

Оценка соответствия. Аккредитация лабораторий

УДК 663.222

ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ ПРИ КЛАССИФИКАЦИИ СОРТОВОЙ И РЕГИОНАЛЬНОЙ ПРИНАДЛЕЖНОСТИ БЕЛЫХ ВИН

© В. О. Титаренко, А. А. Халафян, З. А. Темердашев,
А. А. Каунова, Е. А. Ивановец¹

Статья поступила 28 марта 2017 г.

Белые вина четырех винодельческих регионов Краснодарского края были исследованы методом ИСП-АЭС. Определяли 19 элементов в 153 образцах белых вин марок Шардоне, Рислинг, Мускат. Методами статистического анализа показано, что несмотря на сложный технологический цикл производства вин (почва – виноград – сок – вино) в совокупности микроэлементов сохраняется информация о сорте винограда и регионе его произрастания. Построены вероятностно-статистические модели, которые позволяют по концентрациям определенного набора микроэлементов, содержащихся в вине, идентифицировать сорт винограда и регион, в котором он был выращен. Предложенные модели основаны на предположении, что происхождение почвы является важным фактором, определяющим микроэлементный состав вина.

Ключевые слова: идентификация региональной и сортовой принадлежности белых вин; содержание микроэлементов в винах; методы деревьев классификации и дискриминантного анализа.

В последние годы официальные контролирующие органы европейских стран по вину сталкиваются с образцами подозрительного качества (фальсификатами), которые сложно идентифицировать по сорту винограда и географическому происхождению [1]. Этот вопрос становится все более актуальным в контексте развития мирового рынка вина. В странах ЕЭС действуют нормативно-техническая и информационно-документальная базы, направленные на борьбу с некачественной и фальсифицированной продукцией [2]. Базовыми нормативными документами при установлении физико-химических показателей виноградных вин в России являются национальные стандарты ГОСТ Р 32030–2013 «Вина столовые и виноматериалы столовые. Общие технические условия» и ГОСТ Р 55242–2012 «Вина защищенных географических указаний и вина защищенных наименований места происхождения. Общие технические условия». Регламентируемые этими документами испытания направлены в основном на контроль безопасности продукции и позволяют установить ее соответствие своей товарной группе, но в полной мере не дают представления о ее подлинности.

Один из подходов к идентификации вин основан на получении большого массива данных по элементному составу напитков, винограда и почв, соответствующих области его произрастания, и установлении

взаимосвязи между ними [1, 3 – 7]. Анализ физико-химических показателей вин с использованием статистических методов обработки данных позволяет установить взаимосвязи между изучаемыми переменными, оценить вклад каждой из них в идентификационную мощность статистической модели и выявить критерии качества, подлинности и региональной принадлежности вин [8 – 12].

Важным аспектом оценки качества вин является выработка подходов по установлению достоверности их отнесения к напиткам защищенных географических указаний и защищенных наименований места происхождения по месту производства, произведенным из винограда одного сорта, обладающим определенной вариабельностью органолептических качеств, в частности, вкусовых характеристик.

В работе рассмотрена возможность идентификации методами классификационного анализа по концентрациям микроэлементов региона произрастания и сорта винограда на примере широко известных в России белых вин марок «Шардоне», «Рислинг» и «Мускат» кубанского происхождения.

Объекты исследования. Для исследования белых вин выбраны 153 образца сухих, полусладких и полусухих вин наименований «Рислинг», «Шардоне» и «Мускат», произведенных на территории Краснодарского края винодельческими предприятиями ЗАО «Запорожское», ООО «Кубань-Вино», ОАО АПФ «Фанагория», ООО АПК «Мильстрим-Черноморские

¹ Кубанский государственный университет, г. Краснодар, Россия; e-mail: temza@kubsu.ru

вина», ЗАО АФ «Кавказ», ЗАО «Абрау-Дюрсо», ЗАО АПК «Геленджик», ЗАО АФ «Мысхако», ООО «Фирма Сомелье», ООО АФ «Саук-Дере», ООО «Союз-Вино», расположенными в Южно-предгорной и Черноморской зонах, Анапской и Таманской подзонах Краснодарского края [13].

Материалы и методики исследований. Для приготовления растворов металлов использовали стандартные образцы: ГСО 7780–2000 (Li), ГСО 8062–94 (Na), ГСО 7767–2000 (Mg), ГСО 7854–2000 (Al), ГСО (K), ГСО 7772–2000 (Ca), ГСО 7205–95 (Ti), ГСО (Cr), ГСО 8056–94 (Mn), ГСО 8032–94 (Fe), ГСО 7784–2000 (Co), ГСО 7785–2000 (Ni), ГСО 7836–2000 (Cu), ГСО 8053–94 (Zn), ГСО 7035–93 (Rb), ГСО 7783–2000 (Sr), ГСО 7874–2000 (Cd), ГСО 7760–2000 (Ba), ГСО 7778–2000 (Pb). Все используемые в работе реактивы имели квалификацию не ниже хч. Содержание микро- и макроэлементов в винах определяли методом атомно-эмиссионной спектроскопии с индуктивно-связанной плазмой (ИСП АЭС) с использованием спектрометра iCAP-6000 (ThermoScientific). Перед анализом образцы вин разбавляли по методике, изложенной в работе [14].

Исследования проведены в среде пакета Statistica [15], использованы методы деревьев классификации и дискриминантного анализа. Модели классификации построены для каждого наименования вин.

Распределение частот 153 образцов вин по сортам указанных наименований и географическим зонам (подзонам) края отображено в табл. 1. Наибольшее количество образцов относится к Таманской подзоне (73), наименьшее — к Анапской (18). При этом вина «Шардоне», «Рислинг» и «Мускат» представлены 56, 49 и 48 образцами соответственно.

