivariate approach integrated with the measurements of *myo*-inositol and its isomer *scyllo*-inositol. The separation between adulterated and genuine samples has been obtained by implementing a multivariate version of the expectation−maximization (EM) algorithm, and the estimates have been later used to derive a classification rule based on generalized Mahalanobis distances. In this way it has been possible to highlight the effects of adulterations, in particular the dilution of polyalcohols, and the shortcomings of the present regulations. As a consequence, especially to solve the intermediate cases where attribution is normally difficult, we suggest a narrowing of the acceptance region by a true multiparametric approach integrated with *scyllo*-inositol.

Keywords: *Concentrated rectified grape musts; adulteration; deuterium; site specific natural isotope fractionation−nuclear magnetic resonance (SNIF-NMR); polyalcohols; expectation−maximization (EM) algorithm; multivariate discrimination*

INTRODUCTION

Concentrated rectified musts (CRM) are grape sugar solutions produced by concentrating grape musts which are depurated through ion-exchange resins (Garoglio, 1957; Pompei, 1982). According to the European Community (EC) regulation (EC, 1987, 1988), the CRM must have a quite high sugar concentration (at least 51.9 °Brix) and, in the meantime, be very low in residual impurities. As a consequence, they are potentially an ideal substrate to make profitable adulterations with inverted sucrose obtained from plants other than grape, especially beet or cane. Such adulterations should be detected following the reference methods of the Office International de la Vigne et du Vin (OIV) and EC (Martin and Brun, 1987; EC, 1990) for grape juices, wines, and byproducts like CRM. The methods are especially suitable against beet sugar addition; for the detection of C_4 plants sugar presence, they are integrated with isotopic ratio mass spectrometry (IRMS) measurement of $^{13}C/^{12}C$ of sugar or alcohol.

Martin's method (Martin et al., 1982; 1986b; Martin and Martin, 1988), reference for those of the OIV and EC, is based on the site specific natural isotope fractionation−nuclear magnetic resonance (SNIF-NMR) analysis of the D/H ratios in the methyl and methylene positions of ethanol, which depend on photosynthetic and physiological factors (Martin, 1988), conditioned by geoclimatic parameters. The D/H ratio at the isotopomer CH_2DCH_2OH of ethanol (D/H_I) gives in fact the most useful information about the botanical source of the sugar; the variation of the other isotopomer CH_3-CHDOH (D/H_{II}) is influenced more by the deuterium content of the fermentation medium (Martin et al., 1986a). Further details about the characteristics of different species and their differentiability through a multiisotopic approach including the ratio $R = 2(D/H_{II})/(D/H_I)$ as well as D/H of water are available in Martin et al. (1991a,b).

In the procedure used to discover sugar adulterations, the values of a suspect sample are compared with those of a reference one obtained under the same conditions of that under study, if available, or with a reference data bank of sound samples (Martin, 1990; EC, 1990), taking into account the natural variability. For the Italian wines, starting from 1987, a data bank has been set up with about 500 samples per annum and, since 1991, the data have been validated by a specific EC committee (EC, 1991). In the Italian wines, D/H_I and D/H_{II} range between 98 and 107 ppm and from about 124 to 137 ppm, respectively (Monetti et al., 1995), with a pattern clearly related to latitude and microclimatical conditions. The standard deviations per political region are between less than 0.5 and 1.9 ppm for D/H_I and from less than 1 up to 4 ppm for D/H_{II}, similar to those observed in some French areas by Martin et al. (1988). The corresponding $^{13}C/^{12}C$ values of ethanol vary between −28‰ to −23‰, with standard deviations per region between 0.4 and up to 1.2‰, with a less clear geographical influence (Versini et al., 1995). The southern (SI) and central (CI) Italian macroregions, which

* Corresponding author. Fax: ++39 461 650872. E-mail: Ambrogio.Monetti@ismaa.it.

† Current address: European Joint Research Centre, Ispra, VA, Italy.

S0021-8561(95)00235-4 CCC: $12.00

usually provide the musts for the production of CRM, are characterized by having the highest isotopic values: e.g., in the vintages of the period 1989−1993 the means of D/H_I were between 102.6 and 104.2 ppm in SI and between 100.8 and 102.8 ppm in CI; those of D/H_{II} ranged from 130.7 to 133.4 ppm in SI and from 131.9 to 132.1 ppm in CI, while the mean of $\delta^{13}C$ varied from −25.25 to −24.5‰ in SI and from −26.55 to −25.7‰ in CI.

Among the possible ones, the most interesting exogenous sugars to make adulterations are those of beet and cane which have a mean D/H_I value of the ethanol of 92.5 and 110.5 ppm (EC, 1990) and a $\delta^{13}C$ of −27‰ ± 2.0 and −11.5‰ ± 1.5, respectively (Martin, 1988). In the case of addition of beet inverted sucrose to the must, probably the easiest one, the isotopic behavior in alcohol differs: D/H_I drops, $\delta^{13}C$ is slightly reduced, D/H_{II} grows but it is counterbalanced by the limited addition of tap water present in the sugar syrup. In such a way, after a shrewd addition, an adulterate CRM coming from some southern Italian regions could be sold as a genuine one of central northern areas because its parameters, which remain within the range of natural variability, are admissible. After some calculations, it is easy to verify that mixtures with one-quarter or more of beet sugars could still mime natural D/H and $\delta^{13}C$ values and could be further improved by a restrained addition of sugar cane with its D/H_I values at the other extreme but with the limits due to the $\delta^{13}C$ variation.

