

ISSN 1561-2430 (Print)
ISSN 2524-2415 (Online)
УДК 535.372+519.237.8

Поступила в редакцию 01.06.2017
Received 01.06.2017

М. А. Ходасевич, Н. А. Саскевич

Институт физики им. Б. И. Степанова Национальной академии наук Беларуси, Минск, Беларусь

МЕТОДЫ ПОСТРОЕНИЯ ОБУЧАЮЩЕГО НАБОРА ДЛЯ КАЛИБРОВКИ С ПОМОЩЬЮ ФЛУОРЕСЦЕНТНОЙ СПЕКТРОСКОПИИ НЕБОЛЬШИХ ВЫБОРОК ОБРАЗЦОВ

Аннотация. При проведении калибровки с помощью многопараметрического спектрального анализа ограниченность количества образцов и невозможность априорного контроля величины искомого параметра обуславливают важность решения задачи построения обучающего набора с целью уменьшения ошибки калибровки. На примере калибровки температуры с помощью проекции на латентные структуры в диапазоне от 66 до 150 °С по спектрам флуоресценции $\text{Yb}^{3+}:\text{CaF}_2$, зарегистрированным в диапазоне 880–1120 нм с разрешением около 0,2 нм и шагом по температуре 2 °С, продемонстрированы возможные варианты построения обучающего набора из небольших по количеству образцов выборок с применением равномерного распределения, алгоритма Кеннарда – Стоуна и методов кластерного анализа в пространстве главных компонент, показано их влияние на точность калибровки. Применение метода главных компонент позволяет проводить отбор спектров без привлечения априорных знаний о температуре, которой соответствуют спектры флуоресценции. Показано, что среднеквадратичная ошибка предсказанной величины температуры при равномерном распределении образцов обучающего набора по пространству первой главной компоненты составляет 3,98 °С, при использовании алгоритма Кеннарда – Стоуна – 1,07 °С. Минимальная среднеквадратичная ошибка 0,98 °С может быть достигнута с помощью иерархического кластерного анализа пространства главных компонент исследуемых спектров.

Ключевые слова: флуоресцентная спектроскопия, калибровка, кластерный анализ, проекция на латентные структуры

Для цитирования. Ходасевич, М. А. Методы построения обучающего набора для калибровки с помощью флуоресцентной спектроскопии небольших выборок образцов / М. А. Ходасевич, Н. А. Саскевич // Вест. Нац. акад. наук Беларусі. Сер. фіз.-мат. навук. – 2017. – Т. 54, № 1. – С. 77–83.

М. А. Khodasevich, N. A. Saskevich

B. I. Stepanov Institute of Physics of the National Academy of Sciences of Belarus, Minsk, Belarus

TRAINING SUBSET SELECTION METHODS FOR CALIBRATION WITH FLUORESCENCE SPECTROSCOPY IN SMALL DATA SETS OF SAMPLES

Abstract. A limited number of samples and an impossible *a priori* control of a desired parameter value stipulate how it is important to solve the problem of selecting a training set for calibration by the multivariate spectral analysis in order to reduce a calibration error. Possible variants of a training subset selection from small data sets are shown for temperature calibration with fluorescence spectra of $\text{Yb}^{3+}:\text{CaF}_2$ recorded in the range of 880–1120 nm with a resolution of about 0.2 nm for the temperature range from 66 to 150 °C and at a step of 2 °C. The methods applied are the uniform distribution, the Kennard and Stone algorithm and the cluster analysis in principal component space. The effect of the method choice on the calibration accuracy has been evaluated. The application of the principal component analysis gives the possibility to select spectra without *a priori* knowledge of a temperature, to which the fluorescence spectra correspond. The root-mean-square error of the predicted temperature value is shown to be 3.98 °C for the uniform distribution of the training subset samples over the space of the first principal component and 1.07 °C for the Kennard and Stone algorithm. The minimum root-mean-square prediction error of 0.98 °C is shown to be achieved with the training subset selection by the hierarchical cluster analysis in the space of the principal components of the spectra studied.

Keywords: fluorescence spectroscopy, calibration, cluster analysis, projection to latent structures

For citation. Khodasevich M. A., Saskevich N. A. Training subset selection methods for calibration with fluorescence spectroscopy in small data sets of samples. *Vesti Natsyional'noi akademii navuk Belarusi. Seriya fizika-matematichnykh navuk = Proceedings of the National Academy of Sciences of Belarus. Physics and Mathematics series*, 2017, vol. 54, no. 1, pp. 77–83 (in Russian).

В последние годы метод калибровки [1] по спектральным данным становится все более востребованным в различных областях науки и производства: в экологии [2], медицине [3, 4], фармацевтике [5, 6], химической [7] и пищевой [8] (в том числе винодельческой [9–11]) промышленности, сельском хозяйстве [12, 13], биоэнергетике [14] и др.

