ISSN 1061-9348, Journal of Analytical Chemistry, 2018, Vol. 73, No. 2, pp. 195–206. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © V.O. Titarenko, A.A. Khalafyan, Z.A. Temerdashev, A.A. Kaunova, A.G. Abakumov, 2018, published in Zhurnal Analiticheskoi Khimii, 2018, Vol. 73, No. 2, pp. 141–152.

ARTICLES

Identification of the Varietal and Regional Origin of Red Wines by Classification Analysis

V. O. Titarenko, A. A. Khalafyan, Z. A. Temerdashev*, A. A. Kaunova, and A. G. Abakumov

Department of Chemistry and High Technology, Kuban State University, Krasnodar, 350040 Russia

**e-mail: temza@kubsu.ru*

Received March 17, 2017; in final form, August 7, 2017

Abstract—We examined 144 samples of red dry wines of varieties Cabernet and Merlot, produced in the territory of the main wine-making enterprises of the geographical zones of Krasnodar krai, to consider the possibility of determining the regional and varietal identity of red wines. The concentration of micro- and macroelements in the wines was determined by inductively coupled plasma atomic emission spectrometry. Chemometric studies were performed in the Statistica package environment, using discriminant analysis and classification trees. Adequate mathematical models were proposed for wines to identify the grape variety and the region where it grows. A software product was developed that automates the necessary calculations.

Keywords: identification of the regional and varietal origin of red wines, concentration of elements in wines, classification trees, discriminant analysis, logistic regression

DOI: 10.1134/S1061934818020132

One of the most difficult analytical tasks in determining the authenticity of a wine is the determination of its geographical origin. Single parameters for assessing the quality of wines are not fully sufficient to determine the conformity of the product to its labeling [1]. Analytical approaches based on the determination of mineral and isotopic composition, the study of spectral characteristics, the detection of phenolic and volatile compounds using various methods of analysis are being developed to determine the authenticity and regional origin of wines, as well as the changes occurring during their adulteration [2–4].

The identification of criteria for the authenticity and regional origin of wines is based on obtaining a large body of data on the test objects and then processing it with the use of chemometric algorithms to find the hidden relationships between the variables studied and to evaluate the contribution of each of them to the identification power of the statistical model [2, 3, 5–7]. The combination of modern test methods with the capabilities of chemometric methods offers a more reliable identification of wines according to their regional origin. Information on the elemental composition of wines can be useful also for controlling the technological process and in combination with chemometric methods of processing data enables determining of the origin of wines [1, 8]. Since the concentrations of metals in wines produced in different European wine-growing regions is quite different (Table 1), it seems possible to identify the regional belonging of wines on this basis.

The mineral composition of wines can be affected by various factors; therefore, elements that are least dependent on external effects should be selected for their identification. Sr, Mn, Mg, Li, Co, Rb, B, Cs, Zn, Al, Ba, Si, Pb, and Ca are often used as such elements [10]. Metals are contained in wines in wide ranges of concentrations: macroelements (Ca, K, Na, and Mg) are present in amounts from 10 to 1000 mg/L; minor elements (Al, Fe, Cu, Mn, Rb, Sr, and Zn) in amounts of 0.1–10 mg/L; and microelements (Ba, Cd, Co, Cr, Li, Ni, Pb, V, etc.) may occur in the range of 0.1–1000 μg/L [9].

In the European Union, quality wines are usually given trademarks, such as protected designation of origin, designation of place of origin, or protected geographical indication [11]. Wines with the protected geographical indication and the protected name of a place of origin are now produces in Russia as well. On July 1, 2013, Russia introduced the State Standard *GOST R 55242–2012: Wines of Protected Geographical Indications and Wines of Protected Names of Origin*, which divides them into three categories.

Of greater interest to consumers and connoisseurs is the category, which includes wines of protected geographical indications and protected appellations of origin at the place of production, produced from grapes of one variety, having a certain variability in relation to organoleptic qualities, in particular, taste characteristics. One of the reasons for the natural taste variability of wines is the difference in growing conditions of grapes from which wine is produced, like cli-

Table 1. Concentration (mg/L) of some metals in wines of a number of countries [9]*

Element	Czech Republic	France	Germany	Italy	Spain
K	553–3056	265–426	480–1860	–	338–2032
Na	2.0–110	7.7–14.6	6–25	–	3.5–300
Ca	40–210	65–161	58–200	88–151	12–241
Mg	7.8–138	55–96	56–105	53–60	50–236
Al	–	0.56–1.27	–	–	0.57–14.3
Cu	–	N.d.–0.48	0.02–0.71	–	N.d.–3.1
Fe	0.9–5.2	0.81–2.51	0.4–4.2	–	0.4–17.4
Mn	0.28–3.26	0.63–0.96	0.5–1.3	–	0.1–5.5
Rb	0.56–1.20	0.64–0.72	0.2–2.9	0.50–9.90	0.1–5.3
Sr	0.34–0.53	0.22–0.47	0.12–1.28	0.40–1.16	0.28–1.50
Zn	–	0.44–0.74	0.3–1.5	–	N.d.–4.63
Ba	0.09–0.12	0.025–0.24	0.04–0.26	0.07–0.14	0.01–0.35
Cd	–	N.d.–0.0002	–	–	N.d.–0.019
Co	N.d.–0.018	0.004–0.011	0.004–0.005	0.003–0.006	N.d.–0.040
Cr	0.032–0.037	0.030–0.057	0.022–0.078	0.023–0.034	0.025–0.029
Li	0.015–0.052	0.008–0.036	0.005–0.043	–	0.002–0.13
Ni	–	N.d.–0.052	–	–	0.005–0.079
Pb	–	0.006–0.023	–	–	0.001–0.043
V	0.020–0.054	0.06–0.23	0.01–0.14	–	0.026–0.043

* N.d., not detected.

matic conditions, microelement composition of soil, and technology of cultivation of grapes. The problem of determining of the microelemental "image" of the grapes from which the wine was produced is thus of not only scientific, but also practical interest. On the other hand, wines from one variety of grapes have a high variability in taste characteristics, which are caused by the difference in the places of grapes growing, wine production technologies, vintage, etc. Therefore, in some cases, when there are certain organoleptic similarities between the wines from different grape varieties, for example, color, tart or sour taste, it is important to identify the grape variety by the trace element composition. In fact, the solution of the formulated problem is reduced to the task of determining the concentration of microelements in the sample of an unblended wine of the region (zone) of grapes growing from that it was made and the variety of grapes.

In the present paper, the possibility of using classification analysis to identify the zones of grapes and grape varieties by the concentration of trace elements in two widely known brands of red wines—Cabernet and Merlot of the Kuban origin—is considered in this paper.

EXPERIMENTAL

Objects of research. To consider the possibility of determining the regional and varietal identity of red wines, we examined 144 samples of red dry wines of varieties Cabernet (76 samples) and Merlot (68 samples), produced in the territory of the main wine-making enterprises of the geographical zones of the Krasnodar krai: Zaporozhskoe, Kuban'-Vino, Fanagoriya, Mil'strim–Chernomorskie vina, Kavkaz, Abrau-Dyurso, Gelendzhik, Myskhako, Somel'e, Sauk-Dere, and Soyuz-Vino.

