

ИДЕНТИФИКАЦИЯ ЛЬНЯНОГО МАСЛА ЛИНЕЙНЫМИ МНОГОПАРАМЕТРИЧЕСКИМИ МЕТОДАМИ СПЕКТРАЛЬНОГО АНАЛИЗА

М. А. Ходасевич*, Д. А. Борисевич

УДК 535.372

Институт физики НАН Беларуси,

220072, Минск, просп. Независимости, 68-2, Беларусь; e-mail: m.khodasevich@ifanbel.bas-net.by

(Поступила 6 августа 2019)

Разработана многопараметрическая модель для калибровки содержания льняного масла в бинарной смеси с нерафинированным подсолнечным маслом по спектрам оптической плотности в УФ, видимом и ближнем ИК диапазонах на основе метода главных компонент, кластерного анализа и линейного метода интервальной проекции на латентные структуры. Разработанная модель характеризуется меньшей среднеквадратичной ошибкой (0.12 %) по сравнению с классической содержательной моделью (0.42 %) и многопараметрической моделью (0.63 %), использующей весь измеренный спектральный диапазон.

Ключевые слова: спектральный анализ, метод главных компонент, проекция на латентные структуры, кластерный анализ.

A multivariate model has been developed for calibrating the content of flax oil in a binary mixture with unrefined sunflower oil using the optical density spectra in the UV, visible and near IR ranges. The proposed model, based on the principal component analysis, cluster analysis and the linear method of interval projection to latent structures, is characterized by a smaller root-mean-square error (0.12%) compared to the classical interpretable model (0.42%) and the multivariate model (0.63%) applying the full measured spectral range.

Keywords: spectral analysis, principal component analysis, projection to latent structures, cluster analysis.

Введение. В последнее время во всем мире серьезной проблемой стала фальсификация продуктов питания, которая кроме финансовых потерь производителей оригинальной продукции может привести к непредсказуемому ухудшению состояния здоровья потребителей. При этом почти 68 % поддельной пищевой продукции приходится на продукты животного и растительного происхождения с высоким содержанием жира (27 % мясо, 13 % рыба, 11 % масла, 10 % молочные продукты, 4 % орехи и семена, 3 % других продуктов животного происхождения) [1]. Среди большого разнообразия пищевых растительных масел наибольшим потенциалом для подделки характеризуется оливковое. Однако на первое место по полезности для здоровья человека диетологи ставят льняное масло, которое является самым богатым источником незаменимой α -линоленовой кислоты (омега-3). На ее долю приходится до 64 % суммарного содержания жирных кислот в льняном масле [2]. Уникальные свойства льняного масла и экономические причины обеспечивают ему растущий спрос во многих странах мира, для удовлетворения которого используются значительные посевные площади масличного льна (до 3 млн га), сосредоточенные в Канаде, Аргентине, Китае, Индии, США, России и др. В Республике Беларусь на сегодняшний день масличный лен занимает небольшие посевные площади (1.5—2.5 тыс. га). Востребованность продуктов переработки льна в странах Европейского Союза и растущий спрос на льняное масло являются стимулом к увеличению его производства [3].

IDENTIFICATION OF FLAX OIL BY THE LINEAR MULTIVARIATE METHODS OF SPECTRAL ANALYSIS

M. A. Khodasevich*, D. A. Borisevich (B. I. Stepanov Institute of Physics of the National Academy of Sciences of Belarus, 68-2 Nezavisimosti Prosp., Minsk, 220072, Belarus; e-mail: m.khodasevich@ifanbel.bas-net.by)

Цель настоящей работы — построение регрессионной линейной многопараметрической модели, основанной на методах главных компонент (PCA — Principal Component Analysis) [4], кластерном анализе и проекции на латентные структуры (PLS — Projection to Latent Structures) [5] и позволяющей по спектрам УФ, видимого и ближнего ИК диапазонов количественно определять подделку оригинального льняного масла нерафинированным подсолнечным маслом. Такой подход, при котором аналитические методы, требующие дорогостоящего оборудования и больших затрат труда, заменяются на более быстрые и дешевые косвенные методы, наиболее ярко иллюстрирует современную тенденцию использования спектроскопии [6].