As a consequence, it is necessary to find out other parameters or relationships which could improve the definition of the genuineness characteristics of CRM. Versini (1993) analyzed the $\delta^{13}C$ content of the sugar part of heterosides (mostly monoterpenediols, norisoprenoids, benzyl and phenethyl alcohol derivatives), only partially adsorbed by resins, and a high correlation with the $\delta^{13}C$ of the free sugars was found. However, this correlation has not enough prediction capability to limit satisfactorily the amount of potential adulteration. In the same direction, trace elements have been studied to complement stable isotopes analysis in musts characterization (Day et al., 1994, 1995), but they are not suitable because these tracers are quite totally eliminated in the production process of CRM.

In this context, useful tracers for CRM characterization could be *myo*- and *scyllo*-inositol. These polyalcohols originate in the grape berry and have been already proposed to control the genuininess of the CRM because they are neither retained by the resins nor present in other purified commercial sugars. A minimum content of *myo*-inositol (750 mg/kg sugars) and a ratio less than or equal to 20 between *myo*- and *scyllo*-inositol (Versini et al., 1984) have been suggested as genuineness indexes. In the EC (1990), on the other hand, this proposal has been only partially adopted and the controls are confined to the presence of *myo*-inositol alone without considering its commercial availability, its levels in musts, and the relationship between the two isomers.

In this situation, our aim was to verify the possibility of improving the CRM genuineness control against the most remarkable and frequent adulteration with beet sugar by integrating the results of the reference methods with those of the two polyalcohols jointly considered.

MATERIALS AND METHODS

Some 354 commercial samples of vintages 1987−1992 provided by the Central Bureau against adulterations of the Italian Ministry of Agriculture have been analyzed.

Chemical Methods. The amounts of both *myo*- and *scyllo*-inositol in raw CRMs have been determined by high-resolution gas chromatography (HRGC) after silylation (xylitol as internal standard), in accordance with the EC method no. 2676 (EC, 1990; Versini et al., 1984).

To obtain the alcohol for the NMR analysis, the samples have been fermented by diluting 150 g of CRM at about 65 °Brix with 700 g of tap water in a ratio of about 1:4 w/w and adding 1 g of dried yeasts *Saccharomyces cerevisiae* and 3 g of nutritive complex Bacto yeast nitrogen (DIFCO). The fermented solution has been distilled using a Bullio apparatus equipped with a glass column (110 cm × 2.3 cm i.d.) filled with Rashing rings, adapted to obtain a distillate in accordance with the EC regulation for alcoholic recovery with negligible isotopic fractionation and maximal proof. The alcoholic content has been determined with a Paar densitometer.

Weighed amounts of alcohol and tetramethylurea (TMU) with known isotopic content, provided by the Bureau Communitaire de Référence in Brussels, added with 150 μL of hexafluorobenzene for the lock signal (^{19}F) were analyzed by NMR spectroscopy using a Bruker AMX 400 instrument (line broadening = 0.5 Hz). All the chosen instrumental parameters meet the requirements of repeatability and reproducibility as specified in the official method (EC, 1990), and the results were expressed in parts per million.

However, according to this method, the CRM should be diluted with water having the same isotopic values of the vegetative water of the natural must to be used as reference (EC, 1990), while the OIV method of 1987, contemporary with our initial analyses, allowed one to dilute with tap water. Considering both the practical impossibility of obtaining a sample of a natural must of reference and the inaccuracy associated with the origin attribution of the CRM, we preferred to continue the experimentation diluting with tap water of the same locality (S. Michele all'Adige) also after the officialization of EC method. To make the D/H_I values of CRM comparable to those of wines, the measures of the CRM have been transformed by a linear relationship (with a correlation coefficient $r = 0.94$) previously determined (Ramponi, 1991) where

$$(D/H_I)_{WINE} = -14.35 + 1.16(D/H_I)_{CRM} \quad (1)$$

Statistical Methods. Without any prior information about the origin and the potential presence of adulterations, the first goal of the statistical analysis was the separation of the adulterated group from the other, to develop a rule really applicable in new samples classification. In this context, the problem of estimating the number of populations has been addressed by Aitkin and Rubin (1985), but in the present case the decision has been adopted considering the nature of the problem on hand and was supported by the presence of a distinct pattern with two components evident from the data.

This situation is an example of finite mixture distribution (FMD) which arises naturally when a statistical population is a mixture of k component populations. In this case f, the probability density function of the variables under examination does not depend only on mean and dispersion but on these values in each population and their proportions (Everitt and Hand, 1981):

$$f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{p}) = \sum_{i=1}^{k} p_i g_i(\boldsymbol{x};\boldsymbol{\mu_i},\boldsymbol{\Sigma}_i) \quad (2)$$

where k is the number of populations, p_i are the independent mixing proportions of the mixture such that $\sum_{i=1}^{k} p_i = 1$ and $p_i > 0$, g_i is the probability density function in the ith component which depends on $\boldsymbol{\mu_i}$ and $\boldsymbol{\Sigma}_i$, the mean vector and variance−covariance matrix of the m variables $\boldsymbol{x}$.

Many different estimation methods of the FMD exist (Böhning et al., 1992). The first published works relating to the mixture estimation problem were those of Newcomb in 1886, who used an iterative scheme (Wedel and DeSarbo, 1995), and of Pearson in 1894, based on the method of moments (Redner and Walker, 1984), but a variety of other

methods have been later proposed (Bhattacharya, 1967; Fowlkes, 1979; Quandt and Ramsey, 1978; Ganesalingam and McLachlan, 1981). For a complete review it is possible to refer to Everitt and Hand (1981), Titterington et al. (1985), and McLachlan and Basford (1988).