Materials and methods. The following State Standard Samples (GSOs) were used for the preparation of metal solutions: GSO 7780-2000 (Li), GSO 8062-94 (Na), GSO 7767-2000 (Mg), GSO 7854-2000 (Al), GSO 7771-2000 (K), GSO 7772-2000 (Ca), GSO 7205-95 (Ti), GSO 7834-2000 (Cr), GSO 8056-94 (Mn), GSO 8032-94 (Fe), GSO 7784-2000 (Co), GSO 7785-2000 (Ni), GSO 7836-2000 (Cu), GSO 8053-94 (Zn), GSO 7035-93 (Rb), GSO 7783-2000 (Sr), GSO 7874-2000 (Cd), GSO 7760-2000 (Ba), and GSO 7778-2000 (Pb). Other reagents used in the work were of cp grade or better.

The concentration of micro- and macroelements in wines was determined by inductively coupled plasma atomic emission spectrometry with (ICP AES) using an iCAP-6000 spectrometer (Thermo Scien-

Table 2. Frequency of distribution of wine samples by zones

Grape variety	Anapskaya subzone	Yuzhno-Predgornaya zone	Chernomorskaya zone	Tamanskaya subzone	Total
Cabernet	17	13	18	28	76
Merlot	23	12	0	33	68
Total	40	25	18	61	144

tific); the wine samples were diluted prior to measurements according to the method described in [12].

Chemometric studies were performed in the Statistica package environment [13], using classification trees and discriminant analysis. The classification models were built for each variety of wines.

RESULTS AND DISCUSSION

The main industrial plantations of grapes in the Krasnodar krai are located in five areas of cultivation: Temryukskii region (Taman peninsula, Taman subzone), Anapskii region (Anapskaya subzone), Chernomorskaya zone (regions Gelendzhik and Novorossiysk), Krymskii region (Yuzhno-Predgornaya zone), and Novokubanskii region [14]. The frequency distribution of these 144 samples of Cabernet and Merlot wines by zones and varieties is given in Table 2.

Analysis of the average concentration of elements in wines (Tables 3 and 4) showed that the composition of a wine is significantly dependent on the variety and region of production of grapes. For example, the wines from the Anapskaya subzone are distinguished among others by a high concentration of iron, and in wines from the Yuzhno-Predgornaya zone, high concentrations of Ba, Ti, and V are observed, while in wines from the Tamanskaya subzone contain higher concentrations of Na, Mg, and Rb. The concentrations of many elements differ significantly for the Cabernet wines, and in the samples from the Yuzhno-Predgornaya zone, the concentrations of Li, Na, Al, Ca, Fe, and Sr are the lowest. The Merlot wine from the Yuzhno-Predgornaya zone is characterized by the smallest concentration of Al, Ca, Fe, and Li. As a rule, standard deviations do not exceed half the average value. This indicates a small spread in the concentration of elements; hence, the average values are an informative characteristic of the concentration of metals in wine. Exceptions are Cu, Li, Ni, Na, Rb, and Ti; only in three cases the standard deviations are greater than the mean values, namely, for copper in Cabernet from the Anapskaya subzone and nickel in Cabernet and Merlot from the Yuzhno-Predgornaya zone.

Surely, the visualization of tables with mean values and standard deviations is not a basis for ascertaining the differences in the concentration of elements; therefore, the statistical significance of the difference in the concentration of metals in groups of wine samples was estimated by variance analysis. For most elements, the difference in the average for groups in the varietal and regional origin is statistically significant. For example, the significance levels p of the criterion of the least significance difference (LSD) for magnesium are presented in Table 5. In the “Variety–zone” column, the names of groups of wine samples and their numerical designations (1, 2, ..., 7) are displayed, indicating their varietal and regional origin. The numerical designations of the groups and the average values (M) of the magnesium concentration in these groups are the top of the other columns of the table. If the significance level p in the cell is smaller than 0.05 (marked in bold), then the difference in the magnesium concentration in the groups at the intersection of the corresponding row and column of the table is statistically significant. For example, the difference in the magnesium concentration is statistically significant in the samples of Cabernet from the Anapskaya subzone and one from the Yuzhno-Predgornaya zone; in Cabernet and Merlot from the Tamanskaya subzone; etc. Table 5 is symmetric with respect to empty cells along the diagonal.

Analysis of data on the concentrations of elements in wines of different batches from the same manufacturer showed that there is no significant fluctuation in the chemical composition (Table 6).

Thus, the data of a multielement analysis of the samples under study suggest that the elemental images of wines produced in different geographical zones differ from each other and are similar in the distribution of metals in wines of different producers within the same zone. Therefore, it is useful to create statistical images of the profiles of Cabernet and Merlot wines of the studied geographical zones, which will ensure the differentiation of the wine according to regional characteristics and varietal identity. To solve this problem, discriminant analysis and classification trees were used.

In the terminology of classification analysis, a set of 144 samples of wines with the concentrations of trace elements is called a training sample. The similarity and difference between the objects of the training sample is determined by the distances between them as points in the multidimensional space of attributes characterizing the objects. To solve the problem of determining the belonging of objects to certain groups, objects should form clusters, that is, groups of homogeneity that combine similar objects, in the space of attributes. In the problem being solved, the objects are

Table 3. Mean values and standard deviations (in parentheses) of the concentrations (μg/L) of the elements in the samples of Cabernet wines produced in different geographical regions of the Krasnodar krai

Element	Anapskaya	Yuzhno-Predgornaya	Chernomorskaya	Tamanskaya
Al	761.09 (389.48)	668.69 (353.97)	1073.59 (322.26)	776.95 (222.41)
Ba	91.85 (32.83)	160.34 (69.28)	92.77 (46.28)	100.09 (29.14)
Ca	60042.51 (8432.57)	54707.23 (10123.92)	59863.89 (5570.95)	65516.38 (11564.37)
Cu	112.35 (128.01)	68.87 (35.59)	108.80 (42.73)	65.34 (31.38)
Fe	8098.52 (3017.29)	3397.99 (1150.77)	4188.36 (972.85)	3657.09 (915.35)
K	659037.51 (96738.75)	1190176.9 (394839.15)	1064055.56 (171436.59)	983958.33 (233820.96)
Li	18.48 (4.68)	13.36 (7.54)	24.57 (23.76)	27.97 (14.97)
Mg	74036.62 (12846.51)	116978.08 (21053.00)	96779.63 (27030.09)	141678.27 (59802.26)
Mn	956.44 (318.78)	1585.00 (362.23)	1096.30 (164.85)	1338.22 (435.68)
Na	21433.84 (9680.21)	31760.50 (27186.37)	35267.59 (13784.32)	33698.77 (8590.17)
Ni	24.48 (10.35)	57.93 (65.14)	33.02 (31.55)	20.79 (8.66)
Rb	513.66 (359.92)	976.82 (289.66)	817.38 (236.04)	1514.88 (352.21)
Sr	1206.78 (207.42)	1323.51 (504.23)	1533.48 (719.59)	1269.65 (283.94)
Ti	6.98 (3.32)	31.43 (22.11)	16.20 (10.81)	8.09 (3.07)
Zn	366.20 (140.53)	685.00 (221.94)	756.61 (151.34)	481.00 (143.97)

the wine samples, and the space of attributes is the concentration of 15 trace elements in them.