Эксперимент. В качестве объектов исследования подготовлен 31 образец бинарных смесей подсолнечного (“Жемчужина Поволжья”, изготовитель ООО ТД “Жемчужина Поволжья”, Россия, дата розлива 12.10.2018 г.) и льняного (“Деревенское”, изготовитель ООО “Клуб Фарм-Эко”, Беларусь, дата розлива 04.09.2018 г.) масел с содержанием от 0 до 100 %. Спектры поглощения зарегистрированы с помощью спектрофотометра Shimadzu UV-3101PC (диапазон 190—3200 нм, разрешение 0.1 нм). Измерения проводились с шагом 1 нм в диапазонах 335—870 нм при ширине щели 1 нм и 870—2200 нм при ширине 3 нм. На рис. 1 представлены спектры поглощения образцов с различным содержанием льняного масла. Видно отсутствие в ближнем ИК диапазоне “зашумленного” участка 1698—1766 нм, где оптическая плотность $D > 3$. В этой полосе лежит первый обертоном С-Н колебаний группы $-\text{CH}_2-$ [7, 8]. Коротковолновую часть ИК диапазона измерений можно ограничить $\lambda = 1130$ нм. Штриховкой на оси абсцисс отмечены три спектральных интервала, которые использованы для построения многопараметрической калибровочной модели: 350—690, 1130—1698 и 1766—2200 нм. Следует отметить, что в приведенных на рис. 1, б обзорных ИК спектрах D бинарных смесей масел различия не столь очевидны, как в спектрах видимого диапазона (рис. 1, а), наибольшая разница достигается вблизи максимума на $\lambda = 2145$ нм и составляет 0.36. Однако два учитываемых при построении многопараметрической модели интервала в ИК диапазоне существенны, что показано ниже.

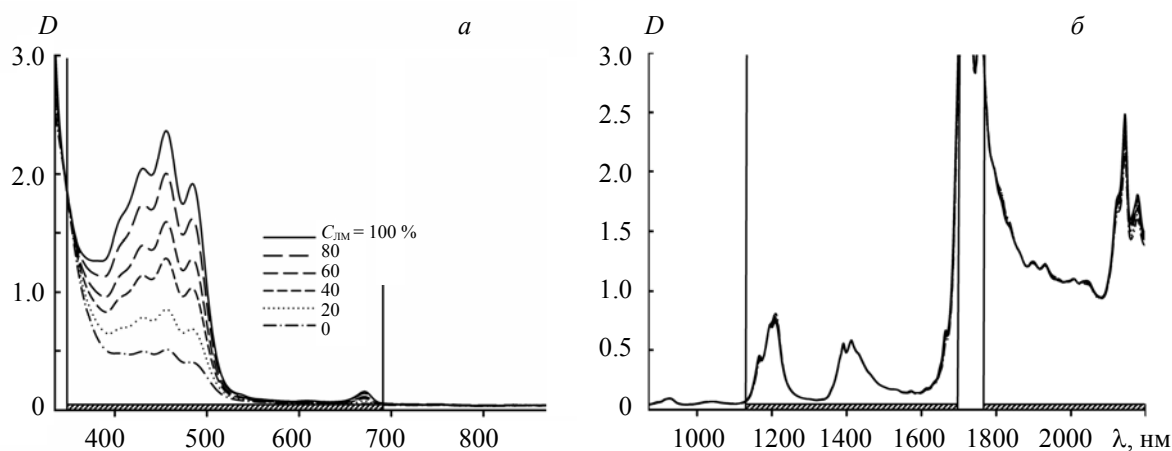


Рис. 1. Спектральная зависимость оптической плотности образцов бинарных смесей льняного масла концентрацией $C_{\text{лм}}$ и подсолнечного масла в различных диапазонах

Обработка спектров и построение калибровочных моделей. После выбора спектральных диапазонов для анализа методом PCA сформирована прямоугольная матрица спектров исследуемых образцов. Линейная аппроксимация зависимости счетов в первую главную компоненту, которая описывает 99.7 % суммарной объясненной дисперсии данных, от концентрации льняного масла в смеси показала, что счета образцов с концентрациями 27,5, 30, 35, 40, 70, 75 и 90 % значительно отклоняются от общей зависимости, а следовательно, являются выбросами и должны быть изъяты из рассмотрения. Построение модели PLS требует разбиения оставшихся образцов на обучающую и проверочную выборки, которое проведено с помощью иерархического кластерного анализа в евклидовом пространстве первой главной компоненты спектров оптической плотности. Такой подход показывает лучшие результаты [9] по сравнению с равномерным разбиением по показателю калибруемого параметра (в рассматриваемом случае по содержанию льняного масла в бинарной смеси) или часто при-

