JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Classification of German White Wines with Certified Brand of Origin by Multielement Quantitation and Pattern Recognition Techniques

MARIA DEL MAR CASTIÑEIRA GÓMEZ,*,†,§ INGO FELDMANN,† NORBERT JAKUBOWSKI,† AND JAN T. ANDERSSON§

Institute of Spectrochemistry and Applied Spectroscopy, Bunsen-Kirchhoff-Strasse 11, D-44139 Dortmund, Germany, and Department of Inorganic and Analytical Chemistry, University of Münster, Correnstrasse 30, D-48149 Münster, Germany

A procedure is proposed for the determination of the authenticity of white wines from four German wine-growing regions (Baden, Rheingau, Rheinhessen, and Pfalz) based on their content of some major, trace, and ultratrace elements. One hundred and twenty-seven white wine samples possessing a certificate of origin, all of the 2000 vintage, were analyzed. The concentrations of 13 elements (Li, B, Mg, Ca, V, Mn, Co, Fe, Zn, Rb, Sr, Cs, and Pb) were determined in wine diluted 1:20 by sector field inductively coupled plasma mass spectrometry (SF-ICP-MS). Indium was routinely used as internal standard. Supervised pattern recognition techniques such as discriminant analysis and classification trees were applied for the interpretation of the data. A quadratic discriminant analysis (QDA) allowed the four regions to be discriminated with 83% accuracy when using only eight variables (Li, B, Mg, Fe, Zn, Sr, Cs, and Pb), and the prediction ability for classifying new samples was 76%. By use of a second method, a decision tree, the classification of samples coming from the four regions could be performed with an accuracy of 84% when only four elements were used: Li (very low in samples from Baden), Zn (abnormally low in the samples from the Rheingau), and Mg and Sr (both important for the differentiation between Pfalz and Rheinhessen samples). For this method, the prediction ability was only 74% in the identification of unknown samples. The robustness of the QDA model was not good enough, and therefore the tree is better recommended for the classification of new wine samples from these areas of German wine production.

KEYWORDS: Wine; SF-ICP-MS; pattern recognition techniques

INTRODUCTION

Chemometric classification procedures are very attractive for the classification of products of certified origin. The extraction of useful information from measured data and the optimum use of such analytical information are important objectives of chemometrics (*1*). In chemometrics, pattern recognition methods are especially useful for the classification of a number of individuals or objects into discrete classes on the basis of measured features or variables (*2*). Pattern recognition methods can be divided into supervised and nonsupervised methods. Supervised methods, such as linear discriminant analysis (LDA), *K* nearest-neighbor (KNN), or soft independent modeling of class analogies (SIMCA), are used when the classes or groups are known and the aim is to find rules for deciding in which class an investigated sample should be placed. Nonsupervised methods (clustering techniques) are those in which the natural classification of the objects characterized by a number of variables is not known and the aim is to define groups of similar objects (*3*). A second classification of pattern recognition methods is between parametric and nonparametric methods. In the parametric techniques, such as LDA, statistical parameters for the distribution of the samples are used in the derivation of the discriminant function. A correct application of the methods demands that the data fulfill two statistical assumptions: (1) a normal distribution and (2) equal variance of the classes, that is, that the classes have equal dispersion of the data in every variable. Nonparametric methods such as the classification and regression trees (C&RT) (*4*) are not explicitly based on distribution statistics. Classification trees include those models in which the variables are categorical, and regression trees include those in which they are continuous.

The use of multielement patterns and pattern recognition techniques has gained ground in recent years for the geographical classification of wines originating from different parts of the world. Pioneering studies on German wines were done by Siegmund and Bächmann (*5*) in 1977; they determined 15 elements in 70 wines using neutron activation analysis (NAA).

* Corresponding author (e-mail castineira@isas-dortmund.de).
† Institute of Spectrochemistry and Applied Spectroscopy.
§ University of Münster.

10.1021/jf035120f CCC: $27.50 © 2004 American Chemical Society
Published on Web 04/20/2004

A hierarchical cluster analysis could be used to classify the samples according to their origin. Maarse et al. (*6*) differentiated between white wines from the regions Mosel and Pfalz. The wines were produced from the same grape variety, Riesling, but from different harvests. Fourteen trace elements were measured, as were volatile and nonvolatile compounds and amino acids. LDA resulted in the best differentiation based on the iron, rubidium, and sodium contents. Thiel et al. (*7*, *8*) analyzed 16 elements and organic components in 165 wines from four German wine-growing regions: Saale-Unstrut, Rheingau, Baden, and Rheinhessen. In that work, two subregions of Baden and two of Rheinhessen were established as groups. Therefore, a total of six groups of classes were found. LDA was applied with the best results for geographical identification (prediction rate of 83%) achieved with B, Ba, Ca, Mg, Pb, Si, Sr, and V. Because some organic volatile compounds allowed the differentiation between wines of various grape varieties, a classification of wines according to geographic origin via varietal differentiation was attempted using the elements Ba, Sr, Si, and V and the organic compounds furfural, 3-methyl-1-pentanol, and 4-methyl-1-pentanol. A recognition ability of 90% was achieved.

In the accompanying paper we reported on the changes in metal concentration throughout two typical winemaking processes for white wines from five geographically defined wine-growing regions in Germany: Mittelrhein, Rheinhessen, Rheingau, Baden, and Pfalz. Of the 63 elements investigated, only 13 elements were robust enough for wine origin studies by showing only minor deviation in their concentration in the must and in the wine. These 13 elements (ultratrace, trace, and major components) are Li, B, Mg, Ca, V, Mn, Fe, Co, Zn, Rb, Sr, Cs, and Pb. The aim of the present work is to check if it is possible to differentiate between wines from four of these previously mentioned wine-growing regions in Germany (Baden, Pfalz, Rheinhessen, and Rheingau) by determining the 13 elements mentioned and to create a simple rule that allows for a fast classification of unknown samples.

EXPERIMENTAL PROCEDURES

Material and Reagents. A multielement homemade stock solution of 13 elements in the concentration range of 25 ng/mL−50 μg/mL was prepared from the following single ICP standards containing 1000 mg/L (all from Merck, Darmstadt, Germany): Cs (Cs_2CO_3), Mn [$Mn(NO_3)_2$], Sr [$Sr(NO_3)_2$], and Rb (Rb_2CO_3), all in 5% HNO_3; V (NH_4VO_3), Li ($LiNO_3$), Fe [$Fe(NO_3)_3$], Co [$Co(NO_3)_2$], Pb [$Pb(NO_3)_2$], and Zn [$Zn(NO_3)_2$] in 2−3% HNO_3; and B (H_3BO_3) in H_2O. From this initial stock solution five standards for calibration were prepared. Ca, 1000 mg/L ($CaCO_3$ in 5% HNO_3, Merck), and Mg, 1000 mg/L ($Mg(NO_3)_2$ in 2−3% HNO_3, Merck) were also separately added to each calibration standard in the corresponding amount. Indium (10 μg/mL in 5% HNO_3) was used as internal standard in each standard solution with a final concentration of 1 ng/mL.

Standards were prepared from the above-mentioned homemade stock solution in polyethylene volumetric tubes (25 mL). HNO_3 (125 μL) and EtOH (125 μL) were added (for partial matrix matching) and diluted with twice-distilled water to the final volume. The concentration range of the standards for the external calibration depended on the element and varied from 0.05 ng/mL to 10 μg/mL. The concentration of the elements in the stock solution was selected to simulate the natural wine concentrations. Blanks were similarly prepared.

A homemade reference wine was prepared for daily quality control of the analysis. Twenty-five milliliters of a white wine (Nierstein, 2000) with characteristics similar to those of the samples were diluted with twice-distilled H_2O to 500 mL. This was called quality control Nierstein (QCN). The wine solution was acidified in the same way as the other wines to a concentration of ∼1% HNO_3. Indium was added to a final concentration of 1 ng/mL.

Table 1. Operating Conditions for the SF-ICP-MS Intrument: Element 2 (Finnigan MAT, Bremen, Germany)

parameter	value
analyzed isotopes per resolution	low ($R = 400$): ^{7}Li, ^{11}B, ^{51}V, ^{66}Zn, ^{68}Zn, ^{85}Rb, ^{88}Sr, ^{115}In, ^{133}Cs, ^{206}Pb, ^{207}Pb, ^{208}Pb medium ($R = 4000$): ^{25}Mg, ^{26}Mg, ^{44}Ca, ^{55}Mn, ^{56}Fe, ^{59}Co, ^{115}In
cooling gas (L/min)	16
auxiliary gas (L/min)	1–1.1
sample gas (L/min)	1.06–1.08
plasma power (W)	1210–1300
guard electrode	yes
sample uptake (mL/min)	0.75
autosampler	ASX 500 (Cetac, Omaha, NE)
spray chamber	Scott (cooled to 4 °C)
nebulizer	Meinhard
cones	Ni-cones, i.d.: sampler, 1 mm; skimmer, 0.7 mm
data acquisition procedure	no. of scans: 15 scan type: EScan dwell time: 10 ms/channel samples per peak: 26 isotope segment duration: 0.3 s acquisition time: 120 s

A cleaning solution containing 2% HNO_3 and 0.5% EtOH was freshly prepared each day for cleaning the system between wine samples. Finally, a tuning solution (1 ng/mL) was prepared from a multielement standard solution with nine elements (10 μg/mL of Be, Bi, Ce, Co, In, Mg, Ni, Pb, and U; multielement solution 5 from Cläritas, Spex, Germany) for the daily tuning of the instrument.

Wine Samples. A set of 127 commercially available German white wines was analyzed. The wines were selected according to the following properties: (a) white wines from one of the German regions [Baden (B), $n = 29$; Rheingau (RG), $n = 36$; Rheinhessen (RH), $n = 30$; or Pfalz (P), $n = 32$]; (b) a certified origin as demonstrated by the official control number (A.P.Nr.; *Amtliche Prüfungsnummer*); this number is given to every wine that passes an official control and ensures (*9*, *10*) that (i) the wine comes from the region mentioned so they are Q.b.A. (*Qualitätswein bestimmter Anbaugebiete*), (ii) the wine was produced from the authorized grape variety, and (iii) the wine has the minimum sugar and alcohol contents required for its quality classification and the wine-growing region]; (c) being of the vintage 2000 (this criterion was used to avoid the possible year-to-year variations in the concentration of some elements); (d) belonging to one of the main German grape varieties (Riesling, Müller-Thurgau, Gutedel, Kerner, Silvaner, Weiss-, or Grauburgunder) (it was not possible to find varied enough samples for only one grape, as would have been desirable); (e) being a dry wine, thereby reducing plasma problems due to the higher sugar content of sweet wines.

