

УДК 663.2:663.256.1

ПРИМЕНЕНИЕ МНОГОМЕРНОГО СПЕКТРОМЕТРИЧЕСКОГО АНАЛИЗА ДЛЯ ИДЕНТИФИКАЦИИ МОЛДАВСКИХ ДИВИНОВ

Ходасевич М.А.¹, канд. физ.-мат. наук,

Синицын Г.В.³, канд. физ.-мат. наук, Роговая М.В.⁵

Институт физики им. Б. И. Степанова НАН Беларуси (Минск, Беларусь)

Скорбанова Е.А.², канд. техн. наук, Камбур Е.И.⁴, Дегтярь Н.Ф.⁶, канд. техн. наук

Научно-практический институт плодоводства, виноградарства
и пищевых технологий (Кишинев, Молдова)

Гаина Б.С.⁷, д-р техн. наук, академик

Академия наук Молдовы (Кишинев, Молдова)

Реферат. Показана возможность установления подлинности молдавских дивinov (коньяков) путем использования спектроскопии УФ-, видимого и ближнего ИК-диапазонов с последующей хемометрической обработкой спектров пропускания. Применение метода главных компонент позволяет определить производителя. Использование метода позволяет упростить дорогостоящую процедуру определения химических показателей дивinov и других аналогичных продуктов.

Ключевые слова: дивин, спектроскопический анализ, метод главных компонент, контроль качества

Summary. The possibility of using of multivariate data analysis of UV-Vis-NIR transmission of spectra is shown for solving the authenticity problem of Moldavian divins (cognacs). Application of principal component analysis allows you to define the manufacturer. The use of method allows you to simplify the expensive procedure for determining the chemical parameters of cognacs and other similar products.

Key words: divin, spectroscopy analysis, principal component, quality control

Введение. В последнее время в странах с развитым виноделием уделяется большое внимание совершенствованию методов контроля качества продукции из винограда. В определенной степени это коснулось винных дистиллятов и приготовленных на их основе бренди, коньяков и других напитков. В этой связи объективная оценка возраста выдержки винных дистиллятов [1], идентификация производителя [2], географическое происхождение напитка [3] являются актуальной проблемой.

В Молдове дивин (коньяк) является традиционным продуктом виноделия. Нами ранее [4, 5] уже была сделана попытка классифицировать молдавские винные дистилляты по возрасту выдержки и производителю методами проекции латентных структур (PLS) [6] и главных компонент (PCA) [7] спектров пропускания дистиллятов. Дивины (коньяки), по сравнению с дистиллятами, характеризуются более сложным и неоднородным химическим составом, так как производятся путем купажирования винных дистиллятов разного возраста, в этой связи классифицировать их по возрасту достаточно сложно.

Целью наших исследований явилось установление возможности классификации и идентификации дивinov (коньяков) с помощью многопараметрического анализа спектров пропускания исследуемых объектов.

Объекты и методы исследований. Для исследований были взяты 24 дивина (коньяка), произведенных в Молдове на двух различных предприятиях. Был создан массив данных, характеризующих их физико-химический состав. Летучие компоненты определяли методом газо-жидкостной хроматографии на хроматографе GC System Agilent Technologies

7890В с пламенно-ионизационным детектором, продукты распада лигнина (ароматические альдегиды и ароматические кислоты) – на жидкостном хроматографе Shimadzu LC-20А. Методы сбора спектральной информации об образцах были описаны нами ранее [8].

После всех процедур каждый исследуемый образец характеризовался номером производителя, пятью химическими параметрами (концентрацией этилацетата, галловой кислоты, сиреневого альдегида, ванилина и ванилиновой кислоты) и спектрами пропускания, содержащими 2698 спектральных отсчета или переменных.

Обсуждение результатов. К матрице зарегистрированных спектров пропускания с предварительным их центрированием применили метод главных компонент PCA (приведение средней величины отсчетов на каждой длине волны к нулевому значению). Метод главных компонент (англ. principal component (PC) analysis, PCA) разлагает 2698-мерное спектральное пространство в двумерное пространство главных компонент с объясненной дисперсией спектров, равной 97,0%.

На рис. 1 изображен график счетов в пространстве первой и второй главных компонент, из которого видно, что 6 образцов дивинов являются возможными претендентами на выбросы.

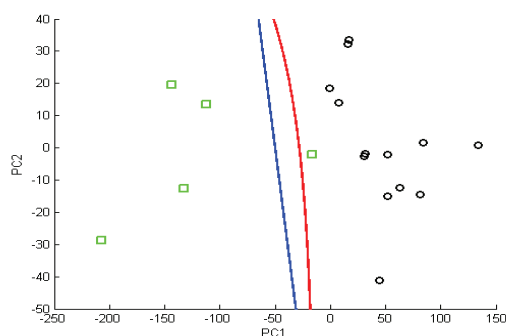


Рис. 1. График счетов в пространстве первой и второй главных компонент

Анализ данных физико-химического состава этих образцов (табл. 1) показал, что один из них значительно отличался высокой концентрацией сиреневого альдегида, второй имел существенно высокое содержание ванилиновой кислоты, три из оставшихся характеризовались небольшим количеством ванилина и один имел малую концентрацию галловой кислоты. Все шесть образцов были удалены из выборки.