В табл. 2 приведены средние значения и стандартные отклонения концентраций определяемых элементов. Из таблицы следует, что, как правило, стандартные отклонения не превышают половины среднего значения, за исключением нескольких случаев: Cu («Шардоне» и «Рислинг» Таманской подзоны); Ni («Шардоне» Таманской подзоны); Ti («Мускат» Анапской подзоны); Zn («Шардоне» Таманской подзоны). Преимущественно малые относительно средних разбросы концентраций элементов говорят о том, что средние величины являются информативной характеристикой содержания элементов в вине. Так как для

большинства марок вин средние величины концентраций элементов существенно отличаются по региональной и сортовой принадлежности винограда, выявленная изменчивость является предпосылкой для разработки вероятностно-статистических моделей классификации вин как по сортовой, так и по зональной принадлежности винограда, из которого они произведены.

Малое количество образцов в некоторых группах — «Шардоне» Анапской подзоны (6) и Черноморской зоны (8), «Мускат» Южно-предгорной (5) и Черноморской (6) зон — не позволило успешно решить задачу классификации одновременно по всем 12 группам вин (см. табл. 1). Поэтому на первом этапе были построены модели для классификации по сортовой принадлежности винограда, а на втором — для каждого сорта винограда (марки вин) были разработаны модели для идентификации регионов его произрастания. Для построения моделей классификации использовали дискриминантный анализ и деревья классификации [15].

Этап 1. Цель дискриминантного анализа состояла в том, чтобы на основе определения различных характеристик (признаков) объекта классифицировать его, т.е. отнести к одной из нескольких заданных групп однородности оптимальным способом. При этом предполагается, что исходные данные наряду с признаками объектов содержат категориальную (группирующую) переменную, которая определяет принадлежность объекта к той или иной группе однородности. На первом этапе группирующей переменной является марка или наименование вина, которые принимают значения «Шардоне», «Рислинг», «Мускат».

Дискриминантный анализ относится к параметрическим методам, построенным на предположении о нормальном распределении количественных признаков объектов. Малое количество образцов вин в группах не позволило проверить справедливость гипотезы о соответствии эмпирического распределения признаков (в нашем случае — концентраций микроэлементов) нормальному распределению, поэтому дополнительно модели классификации были построены непараметрическим методом — деревьями классификации.

В табл. 3 представлены итоговые результаты анализа дискриминантных функций. Основным интегральным критерием качества процедуры дискриминации является статистика «лямбда Уилкса», которая равна отношению детерминанта матрицы межгрупповой дисперсии к детерминанту матрицы общей дисперсии [16] и принимает значения в диапазоне от 0 до 1: чем ближе значение к 0, тем адекватнее модель дискриминации описывает принадлежность объектов к группам. Близкое к 0 значение лямбды Уилкса (0,084) и ее статистическая значимость — уровень значимости p критерия Фишера $F(22,280)$ со степенями свободы 22 и 28 ($p < 0,05$) говорят о высокой адекватно-

Таблица 1. Частоты распределения образцов вин по зонам (подзонам)

Сорт винограда	Зоны произрастания винограда				Всего по стр.
	Таманская подзона	Анапская подзона	Южно-предгорная зона	Черноморская зона	
Шардоне	29	6	13	8	56
Рислинг	19	—	12	18	49
Мускат	25	12	5	6	48
Всего	73	18	30	32	153

Таблица 2. Средние значения и стандартные отклонения (ст. от.) концентраций элементов в образцах вин (мкг/дм³) по маркам вина и зонам

Элемент	Статистика	Шардоне				Рислинг				Мускат			
		Таманская	Анапская	Южно-предгорная	Черноморская	Таманская	Южно-предгорная	Черноморская	Таманская	Анапская	Южно-предгорная	Черноморская	Черноморская
Al	Среднее	714,90	503,62	1325,60	759,32	1588,73	1943,99	1617,35	928,90	957,98	1655,94	5266,38	5266,38
	Ст. откл.	269,61	36,24	213,46	338,87	639,36	166,25	899,25	305,39	526,19	362,56	3415,39	3415,39
Ba	Среднее	85,27	68,80	166,32	110,25	180,94	235,48	123,76	120,95	104,12	145,77	110,67	110,67
	Ст. откл.	40,13	21,42	64,13	55,85	37,85	128,07	32,18	54,08	10,33	36,56	37,13	37,13
Ca	Среднее	51522,84	35535,54	58738,75	57994,49	88414,31	88463,97	77092,22	76726,37	74692,73	74466,09	82599,68	82599,68
	Ст. откл.	18434,75	1664,59	8505,09	17902,10	7558,13	9313,62	9472,79	13557,82	11316,74	6634,65	4839,96	4839,96
Cu	Среднее	187,23	31,68	28,45	57,04	92,162	53,76	231,88	74,88	624,42	49,65	760,13	760,13
	Ст. откл.	252,33	2,63	6,23	5,86	1838,45	17,78	194,55	48,76	52,00	12,67	485,82	485,82
Fe	Среднее	2685,78	2574,12	5022,52	1273,99	2662,02	4546,35	3448,35	2611,95	7941,59	3904,19	5637,54	5637,54
	Ст. откл.	2045,95	297,50	434,77	1073,39	1849,62	837,45	2906,25	1571,58	616,58	526,34	397,12	397,12
K	Среднее	623660,04	736308,67	669531,46	675839,42	446769,62	806629,47	352501,0	541852,35	628852,32	825875,06	934571,39	934571,39
	Ст. откл.	123587,84	45568,91	182078,99	162437,77	186691,77	157734,55	124114,8	143045,12	43514,38	58025,44	297945,63	297945,63
Li	Среднее	37,33	9,58	10,27	14,86	24,86	12,97	20,40	30,67	16,19	20,35	20,98	20,98
	Ст. откл.	16,76	2,54	1,58	13,33	5,15	1,55	13,31	4,91	1,28	4,22	10,49	10,49
Mg	Среднее	84641,01	52971,52	56235,59	51283,28	91468,21	69925,06	58381,73	81506,07	53899,87	74581,80	81897,36	81897,36
	Ст. откл.	44497,78	1447,13	7362,12	4488,04	44242,46	8689,72	12672,97	23240,86	4320,70	10053,59	11839,01	11839,01
Mn	Среднее	930,07	787,86	715,10	1106,31	1361,35	539,22	1139,21	1155,77	682,81	949,69	1173,62	1173,62
	Ст. откл.	675,43	24,51	186,51	539,25	600,67	107,85	460,75	221,40	59,13	130,01	388,84	388,84
Na	Среднее	47779,73	23025,48	58471,24	31534,31	77526,81	38678,84	58775,51	69572,98	35728,72	47416,38	72013,67	72013,67
	Ст. откл.	15636,91	1504,89	6284,14	12785,05	30190,32	16505,50	39243,86	22528,33	2027,66	5028,91	3451,62	3451,62
Ni	Среднее	209,98	11,05	35,30	38,11	55,33	28,02	42,63	31,24	24,92	32,43	102,90	102,90
	Ст. откл.	696,46	3,66	6,82	12,51	19,61	9,00	36,63	8,63	9,02	7,23	65,42	65,42
Rb	Среднее	1795,16	615,65	614,51	1431,09	1004,90	584,87	559,88	621,23	250,98	635,21	521,62	521,62
	Ст. откл.	1754,31	31,02	185,93	574,79	334,52	67,54	86,46	136,69	37,78	101,17	255,90	255,90
Sr	Среднее	755,28	889,11	464,22	830,70	991,49	660,12	821,63	720,76	1373,18	812,97	934,50	934,50
	Ст. откл.	243,60	18,10	64,61	388,28	331,78	52,84	308,08	124,07	61,97	147,39	129,72	129,72
Ti	Среднее	16,47	13,14	20,41	22,36	23,37	31,65	24,18	12,25	14,42	16,98	23,85	23,85
	Ст. откл.	6,13	2,10	2,99	2,62	6,79	9,27	11,85	4,14	17,04	1,06	4,16	4,16
Zn	Среднее	614,19	357,25	521,51	662,74	1153,28	596,40	595,39	353,93	572,58	671,56	754,76	754,76
	Ст. откл.	1403,52	24,40	131,30	120,41	911,10	272,72	380,25	79,79	121,34	97,95	346,05	346,05