For sample preparation, the corks were carefully removed from the bottle and an aliquot was decanted into PP tubes of 50 mL capacity. The tubes were washed twice with the wine before the final filling. The diluted samples were prepared daily in sterile PP centrifuge tubes of 15 mL capacity by taking 0.5 mL from the PP tubes and diluting to 10 mL using HNO_3 subb. (Merck) and EtOH 98% p.a. (Merck) to reach concentrations of 1 and 0.5%, respectively. Every sample of wine was prepared in triplicate. After sample preparation, the PP tubes were dated and stored in the refrigerator as back-up. The wine bottles were closed and stored at room temperature.

ICP-MS Analysis. The ICP-MS was a sector field ICP-MS (Element 2, Finnigan MAT, Bremen, Germany). Typical operating conditions are shown in **Table 1**. The 13 elements for analysis were chosen as in our previous study (accompanying paper). The isotopes of the elements were chosen due to their major natural abundance. Some of these isotopes suffer from spectral interferences, and in such instances medium resolution ($R = 4000$) instead of low resolution ($R = 400$) was used for ^{25}Mg ($\rightarrow {}^{12}C_2{}^{1}H^{+}$), ^{44}Ca ($\rightarrow {}^{12}C^{16}O_2{}^{+}$, $^{88}Sr^{2+}$), ^{55}Mn ($\rightarrow {}^{40}Ar^{14}N^{1}H^{+}$, $^{39}K^{16}O^{+}$), ^{56}Fe ($\rightarrow {}^{40}Ca^{16}O^{+}$), and ^{59}Co ($\rightarrow {}^{43}Ca^{16}O^{+}$).

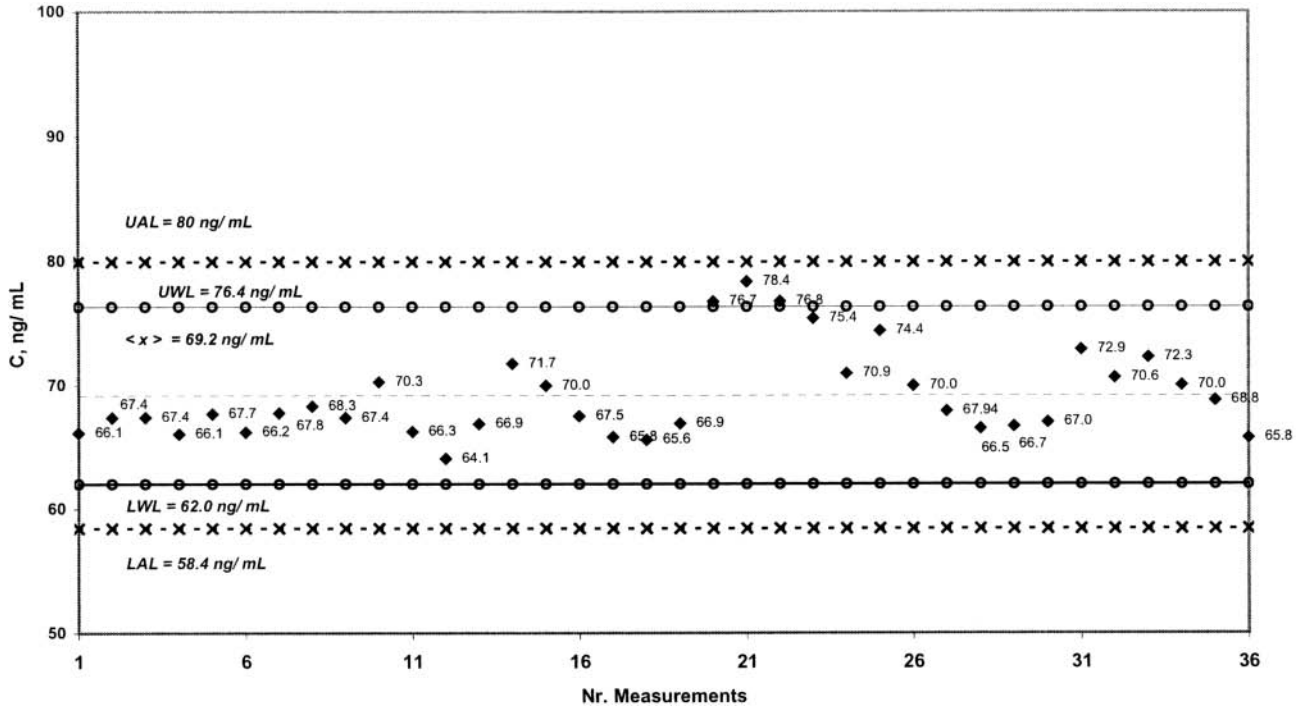

Figure 1. Shewhart chart for the day-to-day variation of lithium concentration in the QCN wine throughout the measurement period: <*x*>, mean value; UAL, upper action limit; UWL, upper warning limit; LWL, lower warning limit; LAL, lower action limit (values are given in ng/mL).

Table 2. Recovery Percentages[a] Obtained for the Spiked Quality Control Nierstein Wine by SF-ICP-MS

element	recovery %[a]	element	recovery %[a]
Li	105.5 ± 9.4	Fe	90.0 ± 7.0
B	96.5 ± 7.6	Zn	81.5 ± 3.4
Mg	89.2 ± 7.3	Rb	91.7 ± 8.2
Ca	96.0 ± 8.0	Sr	97.2 ± 5.1
V	93.5 ± 5.9	Cs	94.2 ± 3.7
Mn	87.4 ± 7.9	Pb	91.8 ± 6.6
Co	82.9 ± 5.2		

[a] Mean ± standard deviation.

Table 3. Quality Control Nierstein Measured by SF-ICP-MS, ICP-OES, and TXRF[a]

	SF-ICP-MS	ICP-OES	TXRF
Li	70.0 ± 4.4	64 ± 5	nd[b]
B*	5.6 ± 3.6		nd
Mg*	92.7 ± 4.4	93 ± 2	nd
Ca*	131.0 ± 5.4	107 ± 3	105 ± 10
V	15.3 ± 0.9	<22	<20
Mn*	1.3 ± 0.4	1.2 ± 0.1	1.2 ±0.1
Fe*	3.6 ±0.1	3.5 ± 0.1	3.4 ± 0.2
Co	5.6 ± 0.3	<17	nd
Zn	1.2 ± 0.05	1.1 ± 0.1	1.1 ± 0.1
Rb	291 ± 11	380 ± 20	300 ± 20
Sr	473 ± 11	410 ± 20	500 ± 30
Cs	1.6 ± 0.1	<1000	nd
Pb	37.8 ± 1.3	<50	<70

[a] Data are given in ng/mL unless denoted by an asterisk (*), which are in μg/mL. [b] Not detectable.

The instrument was tuned daily and compromise instrumental conditions were set to achieve the optimal signal intensity and stability for Li, In, and Ce.

The analyses were routinely performed with the aid of an Autosampler (ASX 500, Cetac, Omaha, NE). Every day, two to three series were measured, each series consisting of five wines (three replicates of each one), five calibration standards, five blanks, and two quality control samples. The system was rinsed with the cleaning solution for 2 min between samples and also 2 min with the sample itself before a measurement was started. As quality control samples, one of the standards with intermediate concentrations and the QCN wine were measured at the end of each series for (i) checking any drift of the system during measurement and (ii) checking the day-to-day reproducibility of the system. By plotting Shewhart charts (*11*) of every element, we could easily check when the measurement should be repeated. For an example of a control chart, see **Figure 1** for the case of Li. The mean value and the associated standard deviation (*s*) were calculated by measuring QCN a number of times under a variety of conditions. Warning limits are set at ±2*s* and action limits at ±3*s*. Measuring points outside the action limits were deleted and the measurements repeated. When points fell outside the warning limits, the measurements were carefully checked for error sources and, depending on the case, repeated or not.

Table 4. Concentration and Day-to-Day Precision of the Quality Control Nierstein Measured over the Period of 3 Months[a]

element	$\bar{x}$ (ng/mL)	*s* (ng/mL)	RSD (%)
Li	70.0	4.4	6.3
B	5640	356	6.3
Mg	92700	4423	4.8
Ca	131000	5372	4.1
V	15.3	0.88	5.8
Mn	1250	42	3.3
Fe	3630	111	3.0
Co	5.59	0.29	5.2
Zn	1190	46	3.9
Rb	291	11	3.8
Sr	473	11	2.4
Cs	1.59	0.05	3.2
Pb	37.8	1.3	3.4

[a] $\bar{x}$ = mean concentration (n = 47); s = standard deviation; RSD = relative standard deviation (values are given in ng/mL).

After analysis of all the samples, 12 of the stored wine samples were randomly selected from all four regions and again measured to test the stability of the samples with time. The results were satisfactory with differences of <10% during a period of 3 months.

Table 5. Limits of Quantification (10*s*, 11 Blanks), Median, Minimum, and Maximum Values of 13 Elements in German White Wines from the Wine-Growing Regions of Baden, Rheingau, Rheinhessen, and Pfalz[a]

		Baden (29 samples)			Rheingau (36 samples)			Rheinhessen (30 samples)			Pfalz (32 samples)		
	LOQ	median	min	max	median	min	max	median	min	max	median	min	max
Li	0.24	3.3	1.5	7.3	13	4.5	49	31	9.5	75	9.0	3.0	150
B*	0.17	4.3	3.3	5.5	4.0	3.2	5.2	4.9	3.2	7	4.1	3.3	7.1
Mg*	0.01	71	61	81	74	66	100	91	75	137	73	60.7	105
Ca*	0.06	88	58	150	110	85	180	120	61	431	110	72.8	190
V	0.03	7.5	1.2	364	58	4.6	260	36	7.2	228	57	3.12	220
Mn*	0.001	1.0	0.7	1.4	0.9	0.3	2.1	1.1	0.1	1.8	1.0	0.7	1.7
Fe*	0.01	1.0	0.4	3.9	1.1	0.4	3.5	3.6	0.5	9.9	2.2	0.5	4.4
Co	0.08	2.7	1.1	4.7	4.2	1.3	15	4.6	0.4	12	5.3	1.1	12
Zn*	0.05	1.1	0.7	1.8	0.3	0.01	3.8	1.1	0.1	3.9	1.2	0.3	2.2
Rb	0.21	380	230	800	360	250	690	430	160	970	380	180	600
Sr	0.27	250	170	380	390	230	900	530	380	970	340	170	590
Cs	0.01	1.9	0.7	6.1	3.8	0.2	50	1.4	0.1	6.3	2.4	0.4	15
Pb	0.47	9.7	3.0	41	23	12	230	32	15	85	23	9.8	63

[a] All concentrations are given in ng/mL unless denoted by an asterisk (*), which are given in μg/mL.