меняемым алгоритмом Кеннарда—Стоуна [10]. На рис. 2 представлены шесть кластеров, группирующих образцы по степени близости счетов в пространстве первой главной компоненты спектров оптической плотности. В проверочную выборку выбраны образцы, счета которых характеризуются минимальным расстоянием до центров соответствующих кластеров. Таким образом, шесть образцов из 24 выделено в проверочную выборку, а оставшиеся 18 сформировали обучающую. Создание проверочной выборки большего размера нецелесообразно из-за ограниченности общего количества образцов.

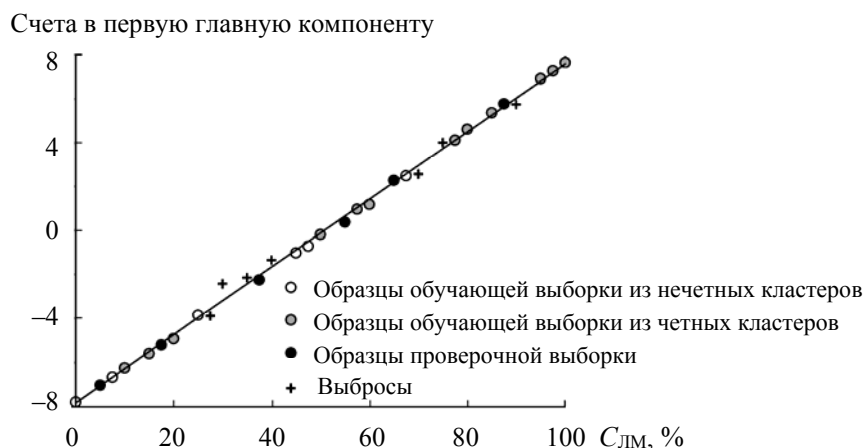


Рис. 2. Обнаружение выбросов и формирование обучающей и проверочной выборок с помощью иерархического кластерного анализа счетов в первую главную компоненту, линейная интерполяция $y = 0.1539x - 7.819$

При построении калибровочных моделей применим три подхода: широкополосный многопараметрический [5], когда используются все имеющиеся измерения или спектральные переменные; многопараметрический с выбором спектральных переменных [11—13], что позволяет в большинстве случаев улучшить качество калибровки; однопараметрический, при котором используются измерения на одной длине волны или в узком спектральном интервале.

Первым применим формальный широкополосный многопараметрический подход, реализованный с помощью метода PLS по всем 1345 спектральным переменным из трех выбранных интервалов. Критерием качества калибровки является величина среднеквадратичной ошибки по проверочной выборке RMSEP (RMSEP — root-mean-square error of prediction). Для широкополосного многопараметрического подхода единственной переменной в модели служит количество латентных структур в PLS, определяемое минимальной RMSEP. Для решаемой задачи $RMSEP_{\min} = 0.63\%$ для шести латентных структур.

Для проведения корректного сравнения результатов калибровки количество латентных структур оставим неизменным на протяжении всего исследования. При выполнении этого условия найденное для широкополосного PLS количество латентных структур определяет и минимальную ширину спектрального интервала для построения многопараметрических моделей с помощью интервальных вариантов PLS (iPLS — interval PLS), улучшающих качество калибровки за счет выбора спектральных переменных [11—13]. Самый простой алгоритм iPLS делит весь используемый спектральный диапазон на последовательные непересекающиеся интервалы, по каждому из которых проводится отдельная калибровка. Содержательный выбор линейной комбинации спектральных переменных в iPLS возможен при условии, что количество латентных структур хотя бы на единицу меньше количества используемых в модели спектральных переменных. Поэтому для рассматриваемого случая минимальная ширина интервала составляет семь спектральных переменных.

Единственный интервал шириной семь спектральных переменных используем и в классической интерпретируемой модели, когда калибровка осуществляется по отдельным характерным максимумам или включающим их узким спектральным интервалам. В видимой области спектра наблюдаются четыре максимума оптической плотности: 431, 456, 484 и 671 нм. Первые три могут быть приписаны каротиноидам [7] и отражают существенную зависимость оптической плотности от состава изучаемого

мой бинарной смеси растительных масел, четвертый — веществам, содержащим хлорофилл. Минимальная величина $RMSEP_{\min} = 0.42\%$ калибровки содержания льняного масла с помощью содержательной модели соответствует интервалу 491—497 нм, расположенному вблизи максимума $\lambda = 484$ нм.