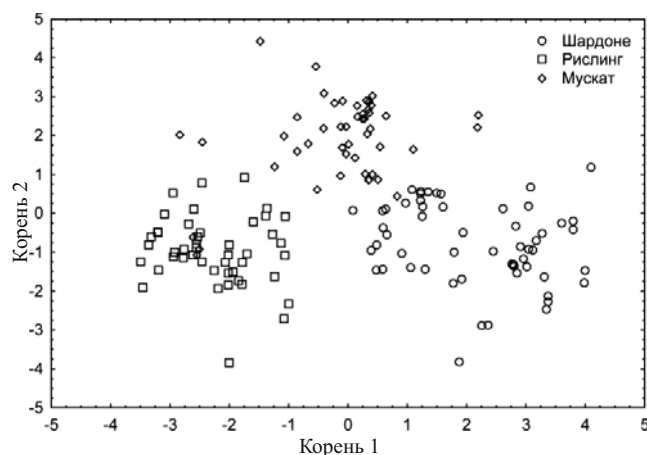


Рис. 1. Диаграмма рассеяния канонических корней образцов вин

сти модели классификации. По значениям показателей дискриминации можно судить о вкладе элементов в процедуру дискриминации вин.

В табл. 3 приведены значения *лямбды Уилкса*, являющиеся результатом исключения соответствующей переменной из модели, поэтому чем больше ее значение, тем более желательно присутствие этой переменной в процедуре дискриминации. *Частная лямбда* характеризует единичный вклад соответствующей переменной в разделительную силу модели, чем меньше эта статистика, тем больше вклад в общую дискриминацию. Элементы упорядочены по мере убывания их вклада в модель дискриминации. F — значение критерия Фишера, p — уровень значимости критерия Фи-

Таблица 3. Результаты дискриминантного анализа ($N = 153$)

Элемент	Лямбда Уилкса = 0,084, $F(22,280) = 30,999$, $p < 0$			
	Лямбда Уилкса	Частная лямбда	$F(2,140)$	p
Ca	0,182	0,464	80,818	0,000
K	0,162	0,520	64,505	0,000
Ti	0,162	0,522	64,018	0,000
Ba	0,126	0,668	34,690	0,000
Cu	0,120	0,700	29,910	0,000
Na	0,116	0,726	26,312	0,000
Rb	0,110	0,769	20,930	0,000
Al	0,094	0,895	8,194	0,000
Mn	0,094	0,901	7,684	0,000
Fe	0,091	0,925	5,639	0,004
Mg	0,091	0,929	5,333	0,005

Таблица 4. Матрица результатов классификации вин

Группа	Уровень верной классификации, %	Шардоне $P = 0,366$	Рислинг $P = 0,320$	Мускат $P = 0,313$
Шардоне	98,214	55	0	1
Рислинг	97,959	0	48	1
Мускат	91,666	1	3	44
Всего	96,078	56	51	46

шера. Чем больше F и меньше p , тем выше вклад переменной в процедуру дискриминации. Переменные, для которых уровень значимости больше чем 0,05, как правило, из модели исключают, поэтому были удалены микроэлементы Li, Ni, Sr, Zn. Последние два столбца таблицы характеризуют меру избыточности переменной в модели.

Об адекватности модели можно также судить по табл. 4, строки которой обозначают исходную принадлежность образцов вин из обучающей выборки к классам (маркам вин), а столбцы — принадлежность по построенной модели. Для каждой марки вин указаны априорные вероятности P , соответствующие доле данной марки вин в выборке. Из таблицы видно, что общее количество неверных классификаций равно 6, общая доля правильной классификации достаточно высока и составила 96,07 %.