First, we built a model of discrimination of wine samples by seven selected groups of varietal and regional origin. An important part of discriminant analysis is canonical analysis, which involves the construction of discriminant functions, also called canonical roots. Canonical roots enable objects of n-dimensional space to be transferred to a two-dimensional space, preserving the order of distances between objects. The wine samples are shown in Fig. 1 with various icons in the coordinate system of Root 1 (ordinate axis)—Root 2 (abscissa axis), depending on the zone and variety. The wine samples of seven groups form clusters localized on the plane, which indicates their some nonuniformity. For example, the Cabernet wines from the Tamanskaya subzone and the Chernomorskaya zone form clusters sufficiently isolated from clusters of other groups. However, the remaining groups are located close enough, mixing with each other. The Cabernet and Merlot wines from the Yuzhno-Predgornaya zone, the Cabernet wines from the Chernomorskaya zone, and the Merlot wines from the Anapskaya subzone are localized in the same part of the plane. The absence of a clear cluster structure of different groups of wines suggests possible errors in the classification of wines by the concentrations of elements. The calculations showed that the total number of incorrectly classified wines was 13, and the total percentage of correct classification was only 91%. This means that if regional and varietal belonging of wine samples not belonging to the training sample is deter-

Table 4. Mean values and standard deviations (in parentheses) of the concentrations (μg/L) of the elements in the samples of Merlot wines produced in different geographical regions of the Krasnodar krai

Element	Anapskaya	Yuzhno-Predgornaya	Tamanskaya
Al	1063.51 (380.89)	925.93 (178.33)	1306.99 (537.01)
Ba	106.47 (41.36)	148.36 (40.08)	162.59 (42.33)
Ca	60116.55 (5279.89)	55629.57 (5823.01)	63211.99 (8701.37)
Cu	53.44 (44.76)	62.19 (33.22)	46.79 (33.03)
Fe	13248.18 (4214.22)	4073.10 (606.44)	4882.86 (1661.30)
K	728426.31 (74086.08)	1264236.4 (369588.21)	695290.34 (106987.53)
Li	25.65 (14.11)	16.83 (7.26)	28.68 (10.90)
Mg	71221.02 (6552.85)	114459.33 (21714.79)	158520.16 (28073.14)
Mn	1181.17 (400.52)	1409.75 (499.32)	1519.22 (398.21)
Na	20698.30 (10063.63)	51380.22 (28798.60)	62561.01 (30852.02)
Ni	28.94 (16.75)	91.81 (100.97)	111.87 (95.55)
Rb	563.32 (531.08)	2063.40 (1786.28)	6157.27 (2706.61)
Sr	1242.35 (171.84)	1453.34 (375.60)	1389.06 (292.37)
Ti	11.11 (11.03)	32.62 (7.39)	27.75 (14.64)
Zn	369.83 (83.22)	754.84 (137.99)	526.47 (139.69)

mined by the discrimination model, the probability of an erroneous classification may be even higher. In this connection, it was decided to carry out the decomposition of the problem to be solved: first, to determine the grape variety from which the wine was produced and then to identify the region of its growth.

The model for determining the varietal origin of Cabernet and Merlot wines. To establish the fact that the grape varieties Cabernet and Merlot are unique in the concentration of trace elements and that their statistical image can be created, the method of classification trees is the most suitable, which enables the selection of the most informative attributes characterizing the objects and the derivation of a classification tree based on them [13]. Classification trees is a method of classification analysis that predicts the belonging of objects to a particular class depending on the corresponding values of the attributes characterizing the objects. The attributes are called independent variables (the concentration of elements), and the variable that indicates the belonging of objects to classes is called dependent (variety of wines). The number of vertices and the accuracy of the prediction of the classification tree were optimized for the aggregate of elements Al, Ba, Fe, K, Ni, and Rb by the classification and regression trees (C&RT) method using the fast algorithm for classification trees (FACT) with a fixed percentage of incorrectly classified observations of 0.035 (Fig. 2). The node or the vertex of the tree is represented as a rectangle. The node number is written in the upper left corner of the rectangle, above which the number of wine samples assigned to this node is dis-

Table 5. Significance levels of the criterion of the least significant difference* for magnesium

Variety and zone	LSD; variable: Mg; differences are noted, significant at the level of $p < 0.05$						
	1 $M = 74037$	2 $M = 117000$	3 $M = 96780$	4 $M = 141700$	5 $M = 71221$	6 $M = 114500$	7 $M = 158500$
Cabernet Anapskaya 1		**0.00**	**0.04**	**0.00**	0.79	**0.00**	**0.00**
Cabernet Yuzhno-Predgornaya 2	**0.00**		0.09	**0.03**	**0.00**	0.85	**0.00**
Cabernet Chernomorskaya 3	**0.04**	0.09		**0.00**	**0.01**	0.15	**0.00**
Cabernet Tamanskaya 4	**0.00**	**0.03**	**0.00**		**0.00**	**0.02**	**0.04**
Merlot Anapskaya 5	0.79	**0.00**	**0.01**	**0.00**		**0.00**	**0.00**
Merlot Yuzhno-Predgornaya 6	**0.00**	0.85	0.15	**0.02**	**0.00**		**0.00**
Merlot Tamanskaya 7	**0.00**	**0.00**	**0.00**	**0.05**	**0.00**	**0.00**	

* For all varieties, the significance level p is given accurate to the second decimal place, which in the third and/or fourth digits is nonzero and statistically significant.

Table 6. Variations of the chemical composition of wine samples by the example of Cabernet (Fanagoriya)