Table 6. Pearson Correlations Obtained for Average Multielement Concentrations in the Different Wines of the Different Regions: Pfalz (P), Baden (B), Rheinhessen (RH), and Rheingau (RG)

		Li	B	Mg	Ca	V	Mn	Fe	Co	Zn	Rb	Sr	Cs	Pb
Li	P	1	0.498	0.084	–0.044	0.26	–0.095	–0.327	–0.144	0.368	0.427	0.55	0.657	–0.16
	B	1	0.126	0.382	0.326	0.373	0.037	0.122	0.532	0.571	–0.18	0.412	0.009	0.41
	RG	1	–0.079	0.288	0.012	0.205	0.278	0.011	–0.059	0.18	0.581	0.298	0.448	0.117
	RH	1	0.324	–0.028	–0.273	0.037	0.353	0.391	0.284	0.112	0.357	0.153	0.317	–0.08
B	P	0.498	1	0.421	0.202	0.348	0.096	0.01	0.061	0.529	0.166	0.557	0.33	0.203
	B	0.126	1	0.425	0.286	0.247	0.475	0.127	0.025	0.066	0.339	0.625	–0.219	0.143
	RG	–0.079	1	0.514	0.104	–0.102	0.269	–0.167	0.006	–0.076	0.256	0.48	–0.035	–0.074
	RH	0.324	1	0.681	0.242	0.196	0.033	–0.132	–0.246	–0.156	0.177	0.597	0.309	–0.086
Mg	P	0.084	0.421	1	0.306	–0.016	0.525	0.545	0.265	0.38	0.259	0.607	–0.178	0.76
	B	0.382	0.425	1	0.584	0.123	0.295	–0.038	0.246	0.402	0.385	0.592	0.334	0.381
	RG	0.288	0.514	1	0.443	–0.242	0.753	0.134	0.133	0.463	0.416	0.839	0.22	0.429
	RH	–0.028	0.681	1	0.673	0.166	0.143	–0.078	–0.077	–0.084	0.138	0.805	0.18	0.183
Ca	P	–0.044	0.202	0.306	1	0.463	0.188	0.239	0.036	0.149	0.046	0.392	–0.276	0.109
	B	0.326	0.286	0.584	1	0.05	0.306	0.145	0.172	0.223	0.031	0.354	0.336	0.271
	RG	0.012	0.104	0.443	1	–0.186	0.382	0.247	–0.014	0.327	0.085	0.338	0.031	0.316
	RH	–0.273	0.242	0.673	1	0.06	–0.142	–0.212	–0.114	–0.211	–0.173	0.643	0.039	0.169
V	P	0.26	0.348	–0.016	0.463	1	0.016	–0.172	–0.01	0.294	–0.079	0.262	0.184	–0.302
	B	0.373	0.247	0.123	0.05	1	–0.333	0.085	0.331	0.362	–0.115	0.195	–0.067	0.073
	RG	0.205	–0.102	–0.242	–0.186	1	–0.18	0.282	–0.122	–0.093	–0.07	0.071	–0.006	–0.089
	RH	0.037	0.196	0.166	0.06	1	–0.193	–0.246	–0.183	–0.34	–0.224	0.201	0.383	–0.216
Mn	P	–0.095	0.096	0.525	0.188	0.016	1	0.533	0.304	0.374	0.153	0.316	–0.202	0.537
	B	0.037	0.475	0.295	0.306	–0.333	1	0.222	–0.043	–0.358	0.24	0.646	–0.026	0.166
	RG	0.278	0.269	0.753	0.382	–0.18	1	0.407	0.484	0.637	0.333	0.753	0.364	0.597
	RH	0.353	0.033	0.143	–0.142	–0.193	1	0.367	0.439	0.446	0.687	0.169	0.177	0.035
Fe	P	–0.327	0.01	0.545	0.239	–0.172	0.533	1	0.401	0.148	0.189	0.284	–0.564	0.708
	B	0.122	0.127	–0.038	0.145	0.085	0.222	1	0.36	–0.098	0.381	0.193	–0.177	0.276
	RG	0.011	–0.167	0.134	0.247	0.282	0.407	1	0.4	0.368	–0.102	0.247	0.201	0.416
	RH	0.391	–0.132	–0.078	–0.212	–0.246	0.367	1	0.726	0.363	0.246	–0.08	–0.141	0.336
Co	P	–0.144	0.061	0.265	0.036	–0.01	0.304	0.401	1	0.222	–0.094	0.145	–0.104	0.311
	B	0.532	0.025	0.246	0.172	0.331	–0.043	0.36	1	0.402	–0.09	0.44	–0.026	0.273
	RG	–0.059	0.006	0.133	–0.014	–0.122	0.484	0.4	1	0.348	0.045	0.361	0.272	0.268
	RH	0.284	–0.246	–0.077	–0.114	–0.183	0.439	0.726	1	0.275	0.434	–0.163	–0.002	0.041
Zn	P	0.368	0.529	0.38	0.149	0.294	0.374	0.148	0.222	1	0.274	0.385	0.245	0.113
	B	0.571	0.066	0.402	0.223	0.362	–0.358	–0.098	0.402	1	0.036	0.087	0.001	–0.009
	RG	0.18	–0.076	0.463	0.327	–0.093	0.637	0.368	0.348	1	–0.002	0.42	0.517	0.705
	RH	0.112	–0.156	–0.084	–0.211	–0.34	0.446	0.363	0.275	1	0.206	0.061	0.223	0.428
Rb	P	0.427	0.166	0.259	0.046	–0.079	0.153	0.189	–0.094	0.274	1	0.414	0.375	0.157
	B	–0.18	0.339	0.385	0.031	–0.115	0.24	0.381	–0.09	0.036	1	0.179	–0.023	0.219
	RG	0.581	0.256	0.416	0.085	–0.07	0.333	–0.102	0.045	–0.002	1	0.354	0.341	–0.124
	RH	0.357	0.177	0.138	–0.173	–0.224	0.687	0.246	0.434	0.206	1	0.009	0.338	–0.253
Sr	P	0.55	0.557	0.607	0.392	0.262	0.316	0.284	0.145	0.385	0.414	1	0.084	0.476
	B	0.412	0.625	0.592	0.354	0.195	0.646	0.193	0.44	0.087	0.179	1	0.004	0.23
	RG	0.298	0.48	0.839	0.338	0.071	0.753	0.247	0.361	0.42	0.354	1	0.314	0.505
	RH	0.153	0.597	0.805	0.643	0.201	0.169	–0.08	–0.163	0.061	0.009	1	0.322	0.21
Cs	P	0.657	0.33	–0.178	–0.276	0.184	–0.202	–0.564	–0.104	0.245	0.375	0.084	1	–0.373
	B	0.009	–0.219	0.334	0.336	–0.067	–0.026	–0.177	–0.026	0.001	–0.023	0.004	1	0.399
	RG	0.448	–0.035	0.22	0.031	–0.006	0.364	0.201	0.272	0.517	0.341	0.314	1	0.349
	RH	0.317	0.309	0.18	0.039	0.383	0.177	–0.141	–0.002	0.223	0.338	0.322	1	–0.241
Pb	P	–0.16	0.203	0.76	0.109	–0.302	0.537	0.708	0.311	0.113	0.157	0.476	–0.373	1
	B	0.41	0.143	0.381	0.271	0.073	0.166	0.276	0.273	–0.009	0.219	0.23	0.399	1
	RG	0.117	–0.074	0.429	0.316	–0.089	0.597	0.416	0.268	0.705	–0.124	0.505	0.349	1
	RH	–0.08	–0.086	0.183	0.169	–0.216	0.035	0.336	0.041	0.428	–0.253	0.21	–0.241	1

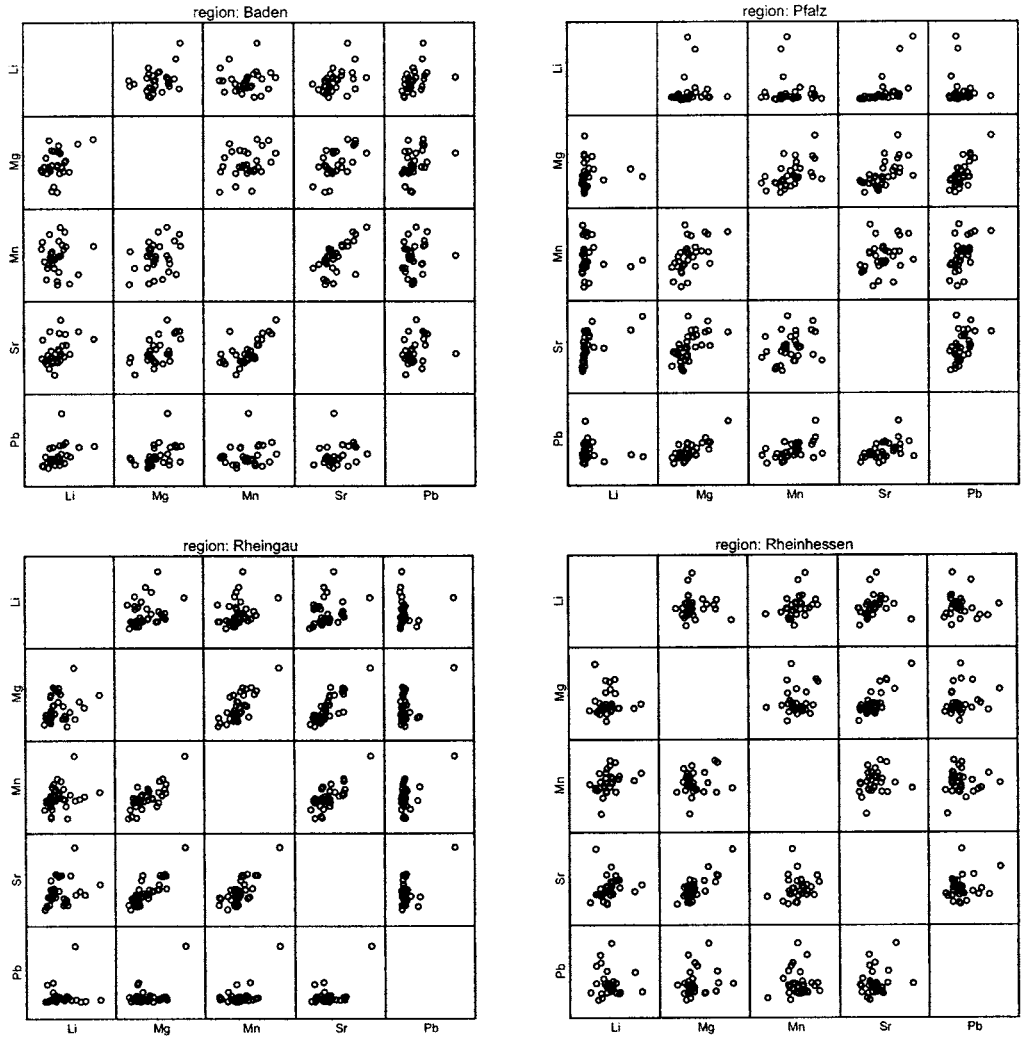

Figure 2. SPLOM (scatter plot matrix) of some elements in the samples from B, P, RG, and RH. The concentrations of the elements are presented against each other for each group.