In this work the maximum likelihood approach has been adopted because, under very general conditions, it has some useful properties like consistent estimators, which converge in probability to the true parameter values, and asymptotic normality (Everitt, 1985). At the same time, the complex dependence of the likelihood function on the parameters to be estimated originates computational difficulties which cannot be explicitly solved and need an approximate solution through an iterative procedure (Redner and Walker, 1984) like the expectation−maximization algorithm (EM). This has been widely applied since the use of high-speed computers became widespread in the 1960s (e.g., Day, 1969, to a mixture of multivariate normal densities with common covariance matrix) but Dempster et al. (1977) were the ones who interpreted the mixture density estimation problem as an estimation involving incomplete data by regarding an unlabeled observation on the mixture as "missing" the group code. In this way, it was possible to relate the FMD problems to a broader class of statistical problems and to see that the EM algorithm for FMD is a specialization of a more general algorithm possessing desirable theoretical properties (Wu, 1983) which account for its global convergence behavior (Redner and Walker, 1984).

Starting with some initial guesses, the EM algorithm for maximizing a likelihood function with missing data consists of iterated applications of the following two steps: (1) in the expectation step, the membership probabilities of each observation for each component are estimated replacing the missing data (group code) by their expected values, given the nonmissing data and the current parameter values; (2) in the maximization step, using the data substituted in the previous step, the likelihood equations are solved for the complete data likelihood and the updated estimates are used to improve the parameters.

This is equivalent to k separate estimation problems with each observation contributing to the log likelihood associated with the separate components with a weight given by the estimated membership probability (Everitt and Hand, 1981). These two steps are repeated until a convergence criterion is satisfied, e.g., a negligible improvement of the likelihood or, as in the present work, a maximum absolute difference between all parameters in two successive iterations less than 10^{-5}. In the EM algorithm, the solution is faster if the populations are well separated and the chosen starting values are suitable (Titterington et al., 1985), the iterations are attractive and convergence is ensured, but the algorithm could require many iterations and could converge to local optima (Wedel and DeSarbo, 1995). These problems could originate from poor initial guesses or from the adoption of an homoscedastic model in the presence of some heteroscedasticity. A reformulation of the algorithm to improve its robustness when the initial estimates are inaccurate has been developed by Hathaway (1985 and 1986), while Basford and McLachlan (1985) have examined the problem of homogeneity of variances.

In the present work, to ensure consistent maximum likelihood estimates, *myo*- and *scyllo*-inositol have been transformed logarithmically. Furthermore, to also take into account the covariation between variables, a multivariate approach has been used and the estimation procedure has been implemented with the interactive matrix language of SAS (SAS, 1989) on a DEC VAX 4000-610 AXP computer with the OpenVMS 1.5 operating system. Examples of univariate implementation of the algorithm are available in Fortran (DerSimonian, 1986, 1990) and in Genstat (Whitaker, 1992). Recent developments are represented by the inclusion of the FMD analysis and the EM algorithm into the wider framework of generalized linear models (GLMs) (e.g., Jansen, 1993; Wedel and DeSarbo, 1995).

Through the application of the EM algorithm, the parameters of the two populations were estimated and the attention was focused on the not-adulterated group to derive a classification rule of practical interest. This criterion relies on the Mahalanobis generalized distance (Krzanowski, 1988), which allows one to classify new observations and to minimize the probability error by assigning an unknown observation to the group from which the distance on all the variables is lower than a prior established limit at a certain significance level.

Finally, to verify whether the genuine CRM are really not adulterated, the values of D/H_I, previously transformed with eq 1, have been compared with the values of genuine wine samples belonging to the Italian data bank (Monetti et al., 1993, 1994). With a one-way analysis of variance (ANOVA) on D/H_I and considering that the CRM are mainly produced with grapes from southern Italian regions, the not-adulterated candidates have been compared with the genuine wines of the same regions. Only this parameter has been considered because in the samples of the data bank the polyalcohol content is not determined and D/H_{II} is influenced by the characteristics of the tap water used for the dilution.

RESULTS AND DISCUSSION

Nine samples were initially excluded from the data set: seven of them proved to be adulterated because they were without *myo*-inositol; one of the others had a null content of *scyllo*-inositol, and the last one was probably adulterated with sugar cane, low polyalcohols (632 mg/kg sugars *myo*-inositol, 81 mg/kg sugars *scyllo*-inositol) and high D/H_I (108.40 ppm after transformation with eq 1). At the end, after derivation of the classification rule, they were all reintroduced into the data set to test the allocation.

Among the four variables, the behavior of which is showed in Figure 1, D/H_{II} has been discharged for its poor discriminating power (Monetti et al., 1994), also confirmed by its marginal frequency distribution. On the contrary, D/H_I was evidently bimodal, with a maximum at about 95 ppm for the adulterated group and 103 ppm for the others. In the same way, the polyalcohols were different in the two groups, with modes positioned at 750 and 1400 mg/kg sugars for *myo*-inositol and around 30 and 230 mg/kg sugars for *scyllo*-inositol in the adulterated and not-adulterated musts, respectively. *scyllo*-Inositol, apparently less different in the two groups, has been retained because it could be arbitrary to ignore its covariation with the other isomer and, more important, because it is not commercially available. In any case, looking at Figure 1, it must be kept in mind that the behavior of the frequency histograms is heavily influenced by the chosen number of classes.

At the same time, in the figure it is evident that the samples belong to two groups depending on the presence or absence of adulterations, but these two groups are not immediately identifiable because they overlap. As a consequence, without any prior information, the first analysis was directed toward the separation of the groups through the EM algorithm.