Element	Concentration of the element in 120 wine samples in different batches (by date of their production), μg/L				
	February 7, 2014 ($n = 18$)	February 24, 2014 ($n = 24$)	April 12, 2014 ($n = 28$)	March 24, 2014 ($n = 23$)	February 19, 2014 ($n = 27$)
Li	25 ± 3	24 ± 3	25 ± 3	24 ± 3	21 ± 3
Na	$(4.0 \pm 0.4) \times 10^4$	$(3.4 \pm 0.3) \times 10^4$	$(3.5 \pm 0.3) \times 10^4$	$(3.4 \pm 0.3) \times 10^4$	$(3.5 \pm 0.3) \times 10^4$
Mg	$(9 \pm 1) \times 10^4$	$(9 \pm 1) \times 10^4$	$(9.3 \pm 1.2) \times 10^4$	$(9.2 \pm 1.2) \times 10^4$	$(10 \pm 1) \times 10^4$
Al	789 ± 166	814 ± 171	822 ± 173	804 ± 169	557 ± 117
K	$(8.5 \pm 0.9) \times 10^5$	$(8.5 \pm 0.9) \times 10^5$	$(8.5 \pm 0.9) \times 10^5$	$(8.5 \pm 0.9) \times 10^5$	$(8.4 \pm 0.9) \times 10^5$
Ca	$(6.3 \pm 0.8) \times 10^4$	$(6.2 \pm 0.7) \times 10^4$	$(6.3 \pm 0.8) \times 10^4$	$(6.2 \pm 0.7) \times 10^4$	$(8 \pm 1) \times 10^4$
Ti	7.3 ± 1.5	6.0 ± 1.3	5.3 ± 1.1	6.7 ± 1.4	7 ± 1
Mn	1067 ± 117	1071 ± 118	1074 ± 118	1058 ± 116	1278 ± 141
Fe	4583 ± 779	4381 ± 745	4361 ± 741	4292 ± 730	3511 ± 597
Ni	17 ± 6	14 ± 5	16 ± 6	16 ± 6	13 ± 5
Cu	53 ± 5	62 ± 6	58 ± 6	55 ± 6	46 ± 5
Zn	411 ± 45	431 ± 47	437 ± 48	441 ± 49	415 ± 46
Rb	1516 ± 258	1505 ± 256	1528 ± 260	1508 ± 256	1512 ± 257
Sr	1013 ± 101	1034 ± 103	1033 ± 103	1020 ± 102	1288 ± 129
Ba	100 ± 17	89 ± 13	87 ± 10	85 ± 13	111 ± 19

Mean values ± standard deviations are given.

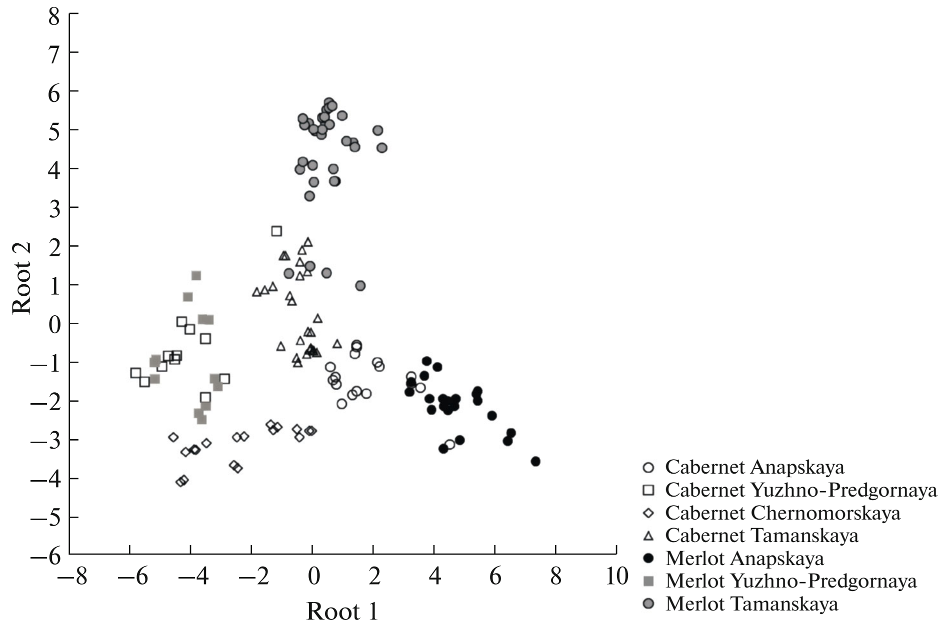

Fig. 1. Scattering diagram of canonical values for wine samples, depending on the zone and name.

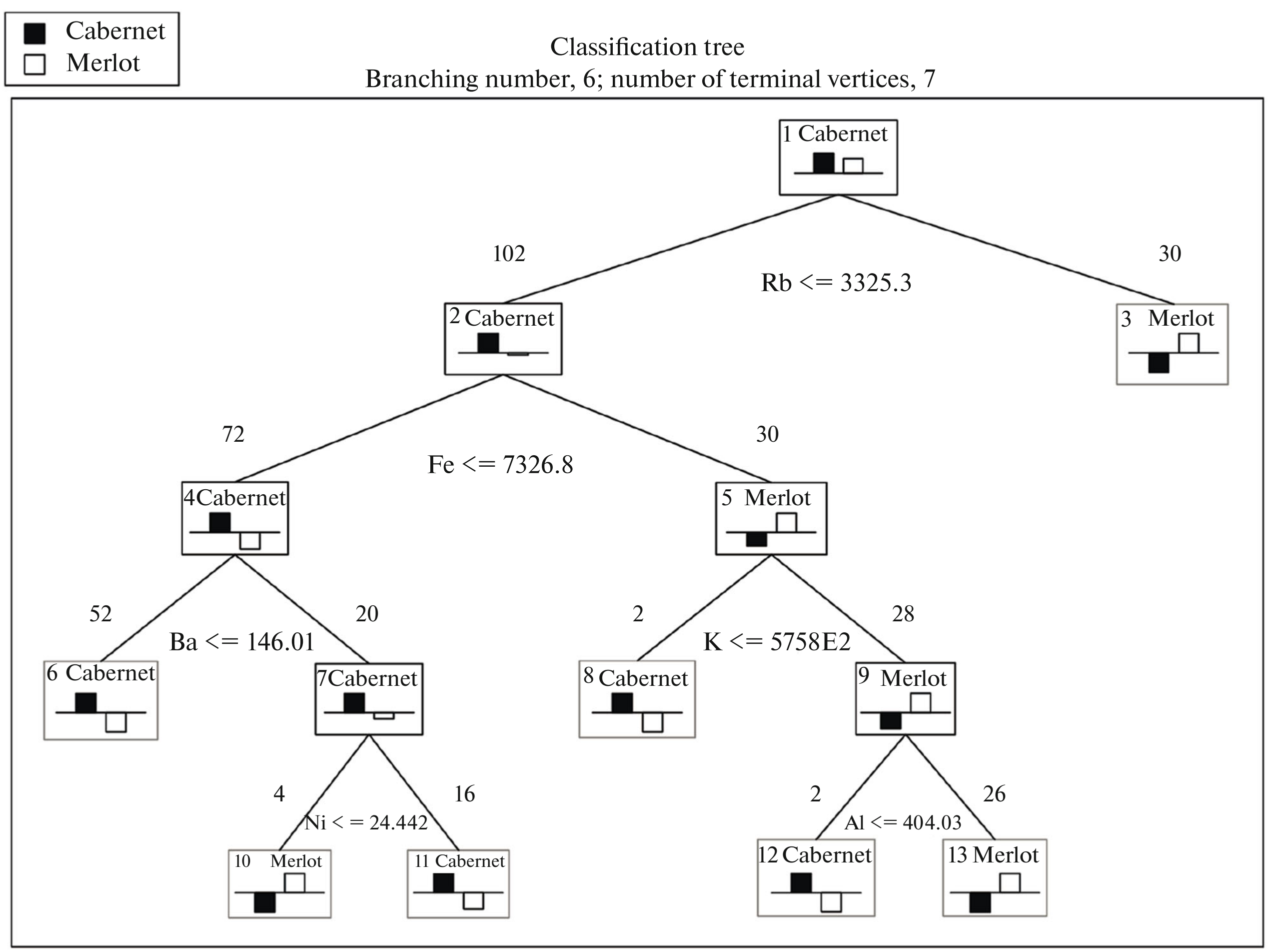

Fig. 2. Graph of the classification tree for the variety of wines.