Канонический анализ предусматривает нахождение коэффициентов дискриминантных функций (канонических корней), которые позволяют объекты n -мерного пространства (в нашем случае образцы вин — это объекты 11-мерного пространства) перенести в пространство размерности 2, сохранив порядок расстояний между ними. Если изобразить образцы вин в виде точек на плоскости в системе координат: ось OX — корень 1, ось OY — корень 2, то будет построена диаграмма рассеяния канонических значений (рис. 1). Видно, что образцы вин определенной марки образуют кластеры, которые локализованы в разных частях плоскости. При этом образцы вин «Шардоне» и «Рислинг» находятся на незначительном расстоянии от образцов вин «Мускат» и на достаточном удалении друг от друга, что и объясняет наличие ошибок при классификации обучающей выборки.

Функции классификации — это линейные функции, которые могут быть использованы для классифи-

Таблица 5. Коэффициенты функций классификации образцов вин

Элемент	Шардоне $P = 0,366$	Рислинг $P = 0,320$	Мускат $P = 0,313$
Al	−0,003061	−0,001078	−0,002077
Ba	−0,041826	−0,009548	−0,065435
Ca	0,000253	0,000549	0,000607
Cu	−0,003453	0,001064	−0,003397
Fe	0,000542	−0,000053	0,000574
K	0,000035	0,000008	0,000032
Mg	−0,000011	0,000049	0,000018
Mn	0,002485	−0,000890	−0,001259
Na	0,000120	−0,000001	0,000141
Rb	0,001234	−0,001984	−0,001610
Ti	−0,248036	−0,104560	−0,600829
Свободный член	−18,182553	−24,016053	−29,487708

Примечание. Значения классификационных функций чувствительны к точности значений переменных, поэтому при вычислении желательно использовать значения, приведенные с точностью до 6 знаков после запятой.

кации объектов (образцов вин). В табл. 5 приведены их угловые коэффициенты и свободные члены. Подставляя значения концентраций микроэлементов в функции классификации, можно вычислить их значения. Образец вина следует отнести к той марке вин, для которой функция имеет наибольшее значение.

Обозначим функции классификации для вин «Шардоне», «Рислинг», «Мускат» соответственно $F_{\text{шард}}$, $F_{\text{рисл}}$, $F_{\text{муск}}$, тогда:

$$F_{\text{шард}} = -18,183 - 0,003\text{Al} - 0,042\text{Ba} + 0,0003\text{Ca} - 0,003\text{Cu} + 0,0005\text{Fe} + 0,00004\text{K} - 0,00001\text{Mg} + 0,002\text{Mn} + 0,0001\text{Na} + 0,001\text{Rb} - 0,248\text{Ti}, \quad (1)$$

$$F_{\text{рисл}} = -24,016 - 0,001\text{Al} - 0,01\text{Ba} + 0,0005\text{Ca} + 0,001\text{Cu} - 0,0005\text{Fe} + 0,000008\text{K} + 0,00005\text{Mg} - 0,0009\text{Mn} - 0,000001\text{Na} - 0,002\text{Rb} - 0,105\text{Ti}, \quad (2)$$

$$F_{\text{муск}} = -24,488 - 0,002\text{Al} - 0,065\text{Ba} + 0,0006\text{Ca} - 0,003\text{Cu} + 0,0006\text{Fe} + 0,00003\text{K} + 0,00002\text{Mg} - 0,001\text{Mn} + 0,0001\text{Na} - 0,002\text{Rb} - 0,601\text{Ti}. \quad (3)$$

Пример 1. Определим наименование для образца со значениями концентраций элементов: Al = 730; Ba = 70; Ca = 54374; Cu = 3; Fe = 1016; K = 490088; Mg = 119066; Mn = 1108; Na = 51759; Rb = 4679; Ti = 17.

Произведя вычисления по формулам (1)–(3), получим: $F_{\text{шард}} = 52,73$; $F_{\text{рисл}} = 27,95$; $F_{\text{муск}} = 37,23$. Наибольшее значение имеет $F_{\text{шард}} = 52,73$, следовательно, образец вина принадлежит к вину марки «Шардоне».

Этап 2. Малое количество образцов вин определенного наименования в группах по регионам (см. табл. 1) снижает адекватность моделей классификации, поэтому для увеличения достоверности идентификации региона произрастания винограда для каждой из трех марок белых вин «Шардоне», «Рислинг», «Мускат» были построены модели классификации двумя методами — дискриминантного анализа и деревьев классификации.

Дискриминантный анализ. Значения лямбды Уилкса для вин марок «Шардоне», «Рислинг», «Мускат» составили соответственно 0,00013; 0,00322; 0,00001, их близость к нулю свидетельствует о высокой адекватности построенных классификационных моделей. Доля правильной классификации образцов вин всех марок по функциям классификации, разработанным программой, составила 100 %. Если обозначить для вин марок «Шардоне», «Рислинг», «Мускат» функции классификации по регионам (Таманская и Анапская подзоны; Южно-предгорная и Черноморская зоны) $F_{\text{анап}}^{\text{ш}}$; $F_{\text{южн}}^{\text{ш}}$; $F_{\text{черн}}^{\text{ш}}$; $F_{\text{там}}^{\text{ш}}$; $F_{\text{южн}}^{\text{р}}$; $F_{\text{черн}}^{\text{р}}$; $F_{\text{там}}^{\text{р}}$; $F_{\text{анап}}^{\text{м}}$; $F_{\text{южн}}^{\text{м}}$; $F_{\text{черн}}^{\text{м}}$; $F_{\text{там}}^{\text{м}}$, то функции классификации примут вид:

$$F_{\text{там}}^{\text{ш}} = -220,12 + 0,066\text{Al} + 0,003\text{Ca} + 0,08\text{Cu} + 0,015\text{Fe} - 0,00004\text{K} + 6,032\text{Li} + 0,002\text{Mg} + 0,04\text{Mn} +$$