The initial estimates, necessary to start the algorithm, were based on both frequency distribution and probability plots of each variable, while the number of populations was established on the basis of the prior knowledge of the problem. The starting values for the adulterated group were $D/H_I = 95$ ppm, *myo*-inositol = 300 mg/kg sugars, and *scyllo*-inositol = 30 mg/kg sugars; those of the other one were $D/H_I = 103$ ppm, *myo*-inositol = 1400 mg/kg sugars, and *scyllo*-inositol = 231 mg/kg sugars. The polyalcohols were transformed logarithmically to obtain consistent estimates of the parameters (Basford and McLachlan, 1985) as well as to avoid the problem of their excess. This surplus, evident in Figure 1 looking at the tails of the distribu-

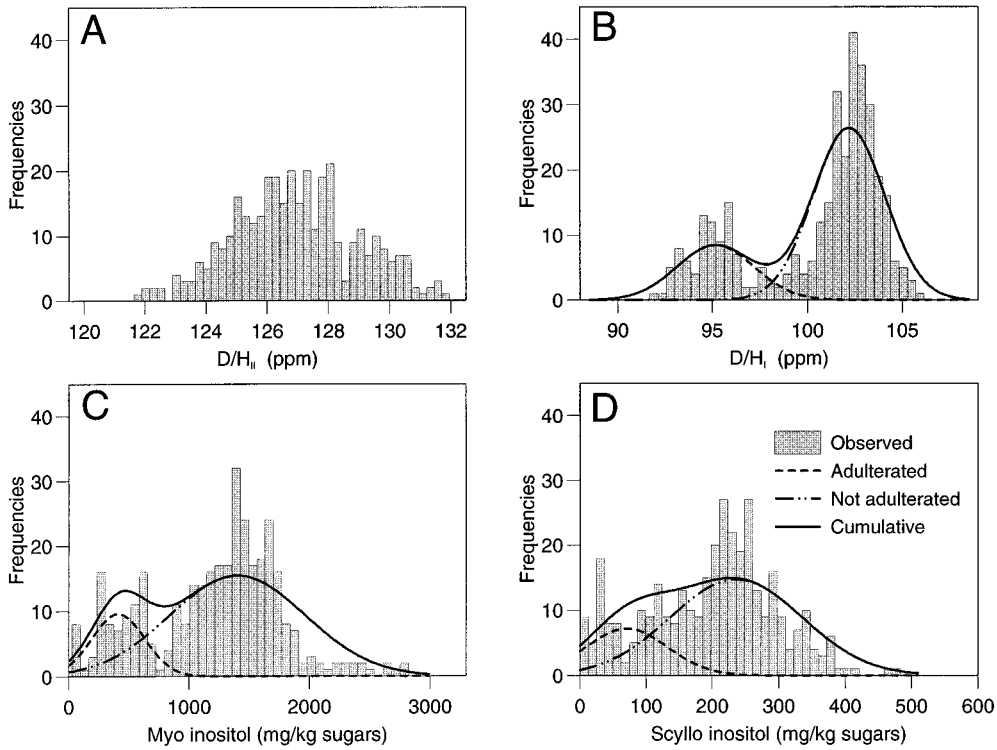

Figure 1. Marginal frequency distributions and probability density functions computed through the expectation−maximization (EM) algorithm on untrasformed data: (A) D/H$_{II}$, (B) D/H$_{I}$, (C) *myo*-inositol, and (D) *scyllo*-inositol.

Table 1. Statistical Parameters (Means, Variance−Covariance Matrices, and Proportions) of the Adulterated and the Genuine Group Estimated through the Expectation−Maximization (EM) Algorithm with Different Combinations of Variables

		adulterated samples			genuine samples		
	EM iterations	proportion (%)	mean	variance−covariance matrix	proportion (%)	mean	variance−covariance matrix
D/H$_{I}$	25	25.72	95.19	2.1470	74.28	102.18	1.7392
myo-inositol	80	26.44	6.13	0.2788	73.56	7.25	0.0569
scyllo-inositol	91	30.07	4.27	0.7285	69.93	5.44	0.0727
D/H$_{I}$ +	5	26.79	95.34	[2.6073 0.5027]	73.21	102.23	[1.6041 0.0183]
myo-inositol			6.12	[0.5027 0.2571]		7.26	[0.0183 0.0534]
D/H$_{I}$ +	11	25.07	95.11	[1.8983 0.5900]	74.93	102.15	[1.8425 0.1640]
scyllo-inositol			4.04	[0.5900 0.5442]		5.44	[0.1640 0.0777]
myo-inositol +	41	24.37	6.06	[0.2300 0.3263]	75.63	7.25	[0.0598 0.0463]
scyllo-inositol			4.03	[0.3263 0.5695]		5.43	[0.0463 0.0813]
D/H$_{I}$ +	11	24.19	94.99	[1.5906 0.2481 0.4889]	75.81	102.10	[2.0031 0.0603 0.1801]
myo-inositol +			6.04	[0.2481 0.2055 0.2930]		7.25	[0.0603 0.0570 0.0434]
scyllo-inositol			4.01	[0.4889 0.2930 0.5218]		5.44	[0.1801 0.0434 0.0794]

tions, was caused by considerable rots (Dittrich, 1987) common in certain vintages and generated by heavy and frequent rains which delayed the harvest.

To confirm that the solution was not a local optimum, other initial values were used, but the solution was unique. Estimated means, dispersion matrices, and proportions for each variable are shown in Table 1. The convergence process was relatively fast: e.g., in the trivariate case 11 iterations were needed and the trivariate joined density function for the two populations (adulterated, A, and genuine, G) was

$$f(\boldsymbol{x}) = 0.2419 f_{\mathrm{A}}(\boldsymbol{x}) + 0.7581 f_{\mathrm{G}}(\boldsymbol{x}) \qquad (3)$$

where

$$f_i(\boldsymbol{x}) = (2\pi)^{-m/2} |\boldsymbol{\Sigma_i}|^{-1/2} \mathrm{e}^{-1/2(\boldsymbol{x}-\boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_i)} \qquad (4)$$

is the *i*th component of the multivariate normal density (Everitt and Hand, 1981).

Since the main goal was the estimation of the parameters of the not-adulterated CRM population, the estimated proportions of each group were secondary, especially considering that they were influenced by the characteristics of the sample. In this case, working with commercial samples, the proportion that was adulterated was inflated by some firms that were subjected to more controls because their samples were frequently found anomalous. As a consequence, the proportions observed here are not representative of the real characteristics of the production of CRM but are specific to these samples under investigation.