Table 7. Results of discriminant analysis of 75 Cabernet samples by the element composition*, **

Element	Wilks' lambda	Particular lambda	F-exclusion (3, 59)
Mg	0.0061	0.350	36.382
Fe	0.0055	0.388	30.970
K	0.0043	0.489	20.520
Rb	0.0043	0.496	19.926
Ti	0.0039	0.535	17.029
Zn	0.0038	0.556	15.655
Na	0.0036	0.586	13.843
Ba	0.0034	0.629	11.569
Mn	0.0031	0.671	9.604
Sr	0.0031	0.681	9.192
Li	0.0030	0.697	8.528
Ca	0.0029	0.721	7.599
Al	0.0029	0.738	6.960

* For all elements, the significance level is $p < 0.000$, but in the fourth and fifth digits, it is different from zero and statistically significant.

** Wilks' lambda is 0.00214; the value of the Fisher test (39, 175) is 31.349.

played. Within each vertex of the tree graph, column diagrams are depicted representing classes (groups of wine samples).

The adequacy of the derived classification model is confirmed by the fact that only one Merlot sample, classified as Cabernet, was mistakenly attributed to the observable classes of the training sample.

The classification algorithm is rather simple and consists of six steps:

Step 1. If Rb ≤ 3325.3, then go to Step 2; otherwise, the sample is Merlot.

Step 2. If Fe ≤ 7326.8, then go to Step 3; otherwise, to Step 5.

Step 3. If Ba ≤ 146.01, the sample is Cabernet; otherwise, go to Step 4.

Step 4. If Ni ≤ 24.442, the sample is Merlot; otherwise, it is Cabernet.

Step 5. If K ≤ 575778.7, the sample is Cabernet; otherwise, go to Step 6.

Step 6. If Al ≤ 404.03, the sample is Cabernet; otherwise, it is Merlot.

Example 1. Determine the grape variety at the following values of the element concentrations: Al = 809, Ba = 57, Fe = 3667, K = 857500, Ni = 4, and Rb = 911.

Step 1. Since Rb ≤ 3325.3, then go to Step 2.

Step 2. Since Fe ≤ 7326.8, then go to Step 3.

Step 3. Since Ba ≤ 146.01, then the sample is Cabernet.

Construction of models to identify the zone for Cabernet wines. The results of discriminant analysis of Cabernet samples by the element composition are presented in Table 7. The information part of the table shows the results of discriminant function analysis: the value of the integral quality criterion for the Wilks' lambda discrimination procedure is 0.00214, the approximate value of the Fisher test F (39, 175) = 31.349 with degrees of freedom 39 and 175, and the significance level $p < 0.000$. The high adequacy of the classification model is indicated by the Wilks' lambda value close to zero. Since the significance level is less than 0.05 (the accepted level of significance of statistical hypotheses [13]) for all elements, the Wilks' lambda criterion is statistically significant. The contribution of elements to the procedure of discrimination of wines by zones can be estimated by the values of statistics. The larger the lambda value (the first column) and the smaller the value of the particular lambda (second column), the greater the contribution of the studied element to the classification procedure. Analysis of the data of Table 7 shows (the elements are ordered in descending order of the lambda, and therefore, in descending order of their role in the wine classification) that the most informative element is Mg, followed by Fe, K, Rb, etc. The statistical significance of the role of elements in the model of wine discrimination can be judged from the values of the Fisher criterion (F-exclusion) and its significance level (p). If all the p levels (in Table 7, they are given accurate to 3 digits) were greater than 0.05, then the elements in the discrimination model would become statistically insignificant, and the model would be unusable for determining the zones of grapes growth, respectively. The variables, for which the level of significance is greater than 0.05, should be deleted from the model. For this reason, there are no elements of Cu and Ni in Table 7. Analysis of the classification matrix showed that the total percentage of correct identification of the zones was 100%.

For the classification of wine samples, classification functions can be used, where the object is assigned to the group for which the function has the greatest value. Denote the classification functions for the Yuzhno-Predgornaya, Chernomorskaya, Anapskaya, and Tamanskaya subzones, respectively, $F_{\text{Anap.}}$, $F_{\text{Yuzhn.}}$, $F_{\text{Chern.}}$, and $F_{\text{Tam.}}$; then, it is easy to compose mathematical expressions for them according to the table of coefficients generated by the program:

$$\begin{aligned} F_{\text{Anap.}} &= -69.411 + 0.007\text{Al} + 0.168\text{Ba} + 0.001\text{Ca} \\ &+ 0.002\text{Fe} + 0.0001\text{K} - 0.352\text{Li} + 0.00002\text{Mg} \\ &- 0.052\text{Mn} - 0.0008\text{Na} + 0.025\text{Rb} \\ &+ 0.027\text{Sr} - 0.0008\text{Ti} + 0.023\text{Zn}, \end{aligned} \tag{1}$$