$$+ 0,002\text{Na} - 0,037\text{Ni} - 0,002\text{Rb} - 0,49\text{Sr} + 3,89\text{Ti}, \quad (4)$$

$$F_{\text{анап}}^{\text{ш}} = -32,799 - 0,009\text{Al} - 0,0009\text{Ca} + 0,03\text{Cu} - 0,006\text{Fe} + 0,00007\text{K} - 0,9\text{Li} - 0,0003\text{Mg} - 0,004\text{Mn} + 0,0002\text{Na} + 0,003\text{Ni} + 0,003\text{Rb} + 0,075\text{Sr} + 0,98\text{Ti}, \quad (5)$$

$$F_{\text{южн}}^{\text{ш}} = -303,2 + 0,116\text{Al} + 0,004\text{Ca} + 0,13\text{Cu} + 0,034\text{Fe} - 0,00008\text{K} + 5,295\text{Li} + 0,002\text{Mg} + 0,027\text{Mn} + 0,001\text{Na} - 0,046\text{Ni} + 0,007\text{Rb} - 0,587\text{Sr} + 3,764\text{Ti}, \quad (6)$$

$$F_{\text{черн}}^{\text{ш}} = -75,928 + 0,032\text{Al} + 0,0009\text{Ca} + 0,07\text{Cu} - 0,0034\text{Fe} + 0,00003\text{K} + 1,862\text{Li} + 0,0003\text{Mg} + 0,006\text{Mn} + 0,0004\text{Na} + 0,006\text{Ni} + 0,005\text{Rb} - 0,142\text{Sr} + 4,398\text{Ti} \quad (7)$$

для вин марки «Шардоне»;

$$F_{\text{там}}^{\text{р}} = -319,81 + 0,006\text{Ca} + 0,055\text{Cu} + 0,0008\text{Fe} - 0,0002\text{K} + 0,004\text{Mg} - 0,264\text{Mn} + 0,0009\text{Na} + 0,138\text{Rb} - 0,022\text{Sr}, \quad (8)$$

$$F_{\text{южн}}^{\text{р}} = -271,07 + 0,004\text{Ca} + 0,047\text{Cu} + 0,011\text{Fe} - 0,0001\text{K} + 0,003\text{Mg} - 0,23\text{Mn} + 0,0003\text{Na} + 0,159\text{Rb} - 0,065\text{Sr}, \quad (9)$$

$$F_{\text{черн}}^{\text{р}} = -139,51 + 0,004\text{Ca} + 0,027\text{Cu} - 0,004\text{Fe} - 0,0002\text{K} + 0,002\text{Mg} - 0,153\text{Mn} + 0,0006\text{Na} + 0,048\text{Rb} + 0,021\text{Sr} \quad (10)$$

для вин марки «Рислинг»;

$$F_{\text{там}}^{\text{м}} = -369,166 + 0,003\text{Ca} - 0,075\text{Cu} - 0,06\text{Fe} + 0,0006\text{K} - 0,003\text{Mg} + 0,48\text{Mn} - 0,002\text{Na} + 0,014\text{Rb} - 0,458\text{Sr} + 15,616\text{Li} + 0,263\text{Ti} - 0,015\text{Zn} - 0,02\text{Al} + 0,585\text{Ba}, \quad (11)$$

$$F_{\text{анап}}^{\text{м}} = -1122,087 - 0,003\text{Ca} + 0,951\text{Cu} + 0,147\text{Fe} + 0,0007\text{K} - 0,003\text{Mg} - 0,967\text{Mn} + 0,003\text{Na} + 0,679\text{Rb} + 0,916\text{Sr} - 7,728\text{Li} + 7,09\text{Ti} + 0,106\text{Zn} - 0,191\text{Al} - 1,341\text{Ba}, \quad (12)$$

$$F_{\text{южн}}^{\text{м}} = -427,449 + 0,002\text{Ca} + 0,085\text{Cu} - 0,004\text{Fe} + 0,001\text{K} - 0,008\text{Mg} + 0,103\text{Mn} - 0,002\text{Na} + 0,513\text{Rb} - 0,293\text{Sr} + 12,639\text{Li} + 2,419\text{Ti} + 0,109\text{Zn} - 0,094\text{Al} - 0,556\text{Ba}, \quad (13)$$

$$F_{\text{черн}}^{\text{м}} = -299,179 + 0,0001\text{Ca} + 0,193\text{Cu} + 0,018\text{Fe} + 0,0006\text{K} - 0,003\text{Mg} - 0,025\text{Mn} + 0,001\text{Na} + 0,036\text{Rb} + 0,051\text{Sr} + 6,886\text{Li} + 1,433\text{Ti} - 0,004\text{Zn} - 0,053\text{Al} - 0,099 \quad (14)$$

для вин марки «Мускат».

Пример 2. Определим зону произрастания винограда для образца вина марки «Мускат», если

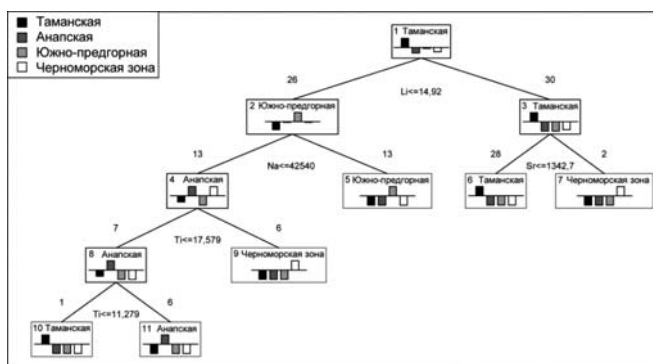


Рис. 2. Граф дерева классификации для вина марки «Шардоне» (число ветвлений — 5, число терминальных вершин — 6)

$Al = 600$, $Ba = 80$, $Ca = 66230$, $Cu = 55$, $Fe = 2460$, $K = 470000$, $Mg = 63350$, $Mn = 945$, $Na = 55900$, $Rb = 430$, $Sr = 540$, $Li = 28$, $Ti = 7$, $Zn = 260$. Воспользуемся формулами (11) – (14) и вычислим значения классификационных функций, подставив концентрации микроэлементов: $F_{там} = 314,36$; $F_{анап} = -1284,52$; $F_{южн} = 132,8$; $F_{черн} = 96,41$. Так как наибольшее значение принимает $F_{там} = 314,36$, то образец вина «Мускат» следует отнести к Таманской подзоне.