Третий используемый нами подход к проведению калибровки — формальное построение многопараметрической модели с помощью одной из модификаций iPLS — метода поиска комбинации сдвигающихся окон (scmwiPLS — searching combination of moving windows iPLS) [13]. Применяемая нами модификация метода scmwiPLS [9] может быть описана следующим образом. Положение первого спектрального интервала определяется аналогично случаю содержательной модели. Второй интервал последовательно сдвигается в пределах всего диапазона измерений и объединяется с первым также при выполнении условия минимальности величины $RMSEP$ для модели, построенной по объединенному множеству спектральных переменных. Процедура продолжается с последующими интервалами до включения в модель полного диапазона измерений, так как в процессе построения модели ведется поиск глобального минимума зависимости $RMSEP$ от количества учитываемых спектральных переменных. В рассматриваемом случае глобальный минимум соответствует объединению 46 интервалов или 322 спектральным переменным, представленным вместе со спектрами оптической плотности образцов на рис. 3.

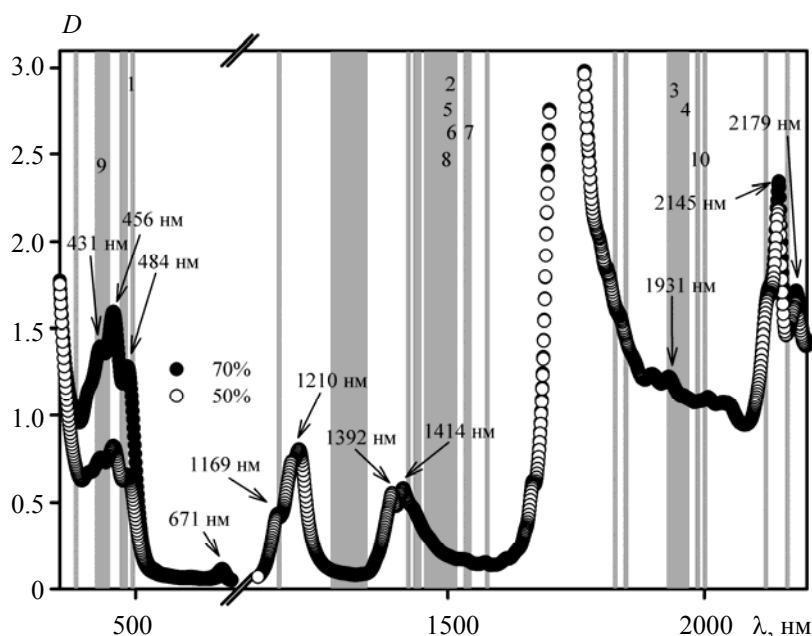


Рис. 3. Спектральные зависимости оптической плотности образцов, содержащих 70 (●) и 50 % (◦) льняного масла, и используемые в модели scmwiPLS 46 спектральных интервалов (первые 10 пронумерованы)

Среднеквадратичная ошибка калибровки концентрации льняного масла в бинарной смеси с подсолнечным маслом для модели scmwiPLS $RMSEP = 0.12\%$. Кроме среднеквадратичной ошибки калибровки важной характеристикой при решении рассматриваемой задачи является предел обнаружения подсолнечного масла (LOD — limit of detection) — наименьшее количество подсолнечного масла в образце, которое может быть обнаружено, но не обязательно точно количественно определено. Величина LOD, определяемая как утроенное отношение отклонения счета в первую главную компоненту при 100 %-й концентрации льняного масла от соответствующего значения указанной на рис. 2 линейной интерполяции и коэффициента наклона этой интерполяции [14], составляет 1.2 %.