In Table 1 it is possible to see that D/H$_{I}$ and *myo*- and *scyllo*-inositol need an increasing number of iterations because the separation between the groups is progressively less evident; in any case, to take into account all the information available and to get a robust discrimination, all the variables were used in subsequent analysis.

The goodness of fit of the EM estimates are shown in Figure 1 where the theoretical density functions, com-

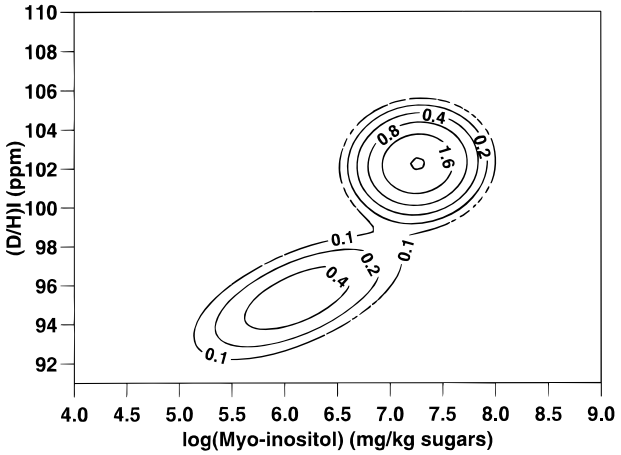

Figure 2. Contour plot of the estimated mixture density for D/H_I and *myo*-inositol.

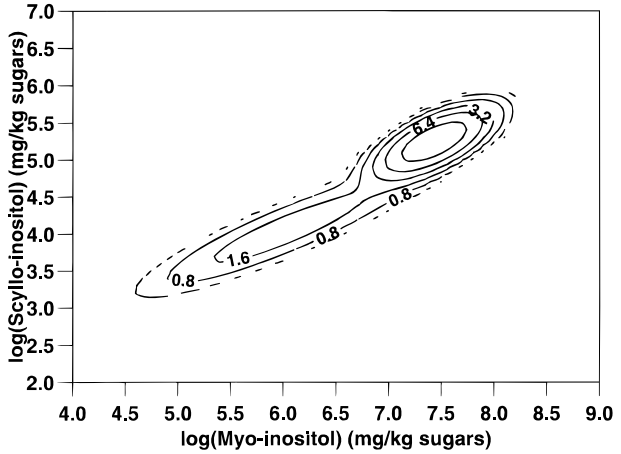

Figure 3. Contour plot of the estimated mixture density for the polyalcohols *myo*- and *scyllo*-inositol.

puted from untransformed data, are plotted onto the empirical frequency histograms.

Figures 2 and 3 show the estimated mixture densities in the form of a contour plot for D/H_I−*myo*-inositol and for the two polyalcohols. In the former, the two groups are well separated because their overlap is minimum and their orientation is different. In fact, in the adulterated group there is a substantial correlation of $r = 0.61$ between the two variables; in the other one, with a coefficient $r = 0.02$ there is nothing: the adulteration introduces a dependency by reducing the values of both variables. The same effect of dilution is appreciable in Figure 3: the polyalcohols are strongly correlated ($r = 0.90$) in the adulterated group and much less in the other one ($r = 0.66$) and the same behavior, with correlations respectively equal to $r = 0.58$ and $r = 0.43$, is observable in the combination D/H_I−*scyllo*-inositol.

With all the necessary parameters available to classify unknown samples, a criterion based on the genuine CRM was developed. The normal trivariate distribution of genuine CRM, completely defined by the mean vector and variance−covariance matrix, was used to derive a classification rule based on generalized Mahalanobis distance, $d_i^2 = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})'\mathbf{S}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$. This distance measures the extent of a sample from its centroid in an m-dimensional space and, knowing that d_i^2 is distributed as χ_m^2 (Krzanowski, 1988), the limit distance after which, with a certain probability, a sample does not belong to a certain population could be computed. In our case, the significance level was fixed at $\alpha = 5‰$ and, with $m = 3$, the corresponding limit distance was $\chi^2_{3;\alpha=5‰} = 12.84$.

The allocation rule for a new sample, indicating with $\boldsymbol{v}_i$ the vector of its values, with π_g the population of genuine CRM, and with π_a the population of the adulterated, could be stated as

$$\text{allocate } \boldsymbol{v}_i \text{ to } \pi_g \text{ if } d_i^2 \le \chi^2_{3;\alpha=5‰} \text{ else allocate } \boldsymbol{v}_i \text{ to } \pi_a$$

where d_i^2 is the generalized Mahalanobis distance $d_i^2 = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})'\mathbf{S}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$, $\bar{\boldsymbol{x}}' = [102.103\ 7.249\ 5.437]$ is the mean vector of D/H_I, *myo*-inositol, and *scyllo*-inositol respectively, and

$$\mathbf{S}^{-1} = \begin{bmatrix} 0.644570 & 0.735789 & -1.862891 \\ 0.735789 & 30.868915 & -18.518280 \\ -1.862891 & -18.518280 & 26.919630 \end{bmatrix}$$

is the inverse of the dispersion matrix of the not-adulterated CRM (Table 1). As an example, a sample with D/H_I, *myo*-inositol, and *scyllo*-inositol, respectively, equal to $\boldsymbol{v}_j' = [100\ 6.90\ 5.30]$ (corresponding to a content of 1000 and 200 mg/kg sugars for the two polyalcohols) will have $d_j^2 = 5.91$, less than the limit value of $\chi^2_{3;\alpha=5‰}$, and will be assigned to π_g; on the contrary, a suspect sample with $\boldsymbol{v}_i' = [98\ 6.21\ 4.38]$ (500 and 80 mg/kg sugars for the two polyalcohols) will have $d_i^2 = 31.77$, greater than the limit distance, and will be allocated to π_a.