Table 7. QDA Results Taking All of the Samples (Eigenvalues, Proportions of Total Dispersion Explained with Each Canonical Discriminant Function, and Coefficients of Those Discriminant Functions)

discriminant function	eigenvalue	proportion of total dispersion explained (%)	unstandardized functions coefficients								
			constant	Li	B	Mg	Fe	Zn	Sr	Cs	Pb
1	3.221	59.1	0.00	1.401	–0.179	0.536	0.498	–0.497	–0.138	–0.139	0.185
2	1.865	93.3	–0.00	0.387	0.583	0.587	0.408	0.982	–1.003	–0.663	–0.489
3	0.367	100	0.00	0.539	0.449	–1.298	0.094	0.431	–0.454	0.119	0.861

Pattern Recognition Techniques. Two statistical programs were used: Systat (version 7.0.1, 1997, SPSS Inc.) for the discriminant analysis and SPSS Answer Tree (version 2.1G, SPSS Inc.) for the classification and regression trees.

RESULTS AND DISCUSSION

Quantitative Analysis and Element Content of German White Wines. Recovery studies were carried out on the QCN wine because no wine reference material is available and the accuracy had to be checked. The QCN wine was spiked in triplicate with the multielement homemade stock solution. The mean and the standard deviation were calculated from the recovery values obtained for the three analyses. The recovery percentages are shown in **Table 2**. They lie between 81.5 ± 3.4% (for Zn) and 105.5 ± 9.4% (for Li) with an average value of 92.1 ± 6.3%. For Zn and Co the recovery percentage was quite low, ∼80%. By the application of other measurement techniques such as ICP-OES and TXRF, SF-ICP-MS results for QCN were validated. Results are compiled in **Table 3**. Unfortunately, neither of the techniques was sensitive enough to quantify the elements V, Co, Cs, and Pb. The elements Li and Mg could be compared only with the ICP-OES values because both elements cannot be measured by TXRF. B was not possible to measure by ICP-OES due to instrumental problems. Otherwise, the results for the other elements agreed quite well. Additionally, the QCN sample was used to verify the day-to-day reproducibility (**Table 4**) of the concentration measurement using the actual method over a 3 month period. In general, the RSD was between 2.4% for Sr and 6.3% for Li or B. Typically, a value of ∼4% was reached.

The results were satisfactory and the proposed methodology was used for the analysis of German white wines.

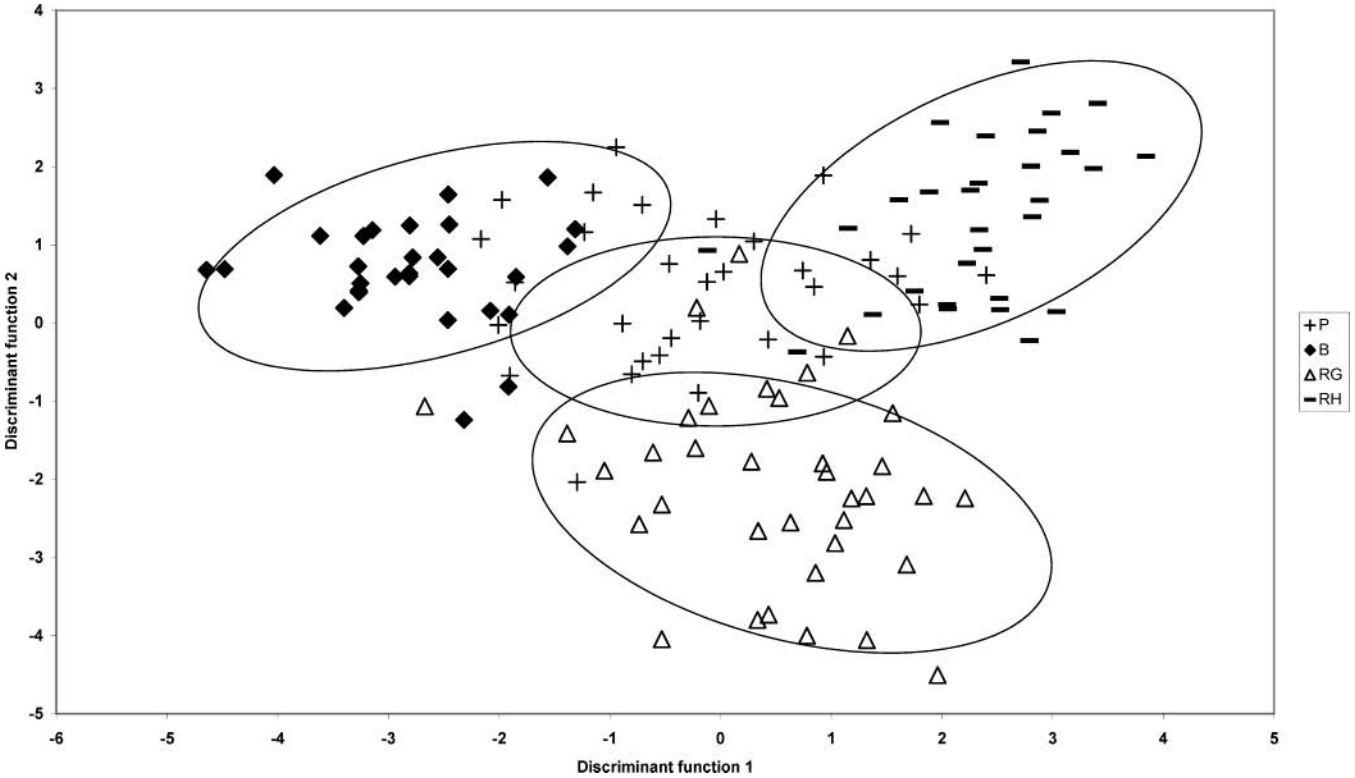

Figure 3. Scatter plot of discriminant functions after quadratic discriminant analysis of the concentrations of Li, B, Mg, Fe, Zn, Rb, Cs, and Pb in German white wines from four wine-growing regions: Baden (B), Rheingau (RG), Rheinhessen (RH), and Pfalz (P).

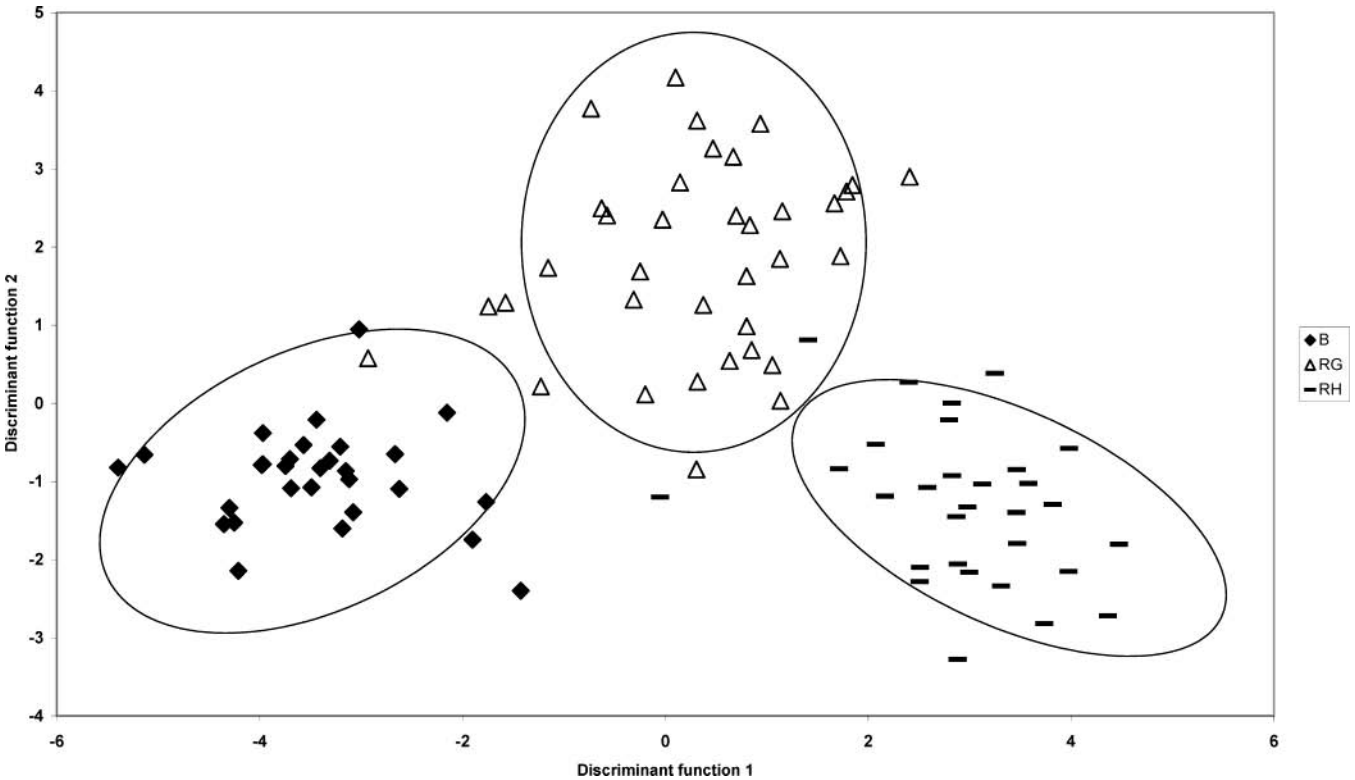

Figure 4. Scatter plot of discriminant functions after quadratic discriminant analysis of the concentrations of Li, B, Mg, Mn, Fe, Zn, Rb, and Cs in German white wines from three wine-growing regions: Baden (B), Rheingau (RG), and Rheinhessen (RH).

The measured metal contents are listed in **Table 5**. Generally the samples from Baden show lowest concentration for the elements Li, Mg, Ca, Fe, Co, Sr, and Pb. Li is present in a concentration up to 10 times and Pb almost 3 times lower than in the other groups. RH samples have the highest content for 10 of the 13 elements, Li, B, Mg, Ca, Mn, Fe, Zn, Rb, Sr, and Pb, but the lowest Cs content. Their low Zn and high Cs contents characterize RG samples. All in all, except for the concentrations of Li in Baden samples and Zn in RG samples, no significant differences in concentrations between groups exist. The range of concentrations within a region vary, with B, Mg, Ca, Mn, Zn, Rb, and Sr quite constant, whereas other elements such as

Li (especially in Pfalz), V, Fe, Co, Cs, and Pb (especially in RG) show a large range.