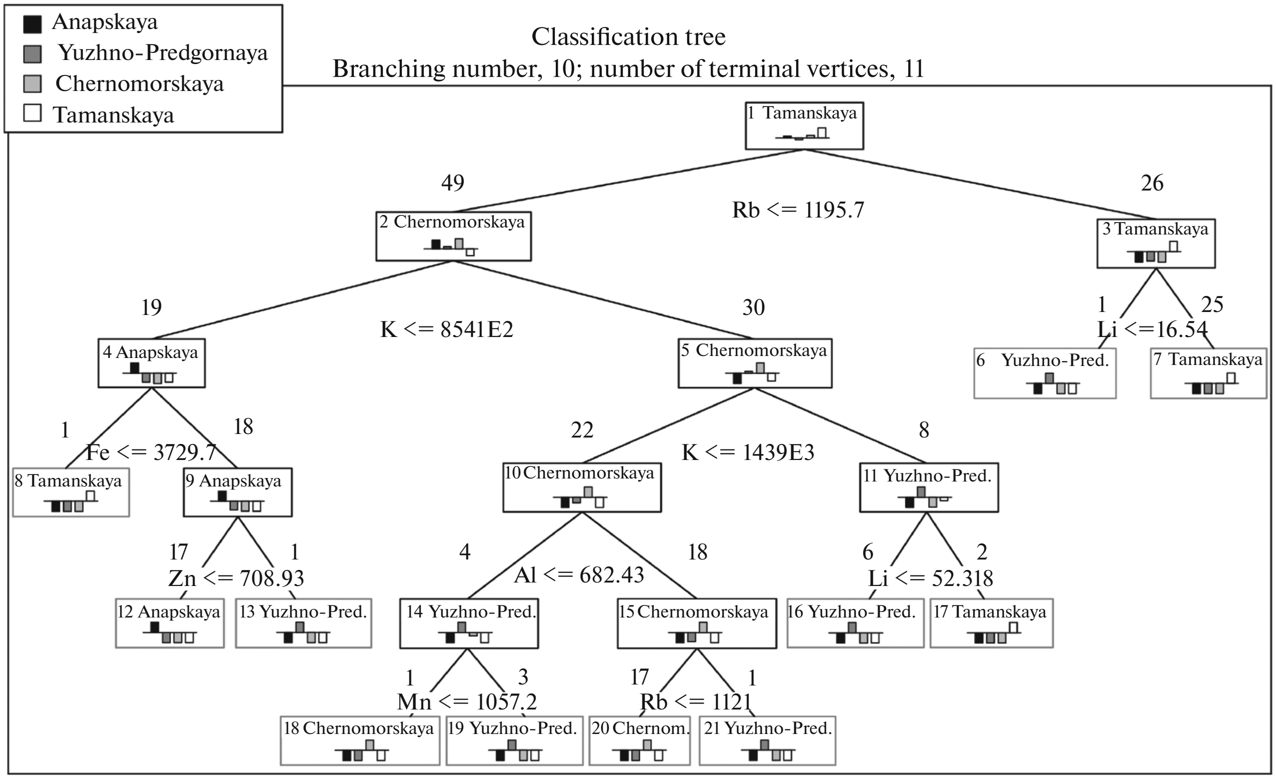

Fig. 3. Classification tree of Cabernet wines for four production zones.

$$\begin{aligned} F_{\text{Yuzhn.}} = -93.66 + 0.022\text{Al} + 0.365\text{Ba} \\ + 0.001\text{Ca} - 0.003\text{Fe} + 0.0001\text{K} \\ - 0.379\text{Li} - 0.0004\text{Mg} - 0.04\text{Mn} - 0.001\text{Na} \\ + 0.042\text{Rb} + 0.031\text{Sr} - 0.497\text{Ti} + 0.027\text{Zn}, \end{aligned} \tag{2}$$

$$\begin{aligned} F_{\text{Chern.}} = -109.100 + 0.031\text{Al} + 0.306\text{Ba} \\ + 0.001\text{Ca} - 0.002\text{Fe} + 0.0001\text{K} - 0.47\text{Li} \\ - 0.0002\text{Mg} - 0.074\text{Mn} - 0.001\text{Na} + 0.055\text{Rb} \\ + 0.046\text{Sr} - 1.204\text{Ti} + 0.069\text{Zn}, \end{aligned} \tag{3}$$

$$\begin{aligned} F_{\text{Tam.}} = -112.241 + 0.021\text{Al} + 0.091\text{Ba} \\ + 0.002\text{Ca} - 0.001\text{Fe} + 0.00005\text{K} + 0.203\text{Li} \\ - 0.0001\text{Mg} - 0.064\text{Mn} - 0.001\text{Na} + 0.059\text{Rb} \\ + 0.01\text{Sr} - 0.357\text{Ti} + 0.032\text{Zn}. \end{aligned} \tag{4}$$

Example 2. Determine the region for a Cabernet wine with the following values of the element concentrations: Al = 809, Ba = 57, Ca = 69833, Fe = 3667, K = 857500, Li = 10, Mg = 74400, Mn = 979, Na = 19550, Rb = 911, Sr = 1097, Ti = 9.7, and Zn = 713.

Calculations by Eqs. (1)–(4) give $F_{\text{Anap.}} = 91.299$, $F_{\text{Yuzhn.}} = 88.427$, $F_{\text{Chern.}} = 115.479$, and $F_{\text{Tam.}} = 99.068$. The greatest value is $F_{\text{Chern.}} = 115.479$; therefore, a sample of Cabernet wine belongs to the Chernomorskaya zone of grapes growing.

To enhance the confidence of identifying the regions where grape was grown, we used the alternative statistical method described above, classification trees [13]. By sequential search of all possible combinations of levels of independent variables, we reached a compromise between the complexity of the tree and the laboriousness of the classification procedure by the classification and regression trees (C&RT) method using the fast algorithm for classification trees (FACT) with a fixed fraction of incorrectly classified observations of 0.04, with the same cost for wine classification errors over all zones under study. The conducted researches made it possible to form the optimal wine classification tree (Fig. 3), constructed using the concentrations of seven trace elements: Rb, K, Li, Fe, Zn, Al, and Mn. The tree graph consists of 11 terminal vertices and does not involve further branching.

According to the graph in Fig. 3, it is easy to compile an algorithm that predicts the zone where grape was grown for an arbitrary sample of Cabernet wine, based on the values of the concentrations of elements.

Step 1. If Fe ≤ 116, then go to Step 3; otherwise, to Step 2.

Step 2. If Li ≤ 16.54, then the sample is from the Yuzhno-Predgornaya zone; otherwise, it is from the Tamanskaya subzone; completion of the algorithm.

Step 3. If K ≤ 854100, then go to Step 4; otherwise, to Step 6.

Step 4. If Fe ≤ 3729.7, the sample is from the Tamanskaya subzone; completion of the algorithm; otherwise, go to Step 5.

Step 5. If Zn ≤ 708.93, then the sample is from the Anapskaya subzone; otherwise, it is from the Yuzhno-Predgornaya zone; completion of the algorithm.

Step 6. If K ≤ 1439 400, then go to Step 7; otherwise, to Step 10.

Step 7. If Al ≤ 682.43, then go to Step 8; otherwise, to Step 9.

Step 8. If Mn ≤ 1057.2, then the sample is from the Chernomorskaya zone; otherwise, it is from the Yuzhno-Predgornaya zone; completion of the algorithm.

Step 9. If Rb ≤ 1121, then the sample is from the Chernomorskaya zone; otherwise, it is from the Yuzhno-Predgornaya zone; completion of the algorithm.

Step 10. If Li ≤ 52.318, then the sample is from the Yuzhno-Predgornaya zone; otherwise, it is from the Tamanskaya subzone; completion of the algorithm.

Example 3. Determine the region for the wine sample from the previous example, taking into account that Zn = 713, K = 857500, Fe = 3667, Mn = 979, Rb = 911, Li = 10, and Al = 809.

Step 1. Since Rb ≤ 116, then go to Step 3.

Step 3. Since Rb ≤ 854100, then go to Step 6.

Step 6. Since K ≤ 1439 400, then go to Step 7.

Step 7. Since Al ≤ 682.43, then go to Step 9.

Step 9. Since Rb ≤ 1121, the sample is from the Chernomorskaya zone; completion of the algorithm.

Construction of models to identify the zone for Merlot wines. For Merlot wines, the adequate classification models were also constructed. In the model of discrimination of wines with elemental composition including Zn, Fe, Mg, K, Mn, Al, Ni, Rb, Cu, Ba, Na, Ti, Sr, and Ca, the value of the Wilks lambda was 0.00212. Compared with the element composition for discriminating Cabernet wines, Ni and Cu are included in the model, but Li is excluded. The number of classification errors for the training sample is 0.

