

## ИНТЕРВАЛЬНЫЕ ЛИНЕЙНЫЕ МНОГОПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СПЕКТРАЛЬНОГО АНАЛИЗА ДЛЯ ИДЕНТИФИКАЦИИ ФАЛЬСИФИЦИРОВАННОГО ЛЬНЯНОГО МАСЛА

*Применение интервальных методов проекции на латентные структуры для калибровки содержания фальсифицирующей добавки нерафинированного подсолнечного масла в льняном масле по спектрам оптической плотности в УФ-, видимом и ближнем ИК-диапазонах характеризуется меньшей ошибкой (0,12%) по сравнению с классической содержательной моделью (0,42%) и по сравнению с многопараметрической моделью, использующей весь измеренный спектр оптической плотности (0,63%).*

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

*To calibrate the content of falsifying additive of unrefined sunflower oil in linseed oil, the use of interval projection to latent structures using the optical density spectra in the UV, visible and near IR ranges is characterized by less root mean square error (0,12%) compared to the classical spectroscopic model (0,42%) and compared with the multivariate model using the full measured spectrum (0,63%).*

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

В последнее время фальсификация продуктов питания, которая кроме финансовых потерь производителей может вызывать ухудшение состояния здоровья у потребителей, превратилась в серьезную проблему во всем мире. Почти 68% объема контрафакта 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 и 3% других продуктов животного происхождения) [1]. Среди большого разнообразия пищевых растительных масел объектом нашего исследования было выбрано промышленно выпускаемое в Республике Беларусь льняное масло, которое содержит в своем составе полезные для человека моно- и полиненасыщенные кислоты и витамины и занимает одно из первых мест среди других растительных масел по своей биологической ценности.

Широко распространенные аналитические методы исследования качества и оригинальности продукции, требующие больших затрат труда, времени, дорогостоящих оборудования и реактивов, могут быть заменены на

косвенные методы, более быстрые и дешевые. Наиболее ярко эта тенденция проявляется при использовании инфракрасной спектроскопии, особенно в ближней инфракрасной области, прежде считавшейся малополезной из-за высокого и трудно устранимого шума, обусловленного интенсивным поглощением воды и эффектом рассеяния в спектрах отражения [2]. В настоящей работе с помощью спектроскопии ультрафиолетовой, видимой и ближней инфракрасной областей и многопараметрических методов разработаны калибровочные модели, основанные на методах главных компонент (PCA – Principal Component Analysis) [3] и проекции на латентные структуры (PLS – Projection to Latent Structures) [4], для количественной оценки содержания фальсификата в льняном масле. В качестве фальсификата использовалось нерафинированное подсолнечное масло.

Итак, в качестве объектов исследования был подготовлен 31 образец бинарных смесей подсолнечного и льняного масел с процентным содержанием первого от 0% до 100%. Спектры оптической плотности регистрировались с помощью спектрофотометра Shimadzu UV-3101PC (спектральный диапазон 190–3200 нм, разрешение 0,1 нм). Измерения проводились с шагом 1 нм в двух диапазонах: от 335 до 870 нм при ширине щели 1 нм и от 870 до 2200 нм при ширине щели 3 нм. На рисунке 1 представлены спектры оптической плотности образцов с указанным в легенде содержанием подсолнечного масла. На графике спектров ближнего ИК-диапазона виден зашумленный участок 1698–1766 нм, где оптическая плотность превышает 3. В этой полосе лежит первый обертон колебаний С-Н группы  $-\text{CH}_2-$  [Li2016, wang2006]. Кроме того, коротковолновую часть инфракрасного диапазона измерений можно ограничить длиной волны 1130 нм. Штриховкой на оси абсцисс отмечены три спектральных интервала, которые использовались для построения многопараметрической калибровочной модели: 350–690 нм, 1130–1698 нм и 1766–2200 нм.