Pearson correlations between the mean values of the 13 elements in each of the wine-growing regions were investigated in order to find eventual interesting relationships. The correlation coefficients appear in **Table 6**. Significant correlations have been found between Mg and Sr in all of the groups, especially in the RG (0.84) and RH (0.81) groups. Other important correlations are as follows: (a) In the RG samples, Zn (present in much smaller concentration than in the other groups) is correlated with Pb (0.7) and Mn (0.64). Also in these samples there is a correlation between Sr and Ca (0.84) and Mn (0.75) and between Mn and Mg (0.75). (b) In the RH samples there is a correlation between Mg and B (0.68) and Ca (0.67), Fe and Co (0.73), and Mn and Rb (0.69). (c) In the Pfalz samples Pb is correlated with Mg (0.76) and with Fe (0.71), and Li is weakly correlated with Cs (0.66).

Data Treatment and Geographic Identification. Some measured values appeared to be inconsistent with the rest of the data. Such wines were twice measured, and in all cases the trueness of the data was confirmed. Therefore, these apparent outliers were treated as genuine data, and we concluded that they only indicated a non-normal distribution of the data. To prepare the data for discriminant analysis (DA), a log transformation for normalization of tailed distributions was performed. Because the chosen elements are present in wine in very different concentration ranges (from ng/L to mg/L levels), an autoscaling by a *z*-transformation was additionally done. For this, a new variable z was defined according to $z_{ik} = (x_{ik} - \bar{x}_k)/s_k$, where z_{ik} is the value i for the variable k after scaling, x_{ik} is the value i for the variable k before scaling, $\bar{x}_k$ is the mean of the variable k, and s_k is the standard deviation of the variable k. The resulting z is a variable with zero mean and unit standard deviation. Detection limits were included for some elements with values below the LOD for completion of the data set, a prerequisite for chemometric methods. The average values of three measurements were used both for DA analysis and for the classification tree.

Discriminant Analysis. DA is a method that maximizes the variance between classes and minimizes the variance within classes or categories. In LDA, one tries to reduce the number of variables that allow us to distinguish between classes by creating new variables, which are linear combinations of the original variables (*canonical variables* or *canonical discriminant functions*).

LDA is a parametric method, which means that it is based on certain statistical assumptions such as equality between the variance and covariance matrices of the groups to be separated. This means that the classes must have equal dispersion. In **Figure 2**, SPLOMs (scatter plot matrix) of some of the variables in the four groups are represented. Because the sample sizes do not differ markedly (29−36 samples per group), the distribution for each pair of variables should have approximately the same shape and tilt across groups if the equality of covariance assumption holds. This is not the case, and therefore not a linear but a quadratic discriminant model is recommended, QDA (*1*). The main disadvantage of this model is the mathematical sophistication of the resulting classification functions; thus, a previous reduction of variables is mandatory. Using the stepwise modeling option (add or delete one variable at each step), it is possible to identify the most relevant variables.

Once a model has been obtained, it is still necessary to demonstrate the quality of the model, and therefore cross-validation techniques are often applied, two of which are used here. In the *K-fold cross-validation* the whole data set is randomly divided in K subgroups as equal in size as possible, constituting the learning or training sample set (i.e., those with known origin). The model of the specified size is computed K times, each time leaving out one of the K groups from the computations and using the group left out as test set (i.e., those with "unknown" origin which have to be classified). Every subgroup is used $K - 1$ times in the learning set and just once as the test set. Every result computed for each of the K testing sets is then averaged to give the K-fold estimate of the cross-validation. When $k = 1$, one has the leaving-out-one or jackknifed method, and only one case is removed and replaced at a time. This approach, however, may still give an overly optimistic situation. In the so-called *test sample cross-validation*, the samples are again divided into two groups: the learning and the test sample sets. The proportion is usually 70% of the total samples must be part of the learning set and 30% part of the test set. The *recognition (or classification) ability* is the percentage of the members of the training sample set that are correctly classified. The *prediction ability* is the percentage of the members of the test sample set that are correctly classified

Table 8. Classification Quote (Jackknifed Cross-Validation Method) and Recognition and Prediction Ability (Test Set Cross-Validation Method) of German White Wines from Baden, Rheingau, Rheinhessen, and Pfalz by Using a QDA with Only Eight Variables: Li, B, Mg, Fe, Zn, Sr, Cs, and Pb

(A) Regional Distribution of the Tested Wine Samples

sample sets	B	P	RG	RH	total
all samples	29	32	36	30	127
learning set (70%)	22	18	26	18	84
test set (30%)	7	14	10	12	43

(B) Classification Quote (%), All Samples (Jackknifed Method)

origin	no. of samples	predicted group B	P	RG	RH	classification quote (%)
B	29	**26**	2	1	0	90
P	32	2	**24**	4	2	75
RG	36	1	2	**31**	2	86
RH	30	0	3	2	**25**	83
						83 (overall)
total	127	29	31	38	29	

(C) Recognition Ability (%), Learning Set (70%)

origin	no. of samples	predicted group B	P	RG	RH	classification quote (%)
B	22	**22**	0	0	0	100
P	18	2	**14**	1	1	78
RG	26	0	1	**24**	1	92
RH	18	0	1	0	**17**	94
						92 (overall)
total	84	24	16	25	19	

(D) Prediction Ability (%), Test Set (30%)

origin	no. of samples	predicted group B	P	RG	RH	classification quote (%)
B	7	**7**	0	0	0	100
P	14	1	**12**	0	1	86
RG	10	0	2	**7**	1	70
RH	12	0	0	1	**11**	92
						86 (overall)
total	43	8	14	8	13	

Table 9. Geographical Wine Classifications through Element Patterns, Organic Compounds, and Classic Enological Wine Parameters

wines	variables	measuring techniques	pattern recognition techniques	important variables for origin determination	ref
51 German white wines from Mosel and Pfalz	14 elements: Mn, Al, Cu, Co, Ni, Cr, Na, K, Fe, Mg, Ca, Zn, Li, Rb	GC-MS FAAS AAS AES	PCA KNN	Rb, Na, Fe	*6*
	volatile and nonvolatile compounds amino acids				
34 French red wines from three regions	anthocyanin and flavonoid constituents	FAAS FAES	F statistics PCA	Rb, Li, Ca, Mg	*14*
	amino acids		SDA	amino acids: arginine, ethanolamine	
	aromatic alcohols 7 elements: Rb, Li, Ca, Mg, Mn, K, Na				
273 Italian white wines from five Venetian regions	6 elements: Na, K, Ca, Mg, Cl, Li	AAS FAAS	EDGC KNN	*cis*-3-hexen-1-ol	*16*
	aroma compounds	GC-MS	PCA	1-hexanol	
	classic enological parameters		SIMCA CVA	K N-compounds, total phenols	
165 German white, red, and rosé wines from Rheinhessen, Rheingau, Baden, and Saale-Unstrut	16 elements: Al, B, Ba, Ca, Cr, Cu, Fe, Mg, Mn, P, Sb, Si, Sn, Sr, V, Zn	ICP-AES	HCA PCA LDA	(a) B, Ba, Ca, Mg, Pb, Si, Sr, V	*7*, *8*
	58 volatile compounds	HS-SPME-GC-MS		(b) Ba, Si, V, Sr, furfural, 3-methyl-1-pentanol, 4-methyl-1-pentanol	
39 red Spanish wines from Ribeira Sacra and non-Ribeira Sacra	volatile and phenolic compounds 9 elements: Li, Na, K, Rb, Zn, Fe, Mn, Ni, Co	AES AAS	HCA PCA LDA KNN SIMCA BASTEP	Li, Fe, Rb 2 polyphenols	*16*
71 rosé Spanish wines from Ribera del Duero, La Rioja, Valdepeñas and La Mancha	classic enological parameters phenolic compounds color parameters	AAS UV–vis	ANOVA SLDA	12 variables: EtOH, total SO_2, proanthocyanidins, color intensity, tonality, ..., etc.	*17*
	7 elements: Ca, Mg, Zn, Na, K, Fe, Cu			Ca, Fe, Zn, Na	
41 red Italian wines from three areas in the Apulia region	^{1}H NMR spectra (0.5–5 ppm)	^{1}H NMR	HCA	amino acids	*18*
	11 elements: Al, B, Cu, Fe, Zn, Mn, Ba, Na, K, Mg, Ca	ICP-OES	PCA LDA		
	classic enological parameters			heavy metals	
33 wines from Apulia and Slovenia	ash, organic acids ^{1}H NMR spectra (0.5–5 ppm)	ICP-OES	HCA PCA DA	amino acids	*19*
	7 elements: Ba, Cu, Al, Zn, B, Mn, Fe	^{1}H NMR		heavy metals	

by using the decision functions or classification rules achieved with the training sample set.

In a first attempt at stepwise modeling (forward and backward variable selection options), in both cases the variables Ca, V, Mn, Co, and Rb did not contribute significantly to the discrimination model, so they were removed from the model. Thus, a quadratic model with Li, B, Ca, Mg, Fe, Zn, Sr, and Pb was chosen for German wine region discrimination. By applying the QDA, three canonical discriminant functions were obtained. Results are summarized in **Table 7**. The first canonical variable is the quadratic combination of the variables that best discriminate among the groups; the second canonical variable is orthogonal to the first and is the next best combination of variables, and so on. Here the coefficients for each variable in every discriminant function and the discriminant power associated with the three functions are shown. To test these functions for statistical significance, the eigenvalues associated with each one were evaluated. Eigenvalues provide an indication of the relative importance of the function. The first eigenvalue (3.221) is quite large relative to the others, indicating that the first function captures most of the differences between groups; it accounts for 59.1% of the total dispersion of the groups. The second function has an eigenvalue of 1.865 and accounts for 34.2% of the dispersion between groups. The third function has an eigenvalue of 0.367 that indicates an almost insignificantly contribution to the discriminating power. Consequently, because >90% of total variance of the data can be explained by using the first and second canonical functions, only the two first discriminant functions were used for computation. Therefore, the simple discriminant function 1 versus discriminant function 2 scatter point plot is adequate to distinguish between the German wines under investigation. As can be seen in **Figure 3**, every region is more or less well-defined except the Pfalz region, which appears to be overlapping the other three.