Деревья классификации представляют собой ветвление по двум возможным направлениям одной переменной, начиная с корневой вершины к вершинам-потомкам, пока вершины-потомки не окажутся терминальными — не способными к дальнейшему ветвлению [11]. Вершины означают группы однородности (классы), соответствующие значениям категориальной переменной региона произрастания: Анапская и Таманская подзоны; Южно-предгорная и Черноморская зоны.

Для вин марки «Шардоне» удалось методом полного перебора при выборе правила прямой остановки с долей неправильно классифицированных объектов, равной 0,05, построить небольшое дерево классификации, в котором достигнут компромисс между сложностью дерева и трудоемкостью процедуры классификации (рис. 2). Вершины дерева изображены в виде прямоугольников, в левом верхнем углу которых написан номер вершины, а над ними указано количество образцов вин, причисленных к классу, название которого написано внутри прямоугольника. Внутри каждой вершины приведены столбчатые диаграммы, изображающие классы вин. Диаграмма для каждого класса обозначена определенным цветом, причем диаграмма для указанного в вершине класса направлена вверх, а остальные диаграммы — вниз. Установлено, что количество ошибок классификации обучающей выборки равно 0, т.е. данная модель столь же адекватна, как и построенная дискриминантным анализом.

По дереву легко составить алгоритм, который позволит для вин марки «Шардоне» установить регион произрастания винограда.

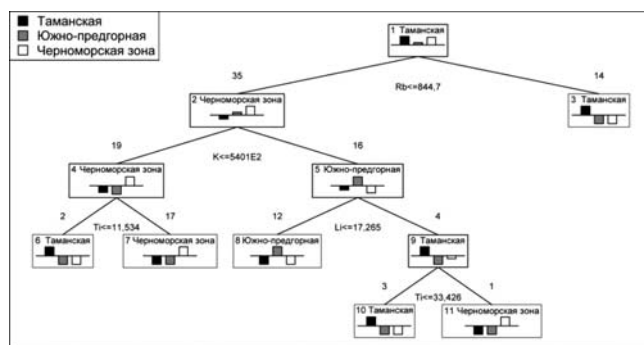


Рис. 3. Граф дерева классификации для вина марки «Рислинг» (число ветвлений — 5, число терминальных вершин — 6)

Шаг 1. Если $Li \leq 14,92$, переходим к шагу 2, в противном случае — к шагу 5.

Шаг 2. Если $Na \leq 42540$, переходим к шагу 3, в противном случае — образец из Южно-предгорной зоны, переходим к завершению работы алгоритма.

Шаг 3. Если $Ti \leq 17,579$, переходим к шагу 4, в противном случае образец из Черноморской зоны, переходим к завершению работы алгоритма.

Шаг 4. Если $Ti \leq 11,279$, образец из Таманской подзоны; в противном случае — из Анапской подзоны, в обоих случаях переходим к завершению работы алгоритма.

Шаг 5. Если $Sr \leq 1342,7$, образец из Таманской подзоны, в противном случае — из Черноморской зоны, в обоих случаях переходим к завершению работы алгоритма.

Для вин марки «Рислинг» при тех же установках параметров ветвления построено адекватное дерево (рис. 3), алгоритм классификации по которому состоит из 5 шагов.

Шаг 1. Если $Rb \leq 844,7$, переходим к шагу 2, в противном случае — образец из Таманской подзоны, переходим к завершению работы алгоритма.

Шаг 2. Если $K \leq 540098,2$, переходим к шагу 3, в противном случае — к шагу 4.

Шаг 3. Если $Ti \leq 11,534$, образец из Таманской подзоны, в противном случае — из Черноморской зоны, в обоих случаях переходим к завершению работы алгоритма.

Шаг 4. Если $Li \leq 17,285$, образец из Южно-предгорной зоны, завершение работы алгоритма; в противном случае переходим к шагу 5.

Шаг 5. Если $Ti \leq 33,426$, образец из Таманской подзоны, в противном случае — из Черноморской зоны, в обоих случаях переходим к завершению работы алгоритма.

Для марки вин «Мускат» при тех же установках параметров ветвления также построено адекватное дерево (рис. 4), алгоритм классификации по которому состоит из 4 шагов.

Шаг 1. Если $Cu \leq 2401,23$, переходим к шагу 2, в противном случае — к шагу 4.

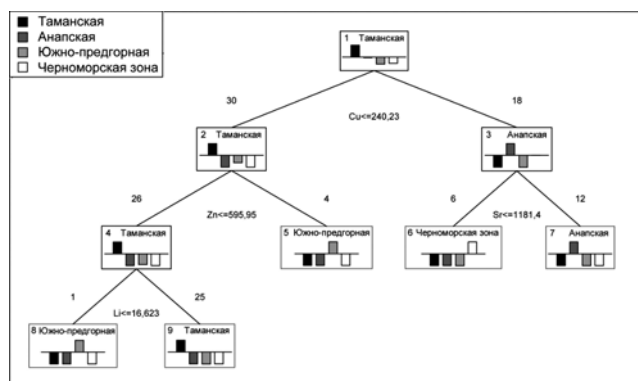


Рис. 4. Граф дерева классификации для вина марки «Мускат» (число ветвлений — 4, число терминальных вершин — 5)

Шаг 2. Если $Zn \leq 595,95$, переходим к шагу 3, в противном случае образец из Южно-предгорной зоны.