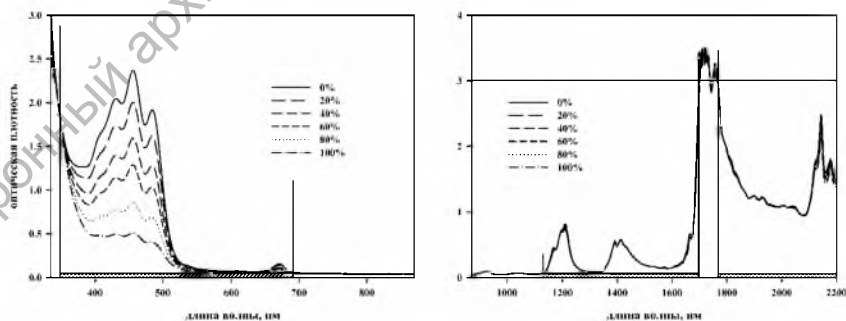


Рис. 1. Спектры оптической плотности образцов бинарных смесей подсолнечного и льняного масел

После выбора спектральных диапазонов сформированная матрица спектров была проанализирована методом PCA на наличие выбросов с помощью графика счетов в первую главную компоненту, которая описывает 99,72% суммарной объясненной дисперсии данных, в зависимости от концентрации подсолнечного масла в смеси (рис. 2). Видно, что выбросами являются 7 отмеченных крестиками образцов с концентрациями 10%, 25%, 30%, 60%, 65%, 70% и 72,5%. Для дальнейшего моделирования их спектры были изъяты из рассмотрения.

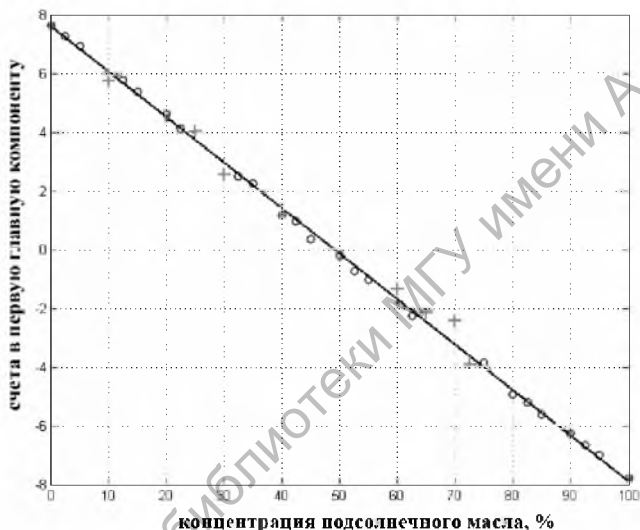


Рис. 2. График зависимости счетов в первую главную компоненту от концентрации подсолнечного масла в смеси

Построение проекции на латентные структуры требует разбиения оставшихся спектров на обучающую и проверочную выборки. С помощью иерархического кластерного анализа в пространстве первой главной компоненты с евклидовой метрикой 6 спектров из 24 было выделено в проверочную выборку, а 18 – в обучающую. На рисунке 3 представлены кластеры и крестиками отмечены выбранные в проверочную выборку спектры, счета которых характеризуются минимальным расстоянием до центров соответствующих кластеров.

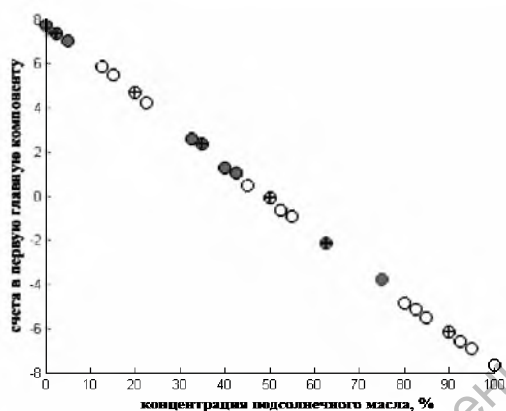


Рис. 3. Формирование обучающей и проверочной выборок с помощью иерархического кластерного анализа счетов в первую главную компоненту

Для применения метода поиска комбинации движущихся окон для интервальной проекции на латентные структуры (scmwiPLS – searching combination of moving windows interval PLS) [5; 6] следует определить ширину минимального спектрального интервала с помощью PLS по всему диапазону спектра без выбора спектральных переменных. Искомая ширина в спектральных отсчетах на единицу превышает оптимальное количество латентных структур в широкополосной PLS, которое в свою очередь определяется минимальной величиной среднеквадратичной ошибки калибровки по проверочной выборке (RMSEP – root mean square error of prediction), равной 6. Следовательно, ширина движущегося окна должна быть равна 7.