With this model a classification rate of 92% was achieved. For testing the goodness of this proposed quadratic model, two different cross-validation procedures were performed. **Table 8B** summarizes the classification quote (%) (following the jackknifed cross-validation with all of the samples) and **Table 8C** the recognition and prediction ability (%) (following the test sample cross-validation procedure) for one of the trials. If the model is built with all of the data and checked with the jackknifed method, the overall classification quote was of 83%. After a close look at each group in **Table 8B,** a good recognition ability of 83% was found for the RH samples. In the case of RG samples 86% were correctly classified, but 7 samples from

Table 10. Geographical Wine Classifications through Elemental Patterns

wines	variables	measuring technique	pattern recognition technique	important variables for origin determination	ref
40 wines from France and USA	17 elements: Al, Ba, B, Cd, Ca, Cr, Cu, Pb, Mg, Mn, Mo, Ni, P, K, Si, Na, Sr	AES	LEAST KNN SIMCA	Al, Ba	*20*
42 white Spanish wines from Rias Baixas, Ribeiro, and Valdeorras	7 elements: Li, Na, K, Rb, Ca, Fe, Mn	AAS AES	PCA LDA KNN SIMCA HCA	Li, Rb	*21*
55 Spanish wines (22 from Somontano, 14 from La Rioja, and 19 from Cariñena) and 57 English wines	65 elements: Li, Be, Al, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Y, Zr, Nb, Mo, Ru, Rh, Pd, Ag, Cd, In, Sn, Sb, Te, I, Cs, Ba, REEs, Hf, Ta, W, Re, Os, Ir, Pt, Au, Hg, Tl, Pb, Bi, Th, U	ICP-MS	QDA	Spanish wines: Cs, Mn, Cd, Cr, Er, Ga, Sr Spanish and U.K. wines: all	*22*
161 Spanish wines from the Canary Islands (Lanzarote, La Palma, Gran Canaria, and Tenerife)	45 variables: REEs, Pb isotope ratios, 23 other metals	ICP-MS	HCA SIMCA	Te, Re, Pt, Au, Tl, Be, Cd, Sn, Sb, Cs, Co, As, Zr, W, V, Ni, Pb, Ti, Cu, Zn, Rb, Sr, Ba	*23*
95 Canadian red wines from Okanagan Valley and Niagara Peninsula	34 trace elements: Li, Be, Mg, Al, P, Cl, Ca, Ti, V, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Br, Rb, Sr, Mo, Ag, Cd, Sb, I, Cs, Ba, La, Ce, Tl, Pb, Bi, Th, U	ICP-MS	PCA LDA	Sr, Rb, Mn, U, Al, V, Zn, Mo, Sb, Co	*24*
83 red Spanish wines from the Canary Islands (Lanzarote, La Palma, Gran Canaria, and Tenerife)	11 elements: K, Ca, Na, Mg, Fe, Cu, Zn, Mn, Sr, Li, Rb	AAS AES	PCA LDA SOMs BP-ANN	Li, Sr, Mg, Ca, K, Fe, Zn	*25*
153 red, white, and rosé Spanish wines from the Canary Islands: Lanzarote, La Palma, Tenerife, El Hierro	39 trace and ultratrace elements: REEs, Te, Re, Be, Cd, Sn, Sb, Cs, Co, As, V, Ni, Ti, Cu, Zn, Rb, Sr, Ba, Pb, Zr, W, Pt, U, Tl	ICP-MS	FA LDA BP-ANN	Sr, Rb, Pb, Be, Ba, Tl, Ti, Au	*26*
68 red wines (grape var. Nebbiolo) from 5 subregions in Cuneo (Piedmont, Italy)	38 elements: REEs, Al, B, Ca, Cr, Cu, Fe, I, K, Li, Mg, Mn, Na, P, Rb, Pb, Si, Sn, Sr, Zn, As, Ba, Be, Cd, Co, Cs, Ga, Ge	ICP-MS	PCA LDA	Mg, Mn, Mo, Si, Ti	*27*

other groups were also included in the RG group; with regard to B samples, they were classified with a 90% success quote, and only 3 samples from other groups were also included as belonging to the B group. The worst classified group was Pfalz (75%); moreover, 7 samples from the other regions were classified as P samples.

Following the second cross-validation method, as can be seen in **Table 8C** for one of the trials, in the learning test a good overall recognition ability of 92% was found; results were especially good for B (100%), RH (94%), and RG (92%). Only in the case of P samples was a poor classification quote of 78% achieved because four Pfalz samples were misclassified. In the test set, the total prediction ability is 86% (**Table 8D**) because 37 of 43 samples were correctly classified. The best prediction abilities were again obtained for B (100%) and RH (92%) samples. All B wines were correctly recognized with only one misclassification case. The worst rates were for P and RG samples because only 70% of the samples were properly classified into the RG group and 86% into the P group. To check the dependence between the results and the test set population, this quadratic model, as it is, was several times performed by varying randomly the members of the test and learning set. Among 10 different trials the overall classification quote for the test set varied between 59 and 86% (on average, 76%): group by group, for the B samples between 56 and 100%; for the P samples, between 50 and 86%; for the RG samples, between 64 and 100%; and finally for the RH samples, between 60 and 100%. The results varied quite significantly and seemed to strongly depend too much on the actual objects in the training set, especially in the case of P samples. It gave not a positive impression about the robustness of this chemometric method when applied to unknown samples.

If the Pfalz samples are eliminated and the classification problem is restricted to the remaining region samples, a new model can be built. In this case and after stepwise modeling, nearly the same eight elements but with Mn instead of Pb were selected for best classification. Again by checking the model after the jackknifed method a total classification quote of 95% was achieved. By the test set cross-validation method and after 10 trials, a recognition ability between 94 and 100% (98% on average) and a prediction ability between 85 and 94% (88% on average) were achieved. The method seems to be quite robust, allowing especially B and RG samples to be well recognized and well predicted. **Figure 4** shows the separation of these regions. Some samples are nevertheless close neighbors to those of the other groups.

The incomplete differentiation between subregions indicates a nonoptimal separation due to the following: (i) The geoclimatic conditions (soils, weather, etc.) of these geographically close wine-growing regions are too similar to reflect this diversity in the wines. (ii) Q.b.A. wines can be sweetened at the end of the winemaking process with a maximum of 15% from a sweet must that comes from the same wine-growing region but can come from another vintage (*12*). If some of these wines were sweetened, its elemental pattern could be slightly altered during this process. (iii) Some samples could be

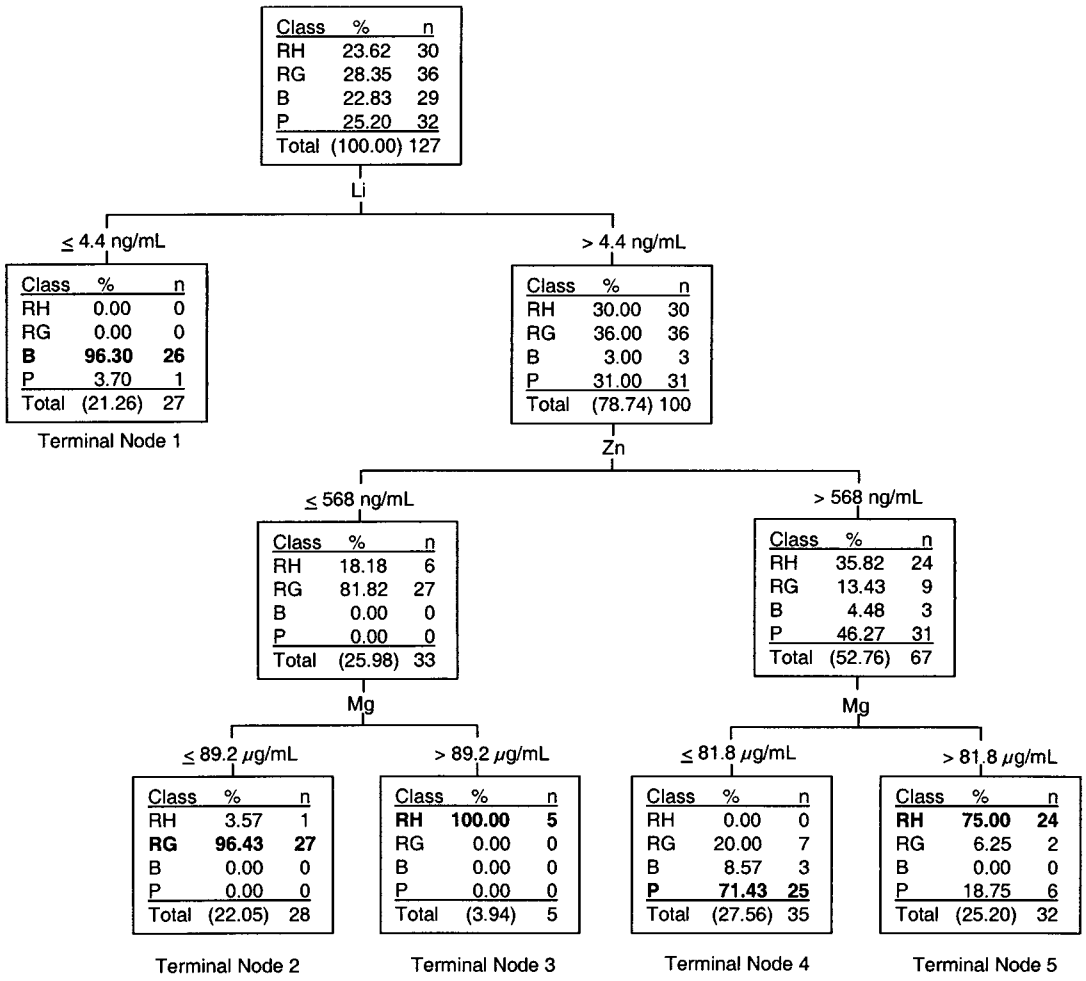

Figure 5. Tenfold cross-validation regression tree for the classification of German white wines from four different wine-growing regions: Baden (B), Rheingau (RG), Rheinhessen (RH), and Pfalz (P). At each node the variable and threshold values are noted. If samples fulfill the condition, the tree is followed on the right branch, which starts from the node. If a sample does not fulfill the condition, the left branch, which starts from the node, is followed. The numbers indicate the composition of each node in terms of samples accepted or rejected. In every terminal node appear the total number of samples classified in this node (also expressed as percentage from the total in parentheses), to which group they belong, how many there are, and how much they mean expressed as percentage from the total samples of this node.

contaminated during the winemaking process or bottling, leading to "abnormal" values for some elements. (iv) Because the wines were purchased from shops, we cannot exclude wrong labeling. (v) The choice of variables is still not the best, and more research should be done for an optimal separation.