The classification functions for the Anapskaya and Tamanskaya subzones and the Yuzhno-Predgornaya zone are, respectively,

$$\begin{aligned} F_{\text{Anap.}} = {} & -170.971 - 0.014\text{Al} + 0.233\text{Ba} + 0.004\text{Ca} \\ & - 0.123\text{Cu} + 0.006\text{Fe} + 0.00006\text{K} + 0.00008\text{Mg} \\ & - 0.022\text{Mn} + 0.0001\text{Na} - 0.196\text{Ni} + 0.006\text{Rb} \\ & - 0.039\text{Sr} + 0.721\text{Ti} - 0.044\text{Zn}, \end{aligned} \tag{5}$$

$$\begin{aligned} F_{\text{Yuzhn.}} = {} & -184.863 - 0.038\text{Al} + 0.305\text{Ba} \\ & + 0.004\text{Ca} + 0.074\text{Cu} - 0.0008\text{Fe} + 0.00009\text{K} \\ & + 0.0004\text{Mg} + 0.0003\text{Mn} + 0.0005\text{Na} - 0.15\text{Ni} \\ & - 0.001\text{Rb} - 0.071\text{Sr} + 0.118\text{Ti} + 0.058\text{Zn}, \end{aligned} \tag{6}$$

$$\begin{aligned} F_{\text{Tan.}} = {} & -220.038 - 0.049\text{Al} + 0.429\text{Ba} + 0.005\text{Ca} \\ & - 0.145\text{Cu} + 0.0001\text{Fe} + 0.000008\text{K} + 0.001\text{Mg} \\ & + 0.029\text{Mn} + 0.0004\text{Na} + 0.004\text{Ni} - 0.0008\text{Rb} \\ & - 0.078\text{Sr} - 0.156\text{Ti} - 0.075\text{Zn}. \end{aligned} \tag{7}$$

Example 4. Determine the region for a Merlot wine with the following values of the element concentrations: Al = 949, Ba = 141, Ca = 59818, Fe = 3749, K = 523461, Mg = 175842, Mn = 1706, Na = 54281, Ni = 37, Rb = 8987, Sr = 1405, Ti = 18, and Zn = 652. Calculations by Eqs. (5)–(7) yield $F_{\text{Anap.}}$ = 161.386, $F_{\text{Yuzhn.}}$ = 151.427, and $F_{\text{Tam.}}$ = 210.93. The greatest value is $F_{\text{Tam.}}$ = 210.93; therefore, a sample of Merlot wine belongs to the Tamanskaya subzone of grapes growing.

By analogy with the Cabernet wines, we constructed an optimal tree for classification of the zones of growth of grapes with a small number of terminal vertices (4) (Fig. 4).

The classification of the training sample was carried out using the constructed tree with respect to the concentrations of only three elements—Rb, Fe, and Ca—with significance ranks, respectively, 90, 68, and 21. According to the branching conditions and the tree graph shown in Fig. 4, it is easy to compile a classification algorithm.

Step 1. If Fe ≤ 8985.7, then go to Step 2; otherwise, the sample is from the Anapskaya subzone; completion of the algorithm.

Step 2. If K ≤ 955878.5, then go to Step 3; otherwise, the sample is from the Yuzhno-Predgornaya zone; completion of the algorithm.

Step 3. If Rb ≤ 818.6, then the sample is from the Yuzhno-Predgornaya zone; otherwise, it is from the Tamanskaya subzone; completion of the algorithm.

Example 5. Determine the region for the wine sample from the previous example, taking into account that Fe = 3749, K = 523461, and Rb = 8987.

Step 1. Since Fe ≤ 8985.7, then go to Step 2.

Step 2. Since K ≤ 955878.5, then go to Step 3.

Step 3. Since Rb ≤ 818.6, the sample is from the Tamanskaya subzone; completion of the algorithm.

To automate the calculations based on the proposed classification algorithms, a software product is developed that has a simple and convenient interface; the user does not need knowledge of classification analysis methods. The start window of the program is shown in Fig. 5. The procedure for classifying wines by the grape variety is carried out by discriminant analysis; the region of origin for each of the wines is determined by two methods (discriminant analysis and classification trees).

It is necessary to enter concentrations of elements in the test wine sample into the corresponding fields. Certainly, in the implementation of classification methods, it is desirable to adhere to the concentration

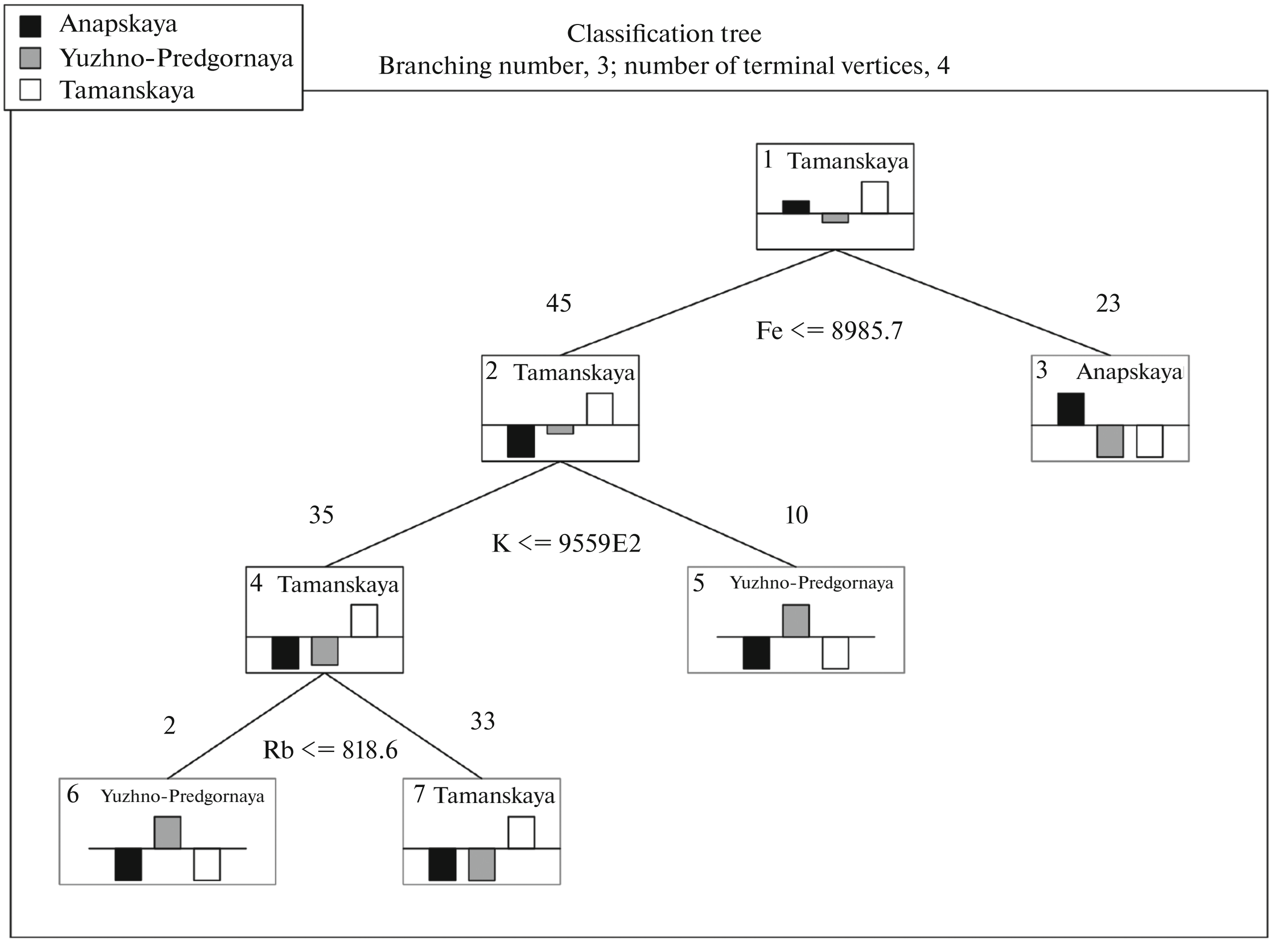

Fig. 4. Classification tree of Merlot wines for three production zones.