Калибровка искомого параметра исследуемой системы по спектральным данным состоит в замене непосредственных измерений этого параметра на измерения спектров системы с последующей многопараметрической обработкой полученной информации. Для проведения калибровки необходимо иметь достаточный набор образцов, для которого величина искомого параметра известна априори. Построение такого обучающего набора является одним из важнейших этапов планирования эксперимента по калибровке. Оптимальный обучающий набор должен достаточно равномерно покрывать весь диапазон возможного изменения параметра. Для проведения тест-валидации математической модели необходим также проверочный набор, не совпадающий с обучающим. При невозможности целенаправленного отбора образцов с априори известным значением параметра на стадии планирования эксперимента построение обучающего и проверочного наборов происходит уже после проведения спектральных измерений. При этом самым надежным способом уменьшить ошибку предсказания величины параметра с помощью калибровки является использование большого и разнообразного обучающего набора [15]. В общем случае ограниченность количества образцов для выполнения спектральных измерений и невозможность априорного контроля величины искомого параметра приводит к необходимости решения задачи построения обучающего набора с целью уменьшения ошибки калибровки.

Мы будем решать эту задачу на примере калибровки температуры допированного иттербием фторида кальция по его спектрам флуоресценции с помощью проекции на латентные структуры – PLS (projection to latent structures) [16]. В рассматриваемом случае каждый спектр флуоресценции соответствует определенной температуре и является отдельным образцом с точки зрения применения многопараметрических методов обработки данных.

Спектры флуоресценции $\text{Yb}^{3+}:\text{CaF}_2$ были зарегистрированы [17] в диапазоне 880–1120 нм с разрешением около 0,2 нм при температуре от 66 до 150 °С с шагом 2 °С. Температура устанавливалась с точностью 0,1 °С с помощью печи PV10 («Covercion Ltd», Великобритания) с контроллером TS-200 («Thorlabs», США). Остальные экспериментальные параметры для целей нашего исследования несущественны.

Спектры флуоресценции были объединены в матрицу X размерами 43×1226 , где 43 – количество спектров, зарегистрированных при разной температуре, 1226 – количество отсчетов на разных длинах волн. Предобработка данных включала в себя приведение к нулю среднего значения отсчетов на каждой длине волны. К матрице спектральных данных вначале был применен метод главных компонент – PCA (principal component analysis) [18], который широко используется как метод исследовательского анализа данных, уменьшения размерности и графического представления в пространстве малой размерности, обнаружения выбросов, классификации и регрессии. В большинстве случаев PCA первым применяется к многопараметрическому набору данных, несмотря на то что впоследствии могут использоваться более сложные подходы.

Величина дисперсии данных в спектрах флуоресценции, объясненной первой главной компонентой (99,95 %), показывает, что практически все изменения в рассматриваемой системе могут быть описаны одним параметром – температурой. Анализ зависимости счетов в первую главную компоненту от температуры выявил выброс в экспериментальных данных при 80 °С. Соответствующий спектр был удален из дальнейшего рассмотрения, и матрица X стала иметь размеры 42×1226 .

Далее к измеренным спектрам флуоресценции был применен метод PLS – билинейный статистический метод, оперирующий матрицей предикторов X и вектором Y откликов. PLS находит в многомерном пространстве предикторов маломерное пространство так называемых латентных переменных, матрицы проекций X и Y в которое имеют максимальную ковариацию. Используем предложенный в [19] подход, где в качестве предикторов рассматривались спектры флуоресценции активированной ионами эрбия свинцово-фторидной наностеклокерамики, а в качестве отклика – температура.

Для отбора спектров в обучающий набор при проведении калибровки часто применяют равномерное покрытие области изменений параметра. В нашей работе вместо диапазона изменений температуры будет рассматриваться область изменений счетов в первую главную компоненту. Это позволяет проводить отбор спектров без привлечения априорных знаний о температуре, которой соответствуют спектры флуоресценции.

Из 42 спектров, оставшихся после изъятия выброса, 7 будут выбираться в обучающий набор, остальные – в проверочный. Малое количество образцов в обучающем наборе в данном случае обусловлено регулярным характером изменений температуры (шаг между измерениями составляет 2 °C).

При равномерном отборе образцов обучающего набора два спектра соответствуют экстремальным значениям счетов, остальные выбраны ближайшими к пяти равномерно распределенным точкам в пространстве первой главной компоненты. Для определения количества латентных структур при нахождении температуры методом PLS была проанализирована зависимость среднеквадратичной ошибки калибровки по обучающему набору RMSEC (root mean square error of calibration) и тест-валидации по проверочному набору RMSEP (root mean square error of prediction). При увеличении количества латентных структур, учитываемых при моделировании, RMSEC уменьшается монотонно, а RMSEP, достигнув локального минимума для трех латентных структур, далее практически не уменьшается. Поэтому всегда в дальнейшем для того, чтобы избежать переопределенности PLS-моделирования, будем учитывать только три латентные структуры.