The overall soil−plant interaction is highly complex. Greenough et al. (*13*) suggested that the wine element concentration could largely be related to solubility factors of the elements present normally in soils. The uptake from the vine depends on the solubility of these elements that, in turn, depends strongly on the soil characteristics. Between the elements found to be important for the classification of wines in this study, the alkali metals Li and Cs and the alkaline earth metals Mg and Sr, Mn and B form part of the so-called geochemical group of silicate-associated litophile elements in soils. They tend to be soluble in water (as a result of extremely low ionic potentials), and therefore the vine plant takes them up from the soil quite easily. Nevertheless, this uptake depends also on important soil parameters such as pH, E_h (oxidation potential), or the presence of complexing agents such as Cl^-, F^-, or S^{2-}, which vary from soil to soil. Presumably these elements are not affected by winegrowers' activities, and because they are not strongly influenced by the winemaking process as already seen in the accompanying paper, their relation in the wine may exactly reflect that in soil. Nevertheless, some questions are still open such as why geochemically different elements such as Fe, Zn, and Pb seem to be important for origin classification as well. Pb has been for years an important factor for detecting anthropological contamination in wine due mainly to nonappropriate storing materials or environmental pollution. Over the past 20 years the atmospheric fallout of anthropogenic lead has significantly decreased as a result of the phasing-out of lead additives from gasoline. With the new enological practices Pb contamination sources have been almost totally reduced as well, and the Pb content that is normally found in wine (10−50 ng/mL) is mainly due to the primary content from the soil. The usefulness of the nine elements (Li, B, Mg, Mn, Fe, Zn, Sr, Cs, and Pb) found here to be important for the wine provenance determination in multivariate analysis has been many times reported in studies with wines from all around the world (see **Tables 9 and 10**). Normally they belong—between other parameters—to sets of selected variables that are best suited for the wine classification problem.

Trees. The tree method does not use a formal statistical method but an algorithm model. Following the method of regression or classification trees (C&RT), it is possible to establish a graphical tree that presents all of the information in a very simple and straightforward way. The tree begins with

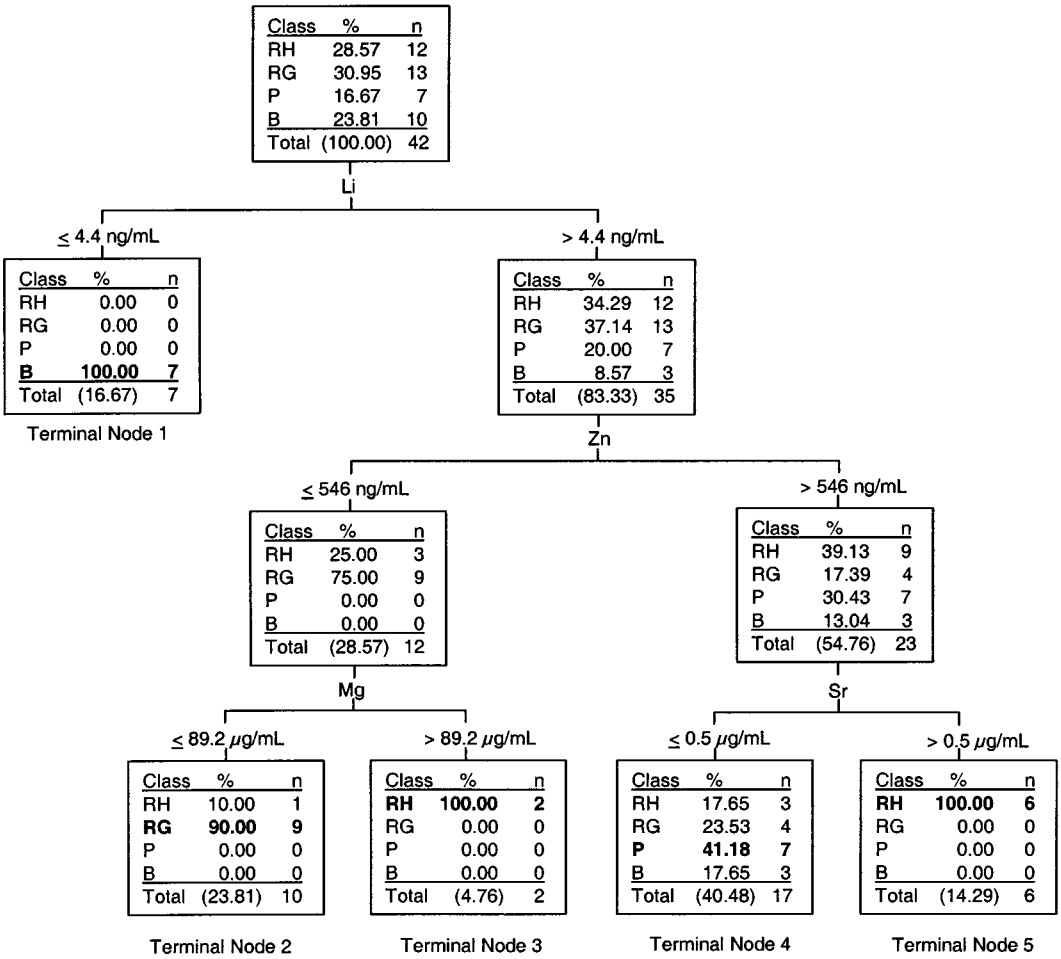

Figure 6. Regression tree for the classification of German white wines from four different wine-growing regions: Baden (B), Rheingau (RG), Rheinhessen (RH), and Pfalz (P). Here the testing sample set (30% total samples population) is classified according to the regression tree previously performed with the learning set samples.

one node, the top decision node or *root node*, where all of the observations are collected. This branches into two groups depending on the values of one of the variables. This variable is among all the one that shows the best discriminating power. Every one of this group is again split into two new groups depending on the value of the same variable or of a second one with better discriminating power. This procedure is repeated with every new group, the procedure resulting in the branching of the root node to many *decision nodes* or *split nodes*. At the end of each branch there is a density display showing the distribution of observations at that point. In this case the tree is binary because each node is split into only two subsamples. *Terminal nodes* are points on the tree beyond which no further decisions are made. It is desired that they are "pure", namely, containing no misclassified observations or objects within the class.

As in DA it is necessary to check the goodness of the classification. According to the first procedure, a 10-fold cross-validation was performed, and the resulting TREE is shown in **Figure 5**. The first variable with the strongest discriminant power is Li, and this metal divides the original group into two subgroups: one is already a terminal node and the other is later split again. In terminal node 1 are present samples with [Li] ≤ 4.4 ng/mL. This group contains 27 wine samples (26 B + 1 P). This terminal node is classified as characteristic for B samples. The second subgroup is divided again according to the Zn content into two new subgroups. The one characterized for concentrations of Zn of <568 ng/mL is again split into two subgroups according to the Mg content. Both subgroups are terminal nodes. In terminal node 2 there are 28 wines (27 RG + 1 RH). This terminal node is classified as characteristic for the RG samples. All of the samples belonging to terminal node 2 have [Li] > 4.4 ng/mL, [Zn] ≤ 568 ng/mL, and [Mg] ≤ 89.2 μg/mL. The subgroup with higher concentrations of Zn is finally split according to the Mg concentration into terminal nodes 4 and 5. Terminal node 4 has 35 wines (25 P + 7 RG + 3 B). Thus, this terminal node is classified as characteristic for P samples. All of the samples belonging to terminal node 4 have [Li] > 4.4 ng/mL, [Zn] > 568 ng/mL, and [Mg] ≤ 81.8 μg/mL. Terminal node 5 contains 32 wines (24 RH + 6 P + 2 RG). Therefore, this node is characterized as typical for RH samples with [Li] > 4.4 ng/mL, [Zn] > 568 ng/mL, and [Mg] > 81.8 μg/mL. On the whole **(Table 11B)**, according to the 10-fold cross-validation method, 84% of the samples were correctly classified. The wines worst classified were those belonging to the RG group, only 75%, although only 4% of the samples identified as RG are not really RG samples. The best classified wines were B samples, with a correct classification rate of 90%. Moreover, of all wines classified as B samples, only 4% were from other regions.

As described for the second procedure, the samples were divided into two sets: the learning sample set with 85 objects (70% of the population) and the test sample set with 42 objects (30% of the population). With the learning set a classification rule resulting in five nodes was derived. This model was tested for application with the test sample, and a new tree (**Figure 6**)

Table 11. Classification Quote (10-Fold Cross-Validation Method) and Recognition and Prediction Ability (Test Set Cross-Validation Method) of German White Wines from Baden, Rheingau, Rheinhessen, and Pfalz by Using a Classification Tree with Only Three Variables: Li, Mg (or Sr), and Zn

(A) Regional Distribution of the Tested Wine Samples

sample set	B	P	RG	RH	total
all samples	29	32	36	30	127
learning set (70%)	19	25	23	18	85
test set (30%)	10	7	13	12	42

(B) Recognition Ability (%), 10-Fold Cross-Validation, All Samples

origin	no. of samples	predicted group B	P	RG	RH	% correctly classified
B	29	**26**	3	0	0	90
P	32	1	**25**	0	6	78
RG	36	0	7	**27**	2	75
RH	30	0	0	1	**29**	97
						84 (overall)
total	127	27	35	28	37	

(C) Recognition Ability (%), Learning Set (70%)

origin	no. of samples	predicted group B	P	RG	RH	% correctly classified
B	19	**19**	0	0	0	100
P	25	1	**22**	0	2	88
RG	23	0	6	**16**	1	70
RH	18	0	2	0	**16**	89
						86 (overall)
total	85	20	30	16	19	

(D) Prediction Ability (%), Test Set (30%)

origin	no. of samples	predicted group B	P	RG	RH	% correctly classified
B	10	**7**	3	0	0	70
P	7	0	**7**	0	0	100
RG	13	0	4	**9**	0	70
RH	12	0	3	1	**8**	67
						74 (overall)
total	42	7	17	10	8	

was obtained. **Table 11C** summarizes the recognition and the prediction ability of the tree in the learning set and test set, respectively; 86% of the learning sample set is correctly classified, but a lower success quote of only 74% was found for the test set. Outstanding in this test set is that although 100% P samples were correctly classified, 59% of the P sample group members come from the other regions. In practice this means that it is possible to exclude wines from the P region with relatively high certainty, but a positive identification is very uncertain. For the other three groups, ~70% of the samples were correctly classified, and they make up 90−100% the samples of each predicted group.

A comparison of the trees obtained from the first and second methods shows that the most important difference is that Sr appears instead of Mg when the population is split into nodes 4 and 5. Both methods give more or less the same total classification rates, 84 and 86%, respectively, and therefore Sr or Mg can be used alternatively. This is confirmed if one takes into account the positive Pearson correlation ($r = 0.71$) between the two variables for the P and RH samples.

All in all, a tree shows that Li, Mg (or Sr), and Zn are very important elements for the origin determination of German wines. All of these elements were also selected by QDA.