Применяемая нами модификация метода scmwiPLS может быть описана следующим образом. Положение первого спектрального окна определяется минимальной величиной RMSEP для моделей, построенных по непересекающимся интервалам шириной 7 отсчетов, и фиксируется. Следующее окно последовательно сдвигается в пределах всего спектрального диапазона измерений и объединяется с первым также при выполнении условия минимальности величины RMSEP для модели, построенной по объединенному множеству спектральных отсчетов. Процедура продолжается до включения в модель полного диапазона измерений, так как идет поиск глобального минимума зависимости величины RMSEP от количества учитываемых в модели спектральных окон. В рассматриваемом случае глобальный минимум соответствует 46 спектральным окнам, представленным вместе со спектрами оптической плотности образцов на рисунке 4.

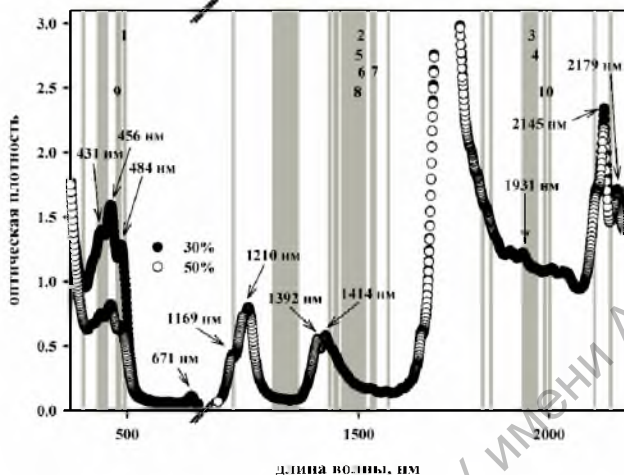


Рис. 4. Спектры оптической плотности образцов с указанными пиками и обозначенные серым цветом спектральные окна, первые 10 из которых пронумерованы

Величина среднеквадратичной ошибки калибровки концентрации подсолнечного масла в бинарной смеси с льняным маслом для модели PLS по выделенным на рисунке 4 322 спектральным отсчетам составляет 0,12%. Величина RMSEP для широкополосной PLS равна 0,63%, а для PLS по одному окну – 0,42%.

В видимой области спектра есть четыре пика оптической плотности: 431 нм, 456 нм, 484 нм и 671 нм. Первые три можно приписать содержащимся в смеси растительных масел каротиноидам, четвертый – веществам, содержащим хлорофилл [1].

Область ближнего ИК-диапазона содержит обертона и комбинационные частоты валентных и деформационных колебаний входящих в состав растительных масел триглицеридов и жирных кислот. Здесь можно выделить пики на длинах волн 1169 нм, 1210 нм, 1392 нм, 1414 нм, 1931 нм, 2145 нм и 2179 нм. Они могут быть отнесены ко второму обертому валентного колебания С-Н группы  $-\text{CH}_3$  ( $8554 \text{ см}^{-1}$  по измеренным спектрам,  $8576 \text{ см}^{-1}$  согласно [7]), второму обертому валентного колебания С-Н группы  $-\text{CH}_2$  ( $8264 \text{ см}^{-1}$  и  $8276 \text{ см}^{-1}$  [7], соответственно), комбинационной частоте валентного колебания 2С-Н и деформационного колебания С-Н группы  $-\text{CH}_2$  ( $7184 \text{ см}^{-1}$  и  $7188 \text{ см}^{-1}$  [7] или  $7185 \text{ см}^{-1}$  [8]), комбинационной частоте валентного колебания 2С-Н и деформационного колебания