В дополнение к указанным выше максимумам оптической плотности исследуемой смеси в видимом диапазоне измеренные спектры содержат обертона и комбинационные частоты валентных и деформационных колебаний входящих в состав растительных масел триглицеридов и жирных кислот в области ближнего ИК диапазона. Можно выделить максимумы при $\lambda = 1169, 1210, 1392, 1414, 1931,$

2145 и 2179 нм. Они могут быть отнесены ко второму обертому валентного колебания С-Н -CH₃-группы (8554 см⁻¹ по измеренным спектрам, 8576 см⁻¹ согласно [7]), второму обертому валентного колебания С-Н -CH₂-группы (8264 и 8276 см⁻¹ [7]), комбинационной частоте валентного колебания 2С-Н и деформационного колебания С-Н -CH₂-группы (7184 и 7188 см⁻¹ [7] или 7185 см⁻¹ [8]), комбинационной частоте валентного колебания 2С-Н и деформационного колебания С-Н -CH₂-группы (7072 и 7080 см⁻¹ [7]), второму обертому валентного колебания С=О (5179 см⁻¹ и по измеренным спектрам и по данным [15]), комбинационной частоте валентных колебаний =С-Н и С=С -НС=СН-группы (измеренное $\nu = 4662$ см⁻¹ и данные [8] совпадают) и комбинационной частоте асимметричного валентного колебания СН и валентного колебания С=С -НС=СН-группы (4589, 4595 см⁻¹ [15] и 4592 см⁻¹ [7]). При этом лишь два из отмеченных максимумов приходятся на спектральные интервалы, включенные в модель scmwPLS (431 и 1931 нм). На рис. 3 представлены серым цветом используемые для калибровки интервалы, а также номера первых 10 из 46 спектральных интервалов, последовательно включаемых в калибровочную модель при выполнении алгоритма scmwPLS (491—497, 1504—1510, 1956—1962, 1963—1969, 1497—1503 нм и т. д.). Следует отметить, что выбранный первым интервал 491—497 нм не пересекается и не примыкает ни к одному другому найденному алгоритмом scmwPLS. Это показывает высокую коллинеарность спектральных данных вблизи максимума поглощения каротиноидов на $\lambda = 484$ нм. Наоборот, выбор второго, пятого, шестого и восьмого по порядку спектральных интервалов в границах самой широкой (70 нм) учитываемой в модели полосы 1273—1342 нм показывает низкую коллинеарность измерений оптической плотности на длинноволновом крыле контура оптической плотности на $\lambda = 1414$ нм.

Закключение. Показано, что использование интервальных многопараметрических методов калибровки содержания льняного масла в фальсификате (смеси льняного и нерафинированного подсолнечного масел) по спектрам поглощения в УФ, видимом и ближнем ИК диапазонах характеризуется меньшей ошибкой (0.12 %) по сравнению с классической содержательной спектроскопической моделью (0.42 %) и многопараметрической моделью, использующей весь измеренный спектр (0.63 %).

Исследование выполнено при поддержке Белорусского республиканского фонда фундаментальных исследований (проект № Ф19-060).

- [1] A. Valdés, A. Beltrán, C. Mellinas, A. Jiménez, M. C. Garrigós. Trends in Food Science & Technology, **77** (2018) 120—130
- [2] И. А. Голуб. Наше сельское хозяйство, № 17 (2016) 44—47
- [3] Д. П. Чирик. Наше сельское хозяйство, № 19 (2016) 21—23
- [4] R. Bro, A. K. Smilde. Analyt. Methods, **6** (2016) 2812—2831
- [5] P. Geladi, B. R. Kowalski. Anal. Chim. Acta, **185** (1986) 1—17
- [6] S. Grassi, C. Alamprese. Curr. Opin. Food Sci., **22** (2018) 17—21
- [7] Z. Li, J. Wang, Y. Xiong, Z. Li, S. Feng. Vibr. Spectrosc., **84** (2016) 24—29
- [8] L. Wang, F. Lee, X. Wang, Y. He. Food Chem., **95** (2006) 529—536
- [9] М. А. Ходасевич, В. А. Асеев. Опт. и спектр., **124**, № 5 (2018) 713—717
- [10] S. Nawar, A. M. Mouazen. Comput. Electron. Agricult., **151** (2018) 469—477
- [11] L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, S. B. Engelsen. Appl. Spectr., **54** (2000) 413—419
- [12] X. Zou, J. Zhao, Y. Li. Vibr. Spectrosc., **44** (2007) 220—227
- [13] Y. P. Du, Y. Z. Liang, J. H. Jiang, R. J. Berry, Y. Ozaki. Anal. Chim. Acta, **501** (2004) 183—191
- [14] A. Shrivastava, V. B. Gupta. Chron. Young Sci., **2** (2011) 21—25
- [15] A. Christy, S. Kasemsumran, Y. Du, Y. Ozaki. Analyt. Sci., **20** (2004) 935—940