Applying this rule to the whole data set, the results were comparable to those obtained with the EM algorithm: 261 samples (75.87%) were classified as genuine and 83 (24.13%) as adulterated.

To visualize these results, a parametric tolerance region with a radius proportional to $(\chi^2_{3;\alpha})^{1/2}$ for the not-adulterated population was computed (Krzanowski and Radley, 1989) after standardization of the values. The results are shown in Figure 4, where the planes defined by every combination of variables are separated to make interpretation easier. The radius of the spherical confidence region, for $\alpha = 5‰$, is $r = 3.58$ and it has been superimposed onto the acceptance region as defined by the present EC regulations and bounded by the values of D/H_I and the presence of *myo*-inositol. To meet the first requirement, a sample is acceptable if its D/H_I value remains within the range defined by the reference data bank of genuine samples; the range normally considered is the mean ±2 standard deviations, which roughly corresponds to a probability level of 95%. The second condition controls the presence of *myo*-inositol, therefore a sample is acceptable if its content is greater than 0. In the figure, the first condition is visualized by the height of the horizontal stripe between ±2.81 standard deviations (corresponding to a probability level of 99.5%) in the planes of D/H_I by the polyalcohols; the second condition is visualized by the extent of the area starting at null content of *myo*-inositol (−30 onto the standardized scale). Finally, it is worth remembering that the acceptance region is not affected by *scyllo*-inositol, which is not considered by the regulation.

From the figure the separation between adulterated and not-adulterated samples should be evident, totally enclosed within the tolerance region. After correction with eq 1, the latter has been validated through ANOVA with the corresponding group of genuine wines of southern Italy available in the data bank: the analysis was not significant ($F_{1,802} = 2.57$; $\alpha = 0.11$); thereby it

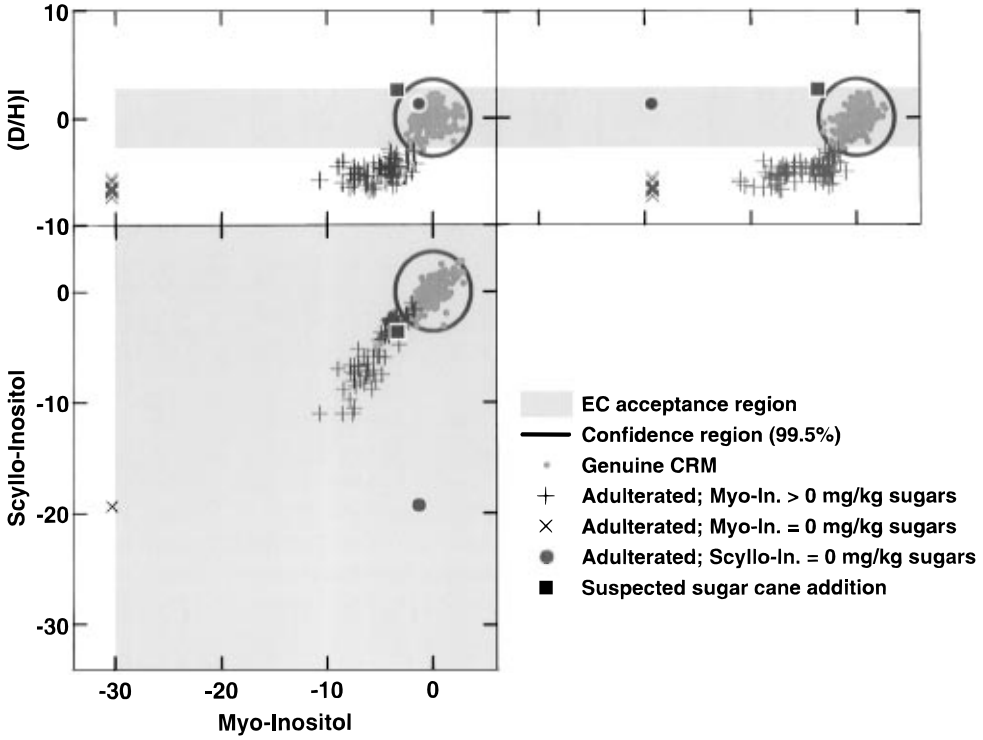

Figure 4. Multivariate tolerance region of acceptability derived from the classification based on generalized Mahalanobis distance superimposed onto the wider present EC acceptance region defined in terms of the mean content of D/H_I and presence of *myo*-inositol.

was possible to conclude that they belong to the same population of genuine samples.

Secondly, it should be evident that the present acceptance region is too indulgent because the genuine samples have a clear pattern and form a distinct cluster so that an acceptance region similar to the spherical one seems more appropriate.

Thirdly and more important, the present acceptance criteria could leave room for well-designed adulterations. In fact, from the figure, it is possible to identify three distinct groups: the genuine samples (•), those adulterated for D/H_I (+), and those adulterated both for *myo*-inositol and D/H_I (×), which were initially excluded. At the same time, an interesting behavior is shown by the other two samples initially considered suspect but genuine for the present EC regulation: the one without *scyllo*-inositol (●) is similar to those lacking in *myo*-inositol, while the other one (■), probably added with sugar cane, is similar to those lacking in both polyalcohols (dilution effect). These two samples were both excluded from the genuine group on the basis of the new classification rule, and in the figure they are out of the tolerance region. Both have probably had *myo*-inositol added, which is purchasable, and this practice is becoming well-known: for example, during routine analyses, we recently discovered a CRM commercial sample with a sucrose residual (determined by GC) of 1.5 g/kg of total sugars but with acceptable parameters: $D/H_I = 103.15$ ppm (after correction with eq 1) and $\delta^{13}C = -25.36‰$, both within the range of the data bank, and *myo*-inositol was present (2549 mg/kg sugars). On the other hand, considering both polyalcohols, this sample was proportionally low in *scyllo*-inositol, 211 mg/kg sugars, which is not considered by the present regulation, and it had a generalized distance from the centroid of the genuine CRM of $d_i^2 = 14.91$, greater than the limit value of 12.84 for $\alpha = 5‰$. Thanks to its sucrose residual, this sample has been rejected, but this case highlights the shortcomings of the present EC regulations because, with a complete inversion of the exogenous sugar, it could appear perfectly normal.