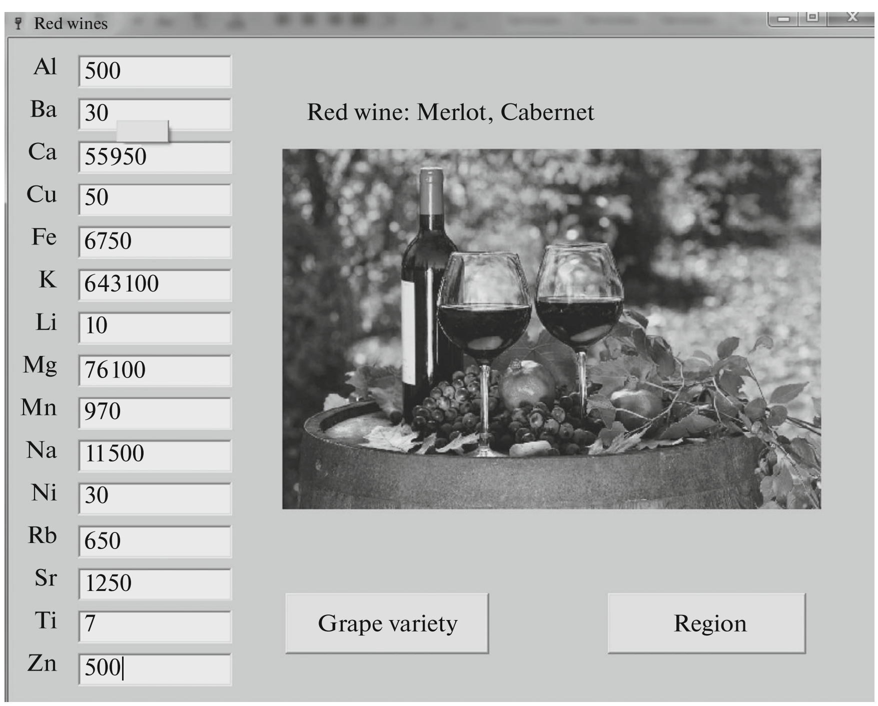

Fig. 5. Start window of the program for identifying the grape variety and region of growth.

ranges of elements in accordance with the training sample. Therefore, when you bring the cursor to the field with the name of the element, the recommended range corresponding to the smallest and largest values in the training sample is highlighted. If you click on the "Grape Variety" button, a window appears indicating the grape variety (Cabernet); if you click on the "Region" button, a window appears showing the region where the wine was grown (for example, the Chernomorskaya zone).

CONCLUSIONS

Thus, it is shown that, in spite of the complicated technological chain of soil–grapes–juice–wine, the elemental statistical image of the grape variety and the terrain in which it was grown remains in the wine. For Cabernet and Merlot wines produced in the territory of the main winemaking enterprises of the geographical areas of the Krasnodar krai, adequate mathematical models are proposed for the identification of the elements of the grape variety and the region where it grows. The models are accompanied by computational examples with the values of element concentrations in samples not belonging to the training sample. A software product is developed that automates the necessary calculations.

ACKNOWLEDGMENTS

The study was supported by the Ministry of Education and Science of the Russian Federation, project no. 4.2612.2017/PCh; experiments were carried out with the use of scientific equipment of the Ecological and Analytical Center of the Kuban State University, unique identifier RFMEFI59317X0008.

REFERENCES

1. Kaunova, A.A., Titarenko, V.O., Temerdashev, Z.A., Sekunova, M.V., and Popandopulo, V.G., *Zavod. Lab., Diagn. Mater.*, 2016, vol. 82, no. 8, p. 69.
2. Schlesier, K., Fauhl-Hassek, C., Forina, M., Cotea, V., Kocsi, E., Schoula, R., van Jaarsveld, F., and Wittkowski, R., *Eur. Food Res. Technol.*, 2009, vol. 230, no. 1, p. 1.
3. Daniel, C. and Smyth, H., *Compr. Anal. Chem.*, 2013, vol. 60, p. 385.
4. Yakuba, Yu.F., Kaunova, A.A., Temerdashev, Z.A., Titarenko, V.O., Khalafyan, A.A., *Analitika Kontrol'*, 2014, vol.18, no. 4, p. 344.
5. Khalafyan, A.A., Yakuba, Yu.F., Temerdashev, Z.A., Kaunova, A.A., and Titarenko, V.O., *J. Anal. Chem.*, 2016, vol. 71, no. 11, p. 1138.
6. Yakuba, Yu.F., Temerdashev, Z.A., and Khalaf'yan, A.A., *J. Anal. Chem.*, 2016, vol. 71, no. 2, p. 205.
7. Titarenko, V.O., Kaunova, A.A., Temerdashev, Z.A., and Popandopulo, V.G., *Analitika Kontrol'*, 2016, vol. 20, no. 2, p. 138.
8. Giaccio, M. and Vicentini, A., *J. Commod. Sci., Technol. Qual.*, 2008, vol. 47, p. 267.
9. Pohl, P., *TrAC, Trends Anal. Chem.*, 2007, vol. 26, p. 941.
10. Hopfer, H., Nelson, J., Collins, T.S., Heymann, H., and Ebeler, S.E., *Food Chem.*, 2015, vol. 172, p. 486.
11. Rapeanu, G., Vicol, C., and Bichescu, C., *Innovative Rom. Food Biotechnol.*, 2009, vol. 5, p. 1.
12. Kaunova, A.A., Petrov, V.I., Tsyupko, T.G., Temerdashev, Z.A., Perekotii, V.V., and Luk'yanov, A.A., *J. Anal. Chem.*, 2013 vol. 68, no. 9, p. 917.
13. Khalafyan, A.A., *STATISTICA 6. Matematicheskaja statistika s elementami teorii veroyatnostei* (STATISTICA 6: Mathematical Statistics with Elements of Probability Theory), Moscow: Binom, 2010.
14. Egorov, E.A., Guguchkina, T.I., Adzhiev, A.M., and Oseledtseva, I.V., *Geograficheskie zony proizvodstva vin i natsional'nykh kon'yakov (brendi) vysokogo kachestva na yuge Rossii* (Geographical Areas for the Production of Wines and National Cognacs (Brandy) of High Quality in Southern Russia), Krasnodar: Prosveshhenie-Yug, 2013.

Translated by O. Zhukova