Матрица спектров пропускания 18 оставшихся образцов была проанализирована вновь с помощью метода главных компонент (PCA). График счетов, представленных на рис. 1, показывает четкую кластеризацию дивинов двух производителей. Однако при этом, ни линейный (синяя прямая), ни квадратичный (красная парабола) дискриминантный анализ не кластеризует рассматриваемых производителей полностью [9]. Поэтому мы применили другой метод – метод построения классификационного дерева [10] в пространстве главных компонент.

В рассматриваемом случае входными переменными являются счета двух первых главных компонент, определенные из спектров пропускания дистиллятов, а целевой переменной – номер производителя. Результат построения однорангового предиктора показан на рис. 2.

Таким образом, в нашем случае достаточно одной главной компоненты и одноуровневого классификационного дерева для классификации выборки из 18 образцов от двух производителей

Таблица 1 – Физико-химический состав образцов дивинов

Номер произво- дителя	Этилацетат, мг/дм ³	Галловая кислота, мг/дм ³	Сиреневый альдегид, мг/дм ³	Ванилиновая кислота, мг/дм ³	Ванилин, мг/дм ³
1	22,5	0,6	2,3	0,7	0,5
1	22,8	0,5	1,3	0,4	0,2
1	15,6	1,3	2,8	23,5	1,6
1	19,0	0,3	2,3	0,3	1,6
1	20,1	4,2	2,9	0,1	1,1
1	24,9	4,3	2,5	0,4	0,9
1	28,5	2,4	3,4	0,9	0,8
1	30,8	1,6	4,1	0,3	0,7
1	23,2	2,0	1,9	0,3	3,3
1	26,3	3,9	3,4	0,3	0,6
1	24,9	4,5	4,0	3,3	0,8
1	28,2	5,4	3,6	0,9	0,4
1	28,5	1,1	3,5	0,3	0,7
1	27,6	3,6	3,8	0,6	1,0
1	30,7	4,4	4,8	0,7	1,0
1	24,2	3,3	8,6	0,1	1,2
1	16,8	2,6	1,9	<0,1	4,0
2	21,5	1,6	5,8	1,6	1,2
2	21,5	1,8	5,1	1,3	1,7
2	21,9	4,2	4,5	1,3	1,3
2	24,8	2,0	3,2	0,4	1,0
2	25,8	4,8	4,5	1,3	1,0
2	29,4	1,4	3,3	0,4	0,2
2	18,5	3,3	0,3	0,2	0,1

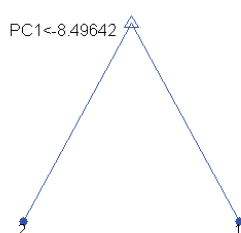


Рис. 2. Идентификация производителей коньяков методом построения классификационного дерева в пространстве главных компонент

Для нахождения корреляционных зависимостей химических характеристик дивинов с их спектральными данными с помощью регрессии мы использовали метод проекции на латентные структуры (PLS) [6]. В отличие от линейного метода главных компонент, метод PLS является билинейным статистическим методом, который одновременно проецирует предикторы (спектры) и отклик (например, концентрация этилацетата) в новое маломерное пространство латентных структур. Для правильного применения многопараметрического анализа данных мы разделили исследуемые образцы дивинов на обучающий и тестовый

наборы. Обучающий набор был использован для моделирования и перекрестной проверки методом одного из оставшихся. Тестовый набор был использован для тестовой проверки и оценки погрешности моделирования. Были рассмотрены следующие химические параметры дивинов: концентрации этилацетата, галловой кислоты, сиреневого альдегида, ванилиновой кислоты и ванилина. Возраст дивинов в выборке был ограничен всего двумя значениями и поэтому нерепрезентативен.

Число латентных структур было выбрано в зависимости от коэффициентов корреляции для обоих наборов. В нашем случае число латентных структур составило: 8 для сиреневого альдегида (14 образцов в обучающем наборе и 3 образца в тестовом наборе), 5 для этилацетата (14 образцов в обучающем наборе и 3 образца в тестовом наборе), 5 для ванилина (12 образцов в обучающем наборе и 2 образца в тестовом наборе) и 4 для галловой кислоты (11 образцов в обучающем наборе и 2 образца в тестовом наборе). Наличие образцов с одинаковыми количественными показателями существенно ухудшает качество моделирования многопараметрическими методами обработки информации.

Этот вывод подтверждается исследованием концентрации ванилиновой кислоты. Из рассмотрения были изъяты образцы с неколичественным параметром ($< 0,1$) и со значением резко отличающимся от остальных (3,3). Анализ данных в табл. 1 показал, что среди 16 оставшихся образцов есть всего 8 различающихся значений содержания ванилиновой кислоты. Следовательно, выборка нерепрезентативна и ее моделирование не проводилось.