Шаг 3. Если $Li \leq 16,623$, образец из Южно-предгорной зоны, в противном случае — из Таманской подзоны, в обоих случаях переходим к завершению работы алгоритма.

Шаг 4. Если $Sr \leq 1181,4$, образец из Черноморской зоны, в противном случае — из Анапской подзоны, в обоих случаях переходим к завершению работы алгоритма.

Пример 3. Определим зону произрастания винограда для образца вина марки «Мускат» из примера 2, т.е. $Cu = 55$, $Sr = 540$, $Li = 28$, $Zn = 260$.

Шаг 1. Так как $Cu \leq 2401,23$, переходим к шагу 2.

Шаг 2. Так как $Zn \leq 595,95$, переходим к шагу 3.

Шаг 3. Так как $Li > 16,623$, то виноград выращен в Таманской подзоне.

Для автоматизации вычислений по предложенным классификационным функциям дискриминантного анализа и алгоритмам деревьев классификации разработан программный продукт с простым и удобным интерфейсом — от пользователя не требуется знаний классификационного анализа (рис. 5). В поля с названиями микроэлементов следует ввести их концентрации для анализируемого образца вина. При подведении курсора к полю с названием микроэлемента высвечивается рекомендуемый диапазон значений (от наименьшего до наибольшего в обучающей выборке). Введем, к примеру, значения концентраций для образца вина марки «Мускат» из примера 2, для которого известно, что виноград выращен в Таманской подзоне.

Если последовательно щелкнуть по кнопкам «Сорт винограда» и «Регион», то появятся окна с указанием сорта винограда — «Мускат» и региона произрастания винограда — Таманская подзона.

Таким образом, в работе показано, что несмотря на сложный технологический процесс в цепочке «почва — виноград — сок — вино» в вине сохраняется информация в виде элементного статистического «образа» сорта винограда и местности, где он был выращен. Для вин марок «Рислинг», «Шардоне» и

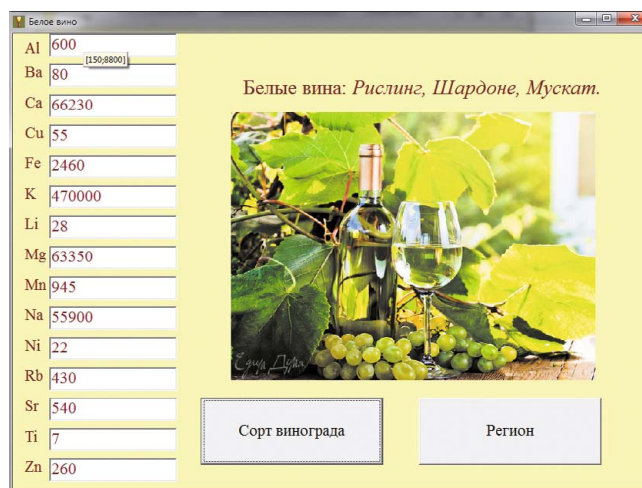


Рис. 5. Интерфейсное окно программы

«Мускат», произведенных на территории Краснодарского края, построены вероятностно-статистические модели для идентификации сорта винограда и региона его произрастания по концентрациям элементов. Адекватность построенных моделей продемонстрирована на примерах для образцов, не принадлежащих обучающей выборке. Разработан программный продукт для автоматизации необходимых вычислений.

ЛИТЕРАТУРА

- Schlesier K., Fauhl-Hassek C., Forina M., et al. Characterisation and determination of the geographical origin of wines. Part I: Overview / European Food Res. Technol. 2009. Vol. 230. N 1. P. 1 – 13.
- Вина и алкогольные напитки. Директивы и регламенты Европейского Союза. — М.: ИПК Издательство стандартов, 2000. — 616 с.
- Титаренко В. О., Каунова А. А., Темердашев З. А., Попандупло В. Г. Исследование взаимосвязи между элементным составом винограда и почвой региона его произрастания / Аналитика и контроль. 2016. Т. 20. № 2. С. 138 – 146.
- Daniel C., Smyth H. Analytical and chemometric-based methods to monitor and evaluate wine protected designation / Comp. Anal. Chem. 2013. Vol. 60. P. 385 – 408.
- Якуба Ю. Ф., Каунова А. А., Темердашев З. А. и др. Виноградные вина, проблемы оценки их качества и региональной принадлежности / Аналитика и контроль. 2014. Т. 18. № 4. С. 344 – 373.
- Khalafyan A. A., Yakuba Yu. F., Temerdashev Z. A., et al. Statistical-probability simulation of the organoleptic properties of grape wines / J. Anal. Chem. 2016. Vol. 71. N 11. P. 1138 – 1144.
- Pohl P. What do metals tell us about wine? / Trends Anal. Chem. 2007. Vol. 26. P. 941 – 949.
- Yakuba Yu. F., Temerdashev Z. A., Khalaf'yan A. A. Application of ranging analysis to the quality assessment of wines on a nominal scale / J. Anal. Chem. 2016. Vol. 71. N 2. P. 205 – 214.
- Giaccio M., Vicentini A. Determination of the geographical origin of wines by means of the mineral content and the stable isotope ratios: A review / J. Commodity Sci., Technol. Quality. 2008. Vol. 47. P. 267 – 284.
- Hopfer H., Nelson J., Collins T. S., et al. The combined impact of vineyard origin and processing winery on the elemental profile of red wines / Food Chem. 2015. Vol. 172. P. 486 – 496.
- Халафян А. А., Якуба Ю. Ф., Темердашев З. А. Применение методов таблицы сопряженности и анализа соответствий для сравнительной оценки качества вин в номинальной шкале / Заводская лаборатория. Диагностика материалов. 2016. Т. 82. № 2. С. 66 – 71.