It must be taken into account that the results provided by this study cannot necessarily be generalized. An extrapolation to all wines from the German regions under study is beset with problems: (i) the relatively small population of objects in each of the groups (the number of samples used in each case might not be large enough to give a good statistical representation of the real situation) and (ii) the commercially purchased samples may not have a totally unambiguous origin. Also, it would be interesting to investigate other vintages to check the year-to-year concentration variation of these elements of interest in the same German regions. Otherwise, the capabilities of this method cannot be extrapolated to other vintages.

Other multivariate techniques such as artificial neuronal networks (ANN) could be applied to look for a better classification among regions.

Other wine characteristics such as Sr isotope ratios, classical enological components, or some organic components are promising tools that could improve up to 100% the differentiation between sub-regions.

CONCLUSIONS

A simultaneous determination of 13 elements (in ultratrace, trace, and major range) in wines was easily performed by SF-ICP-MS. The simple pretreatment of the sample (1:20 dilution) and the use of In as internal standard lead to a reliable quantitative analysis.

Simple inspection of elemental concentrations could not be used to differentiate the growing origin; however, multivariate chemometrical methods were able to detect similarities between samples according to their origin. Use of discriminant analysis and decision tree both successfully classified wines relative to their origin. By applying a QDA model with eight elements [one major element (Mg), four trace elements (B, Fe, Zn, and Sr), and three ultratrace elements (Li, Cs, and Pb)], on average 76% of unknown samples (test samples set) could be classified correctly. The worst case was for the Pfalz samples, with a success of 50% only. If the Pfalz samples are eliminated, a new model with the same elements but with Mn instead of Pb can correctly classify 88% of unknown samples.

By using a decision tree with only three elements (Li, Mg, and Zn), a similar success quote of 74% for the classification of unknown samples was achieved. Again, Pfalz was the worst classified region, because although all Pfalz samples were correctly identified, 59% of all samples assigned to be a Pfalz wine came from other regions.

The robustness of the QDA model was not the very best, so a tree is more recommended and even allows an easier wine classification due to the lower number of elements necessary.

In general, chemometric methods applied here for recognition of elemental pattern seemed to be a useful tool if determination of wine origin is questioned. A surprisingly high correlation between element content and wine origin was found that allows a distinction to be made between white wines even from geographically very close wine-growing regions in Germany. From the strictly legal point of view, none of those methods seem to be sufficiently reliable to be used for checking of falsely labeled wines because unknown samples could be classified with an accuracy of ~75% only. Nevertheless, this method can be applied as a screening tool and as a complement to other methods.

ABBREVIATIONS USED

^{1}H NMR, nuclear magnetic resonance spectrometry; A.P.Nr., official control number of a certified wine in German system (*Amtliche Prüfungsnummer*); AAS, atomic absorption spectrometry; AES, atomic emission spectrometry; ANN, artificial neural networks; ANOVA, analysis of variance; B, Baden; BASTEP, stepwise Bayesian analysis; BP-ANN, back-propagation artificial neural networks; C&RT, classification and regression trees; CVA, canonical variate analysis; DA, discriminant analysis; EDGC, Euclidean distances from group centroids; FAAS, flame atomic absorption spectrometry; F statistics, Fisher statistics; GC-MS, gas chromatography−mass spectrometry; HCA, hierarchical cluster analysis; HS-SPME-GC-MS, headspace solid-phase microextraction gas chromatography−mass spectrometry; ICP-OES, inductively coupled plasma optical emission spectrometry; KNN, *K* nearest-neighbor analysis; QDA, quadratic discriminant analysis; LDA, linear discriminant analysis; LOD, limit of detection; LOQ, limit of quantification; NAA, neutron activation analysis; P, Pfalz; PCA, principal component analysis; PP, polypropylene; Q.b.A., wine with certified origin in German system (*Qualitätswein bestimmter Anbaugebiete*); QCN, quality control Nierstein; RG, Rheingau; RH, Rheinhessen; SLDA, stepwise linear discriminant analysis; SF-ICP-MS, sector field inductively coupled plasma mass spectrometry; SIMCA, soft independent modeling of class analogies; SPLOM, scatter plot matrix; TXRF, total reflection X-ray fluorescence.

ACKNOWLEDGMENT

We are grateful to Alex von Bohlen (ISAS-Dortmund) for supplying the statistical programs and interesting discussions.

LITERATURE CITED

(1) Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. Supervised pattern recognition. In *Chemometrics: a Textbook, Data Handling in Science and Technology*; Elsevier Science Publilshers: New York, 1988; Vol. 2, pp 385−412.

(2) Varmuza, K. *Pattern Recognition in Chemistry*; Lecture Notes in Chemistry 21; Springer-Verlag: Berlin, Germany, 1980.

(3) Massart, D. L.; Kaufman, L. *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis, Chemical Analysis*; Wiley: New York, 1983; Vol. 65.

(4) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and Regression Trees*; Wadsworth International Group: Belmont, CA, 1984.

(5) Siegmund, H.; Bächmann, K. Die Lagezuordnung von Weinen durch Bestimmung des Spurenelementmusters. *Z. Lebensm. Unters. Forsch.* **1977**, *164*, 1−7.

(6) Maarse, H.; Slump, P.; Tas, A. C.; Schaefer, J. Classification of wines according to type and region based on their composition. *Z. Lebensm. Unters. Forsch.* **1987**, *184*, 198−203.

(7) Thiel, G. Atomemissionsspektrometrische und multivariat-statistische Untersuchungen zur Bestimmung der Herkunft von Weinen. Ph.D. thesis, Friedrich-Schiller Universität Jena, 1997.

(8) Danzer, K.; De la Calle Garcia, D.; Thiel, G.; Reichenbächer, M. Classification of wine samples according to origin and grape varieties on the basis of inorganic trace analysis. *Ann. Lab.* **1999**, *31*, 26−33.

(9) Ambrosi, H. *Weinlexikon*; Falken Verlag: München, Germany, 1996.

(10) Foulkes, C.; Broadbent, M. *Wein Enzyklopädie. Die Weinregionen der Welt*; Christian Verlag: München, Germany, 1995 (original version: *Larousse Encyclopedia of Wine*; Hamlyn: 1994.)

(11) Crosby, N. T.; Day, J. A.; Hardcastle, W. A.; Holcombre, D. G.; Treble, R. D. Charting. In *Quality in the Analytical Chemistry Laboratory, Analytical Chemistry by Opening Learning*; Wiley: Chichester, U.K., 1995; pp 152−163.

(12) German wine regulation, revised form, from May 14, 2002: article 40, paragraph 1.

(13) Greenough, J. D.; Longerich, H. P.; Jackson, S. E. Trace element concentrations in wines by ICP-MS: evidence for the role of solubility in determining uptake by plants. *Can. J. Appl. Spectrosc.* **1996**, *41*, 76−80.

(14) Etievant, P.; Schlich, P.; Bouvier, J.-C.; Symonds, P.; Bertrand, A. Varietal and geographic classification of French red wines in terms of elements, amino acids and aromatic alcohols. *J. Sci. Food Agric.* **1988**, *45*, 25−41.

(15) Moret, I.; Scarponi, G.; Cescon, P. Chemometric characterization and classification of five Venetian white wines. *J. Agric. Food Chem.* **1994**, *42*, 1143−1153.

(16) Rebolo, S.; Peña, R. M.; Latorre, M. J.; García, S.; Botana, A. M.; Herrero, C. Characterisation of Galician (NW Spain) Ribeira Sacra wines using pattern recognition analysis. *Anal. Chim. Acta* **2000**, *417*, 211−220.

(17) Pérez-Magariño, S.; Ortega-Heras, M.; González-San José, M. L. Multivariate classification of rosé wines from different Spanish protected designations of origin. *Anal. Chim. Acta* **2002**, *458*, 187−190.

(18) Brescia, M. A.; Caldarola, V.; De Giglio, A.; Benedetti, D.; Fanizzi, F. P.; Sacco, A. Characterization of the geographical origin of Italian red wines based on traditional *and* nuclear magnetic resonance spectrometric determinations. *Anal. Chim. Acta* **2002**, *458*, 177−186.

(19) Brescia, M. A.; Košir, I. J.; Caldarola, V.; Kidrič, J.; Sacco, A. Chemometric classification of Apulian and Slovenian wines using ^{1}H NMR and ICP-OES with HPICE data. *J. Agric. Food Chem.* **2003**, *51*, 21−26.

(20) Kwan, W.-O.; Kowalski, B. R.; Skogerboe, R. K. Pattern recognition analysis of elemental data. Wines of *Vitis vinifera* cv. Pinot noir from France and the United States. *J. Agric. Food Chem.* **1979**, *27*, 1321−1326.

(21) Latorre, M. J.; Garcia-Jares, C.; Medina, B.; Herrero, C. Pattern Recognition Analysis Applied to Classification of Wines from Galicia (Northwestern Spain) with Certified Brand of Origin. *J. Agric. Food Chem.* **1994**, *42*, 1451−1455.

(22) Baxter, M. J.; Crews, H. M.; Dennis, M. J.; Goodall, I.; Anderson, D. The determination of the authenticity of wine from its trace element composition. *Food Chem.* **1997**, *60*, 443−450.

(23) Barbaste, M.; Medina, B.; Sarabia, L.; Ortiz, M. C.; Pérez-Trujillo, J. P. Analysis and comparison of SIMCA models for denominations of origin of wines from the Canary Islands (Spain) builds by means of their trace and ultratrace metals content. *Anal. Chim. Acta* **2002**, *472*, 161−174.

(24) Taylor, V. F.; Longerich, H. P.; Greenough, J. D. Multielement analysis of Canadian wines by inductively coupled plasma mass spectrometry (ICP-MS) and multivariate statistics. *J. Agric. Food Chem.* **2003**, *51*, 856−860.

(25) Díaz, C.; Conde, J. E.; Estévez, D.; Pérez Olivero, S. J.; Pérez Trujillo, J. P. Application of multivariate analysis and artificial neural networks for the differentiation of red wines from the Canary Islands according to the islands of origin. *J. Agric. Food Chem.* **2003**, *51*, 4303−4307.

(26) Pérez-Trujillo, J.-P.; Barbaste, M.; Medina, B. Chemometric study of bottled wines with denomination of origin from the Canary Islands (Spain) based on ultra-trace elemental content determined by ICP-MS. *Anal. Lett.* **2003**, *36*, 679−697.

(27) Marengo, E.; Aceto, M. Statistical investigation of the differences in the distribution of metals in Nebbiolo-based wines. *Food Chem.* **2003**, *81*, 621−630.

Received for review October 1, 2003. Revised manuscript received February 19, 2004. Accepted February 26, 2004. We express our gratitude to the German Bundesministerium für Bildung und Forschung, the Ministerium für Schule, Wissenschaft und Forschung des Landes Nordrhein-Westfalen, and the Deutsche Forschungsgemeinschaft (DFG, Contract Ja 611/ 2-1).

JF035120F