Полученные результаты могут быть также представлены как соизмеренные величины химического параметра с его прогнозируемым количеством.

На рис. 4 показана эта зависимость для сиреневого альдегида. В идеале, все точки должны лежать на диагонали. Отклонение от диагонали показывает относительную погрешность определения концентрации.

Лучший набор образцов, рассмотренных здесь, дает относительную погрешность для сиреневого альдегида не более чем на 3%. Худшие условия для регрессии оценки других химических параметров дают относительную погрешность 6% для этилацетата, 16% для ванилина и 20% для галловой кислоты.

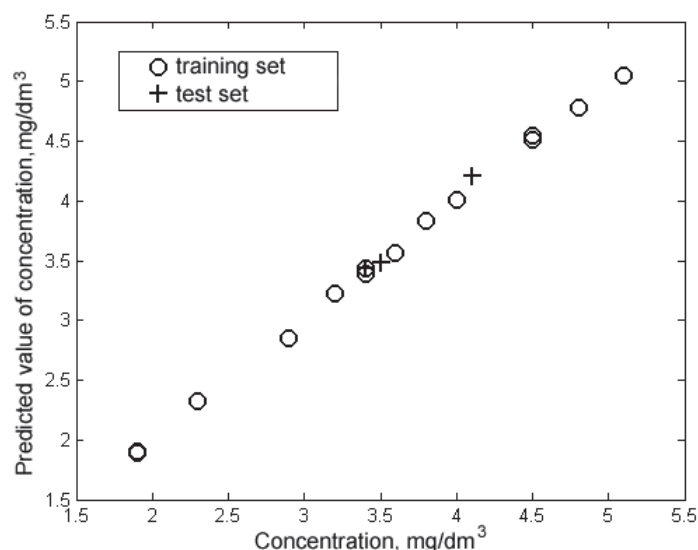


Рис. 4. Соотношение известной концентрации сиреневого альдегида в обучающем и тестовом наборах коньяков и значения прогнозируемой концентрации методом проекции спектров пропускания на 8 латентных структур

Выводы. Показана возможность использования многопараметрического анализа данных (УФ, видимого и ближнего ИК спектров пропускания) для решения проблемы классификации молдавских дивинов (коньяков). Применение метода главных компонент и классификационных деревьев позволяет определить производителя. Проекция спектров коньяков на латентные структуры дает возможность оценить четыре химических параметра с достаточной точностью.

Полученные результаты позволяют упростить процедуру определения химических показателей дивинов (коньяков). Малый размер, высокая эффективность, отсутствие регулировок точности и легкость в использовании являются важными преимуществами для практического применения спектроскопических систем, основанных на применении многопараметрического анализа данных. Метод может быть применен для соответствующего контроля в специализированных лабораториях.

Часть представленных результатов получена в рамках совместного проекта № 13.820.14.07/BA Молдавского Фонда двусторонней программы сотрудничества между Академиями наук Беларуси и Молдовы и проектом № F13MLD-011 Белорусского республиканского фонда фундаментальных исследований.

Литература

1. Contreras U., Barbosa-García O., Pichardo-Molina J.L., Ramos-Ortíz G., Maldonado J.L., Meneses-Nava M.A., Ornelas-Soto N.E., López-de-Alba P.L. // Food Research International. 2010. V. 43. P. 2356.
2. Borba F., Honorato R., Juan A. // Forensic Science International. 2015. V. 249. P. 73.
3. Liu L., Cozzolino D., Cynkar W.U., Damberg R.G., Janik L., O'Neill B.K., Colby C.B., Gishen M. // Food Chemistry. 2008. V. 106. P. 781.
4. Degtyar N.F., Nezalova E.I., Rogovaya M.V., Sinitsyn G.V., Skorbanova E.A., Khodasevich M.A. // Vesti NAN Belarusi. 2014. № 3. P. 113 (in Russian).
5. Rogovaya M. V., Sinitsyn G. V., Khodasevich M. A. // Optics and Spectroscopy. 2014, v. 117. P. 839.
6. Abdi H., Williams L. J. // Wiley Interdisciplinary Reviews: Computational Statistics. 2010, v. 2. P. 433–459.
7. Abdi H. // Wiley Interdisciplinary Reviews: Computational Statistics. 2010, v. 2. P. 97–106.
8. Khodasevich M.A., Sinitsyn G.V., Skorbanov E.A., Rogovaya M.V., Cambur E.I., Aseev V.A. // Optics and Spectroscopy. 2015, in press.
9. Abdi H. // Discriminant correspondence analysis. In: N.J. Salkind (Ed.): Encyclopedia of Measurement and Statistic. Thousand Oaks (CA): Sage. 2007. P. 270.
10. Brown S. D., Myles A. J. // Comprehensive Chemometrics: Chemical and Biochemical Data Analysis. 2009, v. 3. P. 541–569