CONCLUSIONS

At present, the genuineness of CRM sugar content is mainly defined by the presence of *myo*-inositol and the SNIF-^{2}H NMR D/H ratio measured on the methyl site of fermentative ethanol and compared with a reference data bank of genuine samples. At the same time, if these two parameters are considered separately, adulteration discovery becomes very difficult because marginal frequency distributions are misleading and mixed between groups. In fact, CRM are lacking in tracers and extremely purified, while frauds, emulating the parameters of genuine samples, could have different origin (particularly beet but also cane, corn, or mixtures), level, and complexity (e.g., the integration through *myo*-inositol addition). As a consequence, a sharp boundary between adulterated and genuine samples does not exist.

In this situation, as demonstrated by some real cases, the only way of obtaining a confident classification also in the intermediate situations, where attribution is normally difficult, seems to be the improvement of the definition of the genuineness characteristics of CRM. This could be obtained by either narrowing the present acceptance region or employing additional tracers. The first requirement could be achieved if the available parameters are evaluated together, taking into account their covariation, and their measurements become quantitative. This is in particular the problem of *myo*-inositol which, owing to the present regulation, is determined only qualitatively and cannot fully contribute its information. Furthermore, adulterated samples, which have parameters individually acceptable but anomalous relationships among them, could be discov-

ered only taking into account covariation among variables, which allows a more precise acceptance region to be defined.

As regards the second requirement, since the polyalcohols are substantially independent and their content is not affected by the production process, it is worth considering *scyllo*-inositol as a qualified additional tracer of adulterations in CRM.

This tracer, together with the measurement of *myo*-inositol, will supplement reference methods like SNIF-NMR analysis of the D/H ratios and IRMS measurement of $^{13}C/^{12}C$ which are going to enhance their contribution in genuineness characterization of CRM to obtain an objective classification of the most difficult cases of adulteration.

LITERATURE CITED

Aitkin, M.; Rubin, D. B. Estimation and hypothesis testing in finite mixture models. *J. R. Stat. Soc. B.* **1985**, *47*, 67−75.

Basford, K. E.; McLachlan, G. J. Likelihood estimation with normal mixture models. *Appl. Stat.* **1985**, *34*, 282−289.

Bhattacharya, C. G. A simple method of resolution of a distribution into Gaussian components. *Biometrics* **1967**, *23*, 115−135.

Böhning, D.; Schlattman, P.; Lindsay, B. Computer-assisted analysis of mixtures (C.A.MAN): statistical algorithms. *Biometrics* **1992**, *48*, 283−303.

Day, M. P.; Zhang, B. L.; Martin, G. J. The use of trace element data to complement stable isotope methods in the characterization of grape musts. *Am. J. Enol. Vitic.* **1994**, *1*, 79−85.

Day, M. P.; Zhang, B. L.; Martin, G. J. Determination of the geographical origin of wine using joint analysis of elemental and isotopic composition. II-Differentiation of the principal production zones in France for the 1990 vintage. *J. Sci. Food Agric.* **1995**, *67*, 113−123.

Day, N. E. Estimating the components of a mixture of normal distributions. *Biometrika* **1969**, *56*, 463−474.

Dempster, A. P.; Laird, N. M.; Rubin, D. B. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. R. Stat. Soc. B* **1977**, *39*, 1−38.

DerSimonian, R. Maximum likelihood estimation of a mixing distribution. *Appl. Stat.* **1986**, *35*, 302−309.

DerSimonian, R. Correction to algorithm AS221: Maximum likelihood estimation of a mixing distribution. *Appl. Stat.* **1990**, *39*, 176.

Dittrich, H. H. *Mikrobiologie des Weines*; Ulmer: Stuttgart, Germany, 1987.

EC Regulation. Common organization of the market related to wine and wine products. No. 822/87, March 16, 1987.

EC Regulation. Modifications to Regulation (EC) No. 822/87 about the common organization of the market of wine and wine products. No. 2253/88, July 19, 1988.

EC Regulation. Community analytical methods which can be applied in the wine sector. No. 2676/90, Sept 17, 1990.

EC Regulation. Setting up of a data bank of deuterium nuclear magnetic resonance analyses in wine products. No. 2348/91, July 29, 1991.

Everitt, B. S. Mixture distributions. In *Encyclopedia of statistical sciences*; Kotz, S., Johnson, N. L., Read, C. B., Eds.; Wiley: New York, 1985; Vol. 5, 559−569.

Everitt, B. S.; Hand, D. J. *Finite mixture distributions*; Chapman and Hall: London, 1981.

Fowlkes, E. B. Some methods for studying the mixture of two normal (lognormal) distributions. *J. Am. Stat. Assoc.* **1979**, *74*, 561−575.

Ganesalingam, S.; McLachlan, G. J. Some efficiency results for the estimation of the mixing proportion in a mixture of two normal distributions. *Biometrics* **1981**, *37*, 23−33.

Garoglio, G. *Experimental contributions to study the potential applications of ion-exchange resins in oenology* (Italian); Strianti: San Casciano, Firenze, Italy, 1957.

Hathaway, R. J. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Ann. Stat.* **1985**, *2*, 795−800.

Hathaway, R. J. A constrained EM algorithm for univariate normal mixtures. *J. Stat. Comput. Simul.* **1986**, *23*, 211−230.

Jansen, R. C. Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. *Biometrics* **1993**, *49*, 227−231.