12. Халафян А. А., Темердашев З. А., Якуба Ю. Ф., Гутучкина Т. И. Использование многомерного анализа для итоговой оценки результатов экспертных оценок / Заводская лаборатория. Диагностика материалов. 2016. Т. 82. № 10. С. 71 – 78.
13. Егоров Е. А., Гутучкина Т. И., Аджиев А. М., Оселедцева И. В. Географические зоны производства вин и национальных коньяков (бренди) высокого качества на юге России. — Краснодар: ГНУ СКЗНИИСиВ; Просвещение-Юг, 2013. — 155 с.
14. Kaunova A. A., Petrov V. I., Tsiupko T. G., et al. Identification of wine provenance by ICP-AES multielement analysis / J. Anal. Chem. 2013. Vol. 68. N 9. P. 917 – 922.
15. Халафян А. А. Statistica 6. Математическая статистика с элементами теории вероятностей. — М.: Бином, 2010. — 491 с.
16. StatSoft, Inc. (2012). Электронный учебник по статистике. URL: <http://www.statsoft.ru/home/textbook/default.htm> (дата обращения 28.03.17).

Compliance verification. Laboratory accreditation

UDC 663.222

THE USE OF STATISTICAL METHODS IN THE CLASSIFICATION OF VARIETAL AND REGIONAL ORIGIN OF WHITE WINES

© V. O. Titarenko, A. A. Khalafyan, Z. A. Temerdashev,
A. A. Kaunova, and E. A. Ivanovets

Submitted March 28, 2017.

White wines of four wine-growing regions of the Krasnodar Territory were studied using the ISP-AES method. We have analyzed concentrations of 19 elements in 153 samples of white wines of Chardonnay, Riesling, and Muscat. Statistical analysis of the result demonstrated that despite complex technological cycle of wine production (soil – grapes – juice – wine) the microelement composition provides information about the grape variety and region of origin. Probabilistic statistical models are constructed to identify the grape variety and the region of growing proceeding from the concentration of a certain set of trace elements contained in the wine. The proposed models are based on the assumption that the soil origin is an important factor affecting the microelement composition of the wine.

Keywords: identification of regional and varietal origin of white wines; micronutrient content of wines; methods of classification trees and discriminant analysis.

REFERENCES

1. Schlesier K., Fauth-Hasek C., Forina M., et al. Characterisation and determination of the geographical origin of wines. Part I: Overview / European Food Res. Technol. 2009. Vol. 230. N 1. P. 1 – 13.
2. Wines and alcoholic drinks. EU directives and regulations. — Moscow: IPK Izdatel'stvo standartov, 2000. — 616 p. [in Russian].
3. Titarenko V. O., Kaunova A. A., Temerdashev Z. A., Popandopulo V. G. Investigation of the correlation between the elemental content of grapes and the soil of the region of its growth / Analit. Kontrol'. 2016. Vol. 20. N 2. P. 138 – 146 [in Russian].
4. Daniel C., Smyth H. Analytical and chemometric-based methods to monitor and evaluate wine protected designation / Comp. Anal. Chem. 2013. Vol. 60. P. 385 – 408.
5. Yakuba Yu. F., Kaunova A. A., Temerdashev Z. A., et al. Grape wines, problems of their quality and regional origin evaluation / Analit. Kontrol'. 2014. Vol. 18. N 4. P. 344 – 373 [in Russian].
6. Khalafyan A. A., Yakuba Yu. F., Temerdashev Z. A., et al. Statistical-probability simulation of the organoleptic properties of grape wines / J. Anal. Chem. 2016. Vol. 71. N 11. P. 1138 – 1144.
7. Pohl P. What do metals tell us about wine? / Trends Anal. Chem. 2007. Vol. 26. P. 941 – 949.
8. Yakuba Yu. F., Temerdashev Z. A., Khalafyan A. A. Application of ranging analysis to the quality assessment of wines on a nominal scale / J. Anal. Chem. 2016. Vol. 71. N 2. P. 205 – 214.
9. Giaccio M., Vicentini A. Determination of the geographical origin of wines by means of the mineral content and the stable isotope ratios: A review / J. Commodity Sci., Technol. Quality. 2008. Vol. 47. P. 267 – 284.
10. Hopfer H., Nelson J., Collins T. S., et al. The combined impact of vineyard origin and processing winery on the elemental profile of red wines / Food Chem. 2015. Vol. 172. P. 486 – 496.
11. Khalafyan A. A., Yakuba Yu. F., Temerdashev Z. A. Application of table of congruences and correspondence analysis to comparative assessment of wine quality in rating scale / Zavod. Lab. Diagn. Mater. 2016. Vol. 82. N 2. P. 66 – 71.
12. Khalafyan A. A., Temerdashev Z. A., Yakuba Yu. F., Gutuchkina T. I. The use of multivariate analysis for the final evaluation of the results of expert assessments / Zavod. Lab. Diagn. Mater. 2016. Vol. 82. N 10. P. 71 – 78.
13. Egorov E. A., Guguchkina T. I., Adzhiev A. M., Oseledtseva I. V. Geographical areas and national wine production of high quality cognac (brandy) in southern Russia. — Krasnodar: GNU SKZNIISiV; Prosveshchenie-Yug, 2013. — 155 p. [in Russian].
14. Kaunova A. A., Petrov V. I., Tsiupko T. G., et al. Identification of wine provenance by ICP-AES multielement analysis / J. Anal. Chem. 2013. Vol. 68. N 9. P. 917 – 922.
15. Khalafyan A. A. Statistica 6. Mathematical statistics with elements of theory of probability. — Moscow: Binom, 2010. — 491 p. [in Russian].
16. StatSoft, Inc. (2012). Electronic textbook on statistics. URL: <http://www.statsoft.ru/home/textbook/default.htm> (accessed 28.03.17) [in Russian].