C-H группы  $-\text{CH}_2$  ( $7072 \text{ cm}^{-1}$  и  $7080 \text{ cm}^{-1}$  [7]), второму обертому валентного колебания  $\text{C}=\text{O}$  ( $5179 \text{ cm}^{-1}$  по измеренным спектрам и по литературным данным [9]), комбинационной частоте валентных колебаний  $=\text{C}-\text{H}$  и  $\text{C}=\text{C}$  группы  $-\text{HC}=\text{CH}-$  (измеренная величина волнового числа  $4662 \text{ cm}^{-1}$  и литературные данные [8] совпадают) и комбинационной частоте асимметричного валентного колебания  $\text{CH}$  и валентного колебания  $\text{C}=\text{C}$  группы  $-\text{HC}=\text{CH}-$  ( $4589 \text{ cm}^{-1}$ ,  $4595 \text{ cm}^{-1}$  [9] и  $4592 \text{ cm}^{-1}$  [7], соответственно). При этом лишь два из отмеченных пиков приходятся на спектральные интервалы, включенные в модель scmwPLS (431 нм и 1931 нм). Эти интервалы представлены серым цветом на рисунке 4. В верхней части рисунка 4 обозначены порядковые номера первых 10 из 46 спектральных окон шириной 7 отсчетов, последовательно включаемых в калибровочную модель при выполнении алгоритма scmwPLS (491–497 нм, 1504–1510 нм, 1956–1962 нм, 1963–1969 нм, 1497–1503 нм и т. д.). Следует отметить, что выбранное первым спектральное окно 491–497 нм не пересекается и не примыкает ни к одному другому найденному алгоритмом scmwPLS. Это показывает высокую коллинеарность спектральных отсчетов вблизи длинноволнового пика поглощения каротиноидов. И наоборот, выбор второго, пятого, шестого и восьмого по порядку спектральных окон 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 длине волны 1414 нм.

Таким образом, показано, что использование интервальных методов PLS для калибровки содержания фальсификата (нерафинированного подсолнечного масла) в льняном масле по спектрам оптической плотности в УФ-, видимом и ближнем ИК-диапазонах характеризуется меньшей ошибкой (0,12%) по сравнению с классической содержательной спектроскопической моделью (0,42%) и по сравнению с многопараметрической моделью, использующей весь измеренный спектр (0,63%).

Исследование выполнено при поддержке Белорусского республиканского фонда фундаментальных исследований (проект Ф19-060).

### Литература

1. Valdés, A., Beltrán, A., Mellinas, C., Jiménez, A., Garrigós, M.C. // Trends in Food Science & Technology. – 2018. – V. 77. – P. 120–130.
2. Родионова, О.Е. // Российский химический журнал. – 2006. – № 2. – С. 128–144.
3. Esbensen, K.H., Geladi P. // Compr. Chemom. – 2009. – V. 2. – P. 211–226.
4. Abdi, H. // Wiley Interdisciplinary Reviews: Computational Statistics. – 2010. – V. 2. – P. 97–106.

5. Du Y. P., Liang Y. Z., Jiang J. H., Berry R. J., Ozaki Y. // *Anal. Chim. Acta.* – 2004. – V. 501. – P. 183–191.
6. Khodasevich M.A., Aseev V.A. // *Optics and Spectroscopy.* – 2018. – V. 124. – P. 748–752.
7. Li Z., Wang J., Xiong Y., Li Z., Feng S. // *Vibrational Spectroscopy.* – 2016. – V. 84. – P. 24–29.
8. Wang L., Lee F., Wang X., He Y. // *Food Chemistry.* – 2006. – V. 95. – P. 529–536.
9. Christy A., Kasemsumran S., Du Y., Ozaki Y. // *Analytical sciences.* – 2004. – V. 20. – P. 935–940.