Krzanowski, W. J.; Radley, D. Nonparametric confidence and tolerance regions in canonical variate analysis. *Biometrics* **1989**, *45*, 1163−1173.

Krzanowski, W. J. *Principles of multivariate analysis*; Clarendon Press: Oxford, 1988.

Martin, G. J. Les applications de la mesure par resonance magnetique nucleaire du fractionnement isotopique naturel specifique (RMN-FINS) en viticulture et en oenologie. *Rev. Fr. Oenol.* **1988**, *114*, 23−24,53−60.

Martin, G. J. The chemistry of chaptalization. *Endeavour* **1990**, *14*, 137−143.

Martin, G. J.; Brun, S. Application de la résonance magnétique nucléaire du deutérium au contrôle des moûts, des moûts concentrés, du sucre de raisin et des vins (R.M.N.-F.I.N.S./S.N.I.F.-N.M.R.). *Bull. O.I.V.* **1987**, *671−672*, 131−145.

Martin, G. J.; Danho, D.; Vallet, C. Natural isotope fractionation in the discrimination of sugar origins. *J. Sci. Food Agric.* **1991a**, *56*, 419−434.

Martin, G. J.; Guillou, C.; Martin, M. L.; Cabanis, M.-T.; Tep, Y.; Aerny, J. Natural factors of isotope fractionation and the characterization of wines. *J. Agric. Food Chem.* **1988**, *36*, 316−322.

Martin, G. J.; Guillou, C.; Naulet, N.; Brun, S.; Tep, Y.; Cabanis, J.-C.; Cabanis, M.-T.; Sudraud, P. Contrôle de l'origine et de l'enrichssement des vins par analyse isotopique spécifique. (Control of origin and enrichment of wine by specific isotope analysis). *Sci. Aliments* **1986a**, *6*, 385−405.

Martin, G. J.; Martin, M. L. The site-specific natural isotope fractionation-NMR method applied to the study of wines. In *Modern methods of plant analysis*; Linskens, H. F., Jackson, J. F., Eds.; Springer: Berlin, 1988; Vol. 6, Chapter 9.

Martin, G. J.; Martin, M. L.; Mabon, F.; Michon, M. J. Identification of the origin of natural alcohols by natural abundance hydrogen-2 nuclear magnetic resonance. *Anal. Chem.* **1982**, *54*, 2380−2382.

Martin, G. J.; Zhang, B. L.; Naulet, N.; Martin, M. L. Deuterium transfer in the bioconversion of glucose to ethanol studied by specific isotope labeling at the natural abundance level. *J. Am. Chem. Soc.* **1986b**, *108*, 5116−5122.

Martin, M. L.; Martin, G. J.; Guillou, C. A site-specific and multi-element isotopic approach to origin inference of sugars in foods and beverages. *Mikrochim. Acta* **1991b**, *2*, 81−91.

McLachlan, G. J.; Basford, K. E. *Mixture models;* Marcel Dekker: New York, 1988.

Monetti, A.; Reniero, F.; Versini, G. SNIF-NMR ^2H analysis of alcohols of Italian genuine wines: study of interregional and interannual variability: Presented at the 2nd European Symposium on Food Authenticity, Nantes, France, October 1993.

Monetti, A.; Reniero, F.; Versini, G. Interregional and interannual variability of ethanol site-specific deuterium content in Italian wines. *Z. Lebensm. Unters. Forsch.* **1994**, *199*, 311−316.

Monetti, A.; Versini, G.; Reniero F. Classification of Italian wines on a regional scale by means of a multi-isotopic analysis. In *Food flavors: generation, analysis and process influence*; Charalambous, G., Ed.; Elsevier: Amsterdam, 1995; pp 1723−1730.

Pompei, C. Le sucre du raisin. Aspects technologiques (production et utilisation). *Bull. O.I.V.* **1982**, *611*, 25−52.

Quandt, R. E.; Ramsey, J. B. Estimating mixtures of normal distributions and switching regressions. *J. Am. Stat. Assoc.* **1978**, *73*, 730−738.

Ramponi, M. Characterization of some constituents of grape musts and of concentrated rectified grape musts through analytical parameters and the stable isotopes content (Italian). Thesis, The University of Padua, Italy, 1991

Redner, R. A.; Walker, H. F. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **1984**, *26*, 195−239.

SAS Institute Inc. *SAS/IML Software: Usage and Reference,* Version 6, 1st ed.; SAS Institute Inc.: Cary, NC, 1989.

Titterington, D. M.; Smith, A. F. M.; Makov, U. E. *Statistical analysis of finite mixture distributions;* Wiley: Chichester, Great Britain, 1985.

Versini, G. New methods for the evaluation of genuineness of grape derivates. *Proceedings of the 10th International Oenology Symposium*, Montreux, Switzerland; International Association Winery Technology Management: Breisach, Germany, 1993; pp 440−450.

Versini, G.; Dallaserra, A.; Margheri, G. Polyalcohols and secondary sugars in concentrated rectified musts as genuineness parameters (Italian). *Vignevini* **1984**, *11*, 41−47.

Versini, G.; Monetti, A.; Reniero, F. Monitoring authenticity and regional origin of wines by natural stable isotope ratios analysis. Presented at the 210th National Meeting of the American Chemical Society, Chicago, IL, August 1995.

Wedel, M.; DeSarbo, W. S. A mixture likelihood approach for generalized linear models. *J. Classification* **1995**, *12*, 21−55.

Whitaker, C. J. Finite mixture distributions. *Genstat Newsletter* **1992**, *28*, 39−48.

Wu, J. C. F. On the convergence properties of the EM algorithm. *Ann. Stat.* **1983**, *11*, 95−103.

Received for review April 21, 1995. Revised manuscript received March 19, 1996. Accepted May 3, 1996.[⊗]

JF9502358

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1996.