На рис. 1 представлен график предсказанной величины температуры от измеренной для построения равномерного обучающего набора (точки изображены сплошными кружками) спектров флуоресценции при применении PLS по трем латентным структурам. Как видим, распределение образцов обучающего набора в пределах исследуемого диапазона температуры равномерное. Величина RMSEP для результатов калибровки составляет 3,98 °C.

Альтернативой равномерному распределению образцов обучающего набора выступает алгоритм Кеннарда – Стоуна [20]. Первым в обучающий набор выбирается образец, ближайший к центру диапазона изменения параметра, каждый последующий – должен быть наиболее удаленным, согласно используемой метрике, от уже выбранных. Мы применим евклидово расстояние в пространстве первой главной компоненты для реализации алгоритма Кеннарда – Стоуна при условии предварительной выборки двух спектров, соответствующих экстремальным

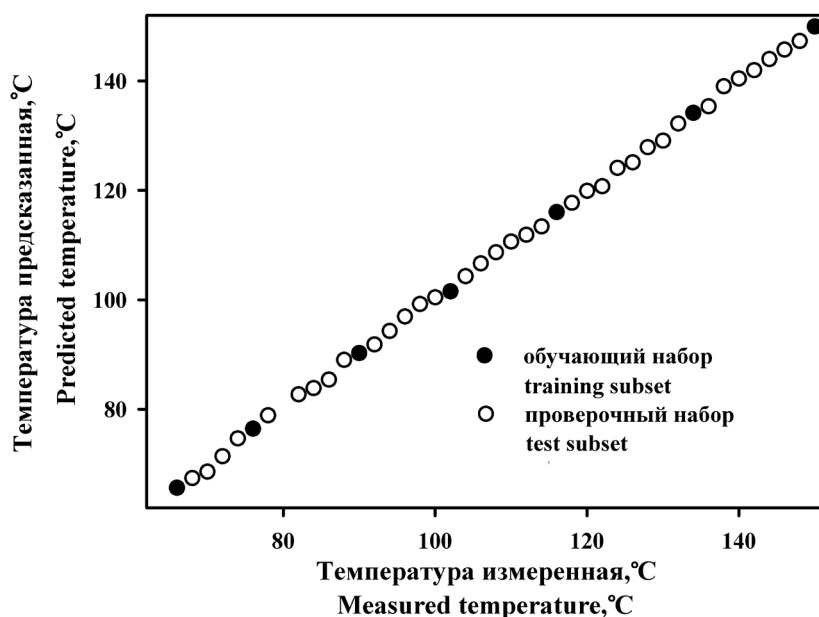


Рис. 1. Предсказанная с помощью PLS величина температуры в зависимости от измеренной для равномерного выбора обучающего набора

Fig. 1. Temperature predicted by PLS vs. the measured temperature for a uniform distribution of a training subset

величинам счетов. Величина RMSEP для результатов калибровки данным образом составляет $1,07\text{ }^{\circ}\text{C}$, что существенно меньше среднеквадратичной ошибки при равномерном выборе образцов.

От описанных выше алгоритмов построения обучающего набора существенно отличаются методы [21], базирующиеся на применении кластерного анализа [22]. Кластерный анализ – многомерная статистическая процедура, упорядочивающая объекты в однородные группы по некоторым общим признакам, – позволяет значительно упростить решение многих прикладных задач, когда исследователи сталкиваются с существенными сложностями при определении степени сходства объектов. В рассматриваемом случае данный подход позволит разделить все зарегистрированные спектры флуоресценции на несколько непересекающихся кластеров. По одному объекту из каждого кластера будет отобрано в обучающий набор.

Мы применим иерархический анализ, поэтапно объединяющий ближайшие объекты (или кластеры) в более крупные кластеры; метод k -средних, перераспределяющий образцы по кластерам с целью минимизации суммарного квадратичного отклонения объектов каждого кластера от его центра; метод разделения гауссовых многопараметрических смесей, относящий элементы выборки случайных и независимых измерений к известному количеству компонент и определяющий центр и дисперсию каждой компоненты в предположении нормального распределения расстояний до центров кластеров.

Все описанные процедуры анализа будут проводиться в пространстве первой главной компоненты спектров флуоресценции при условии обязательного выбора двух экстремальных точек и выделения пяти кластеров. Таким образом, обучающий набор также будет содержать семь образцов. Из спектров, принадлежащих одному кластеру, будет выбран образец, ближайший к центру кластера, в отличие от [21], где для лучшего учета разнообразия брались образцы, наиболее удаленные от центра. Подход из [21] в нашем случае может ухудшить качество калибровки при рассмотрении одномерного пространства главных компонент по причине возможного выбора спектров, соседних по величине счетов в первую главную компоненту.

На рис. 2 представлена зависимость предсказанной при помощи PLS по трем латентным структурам величины температуры от измеренной для построения обучающего набора с помощью иерархического кластерного анализа спектров флуоресценции в пространстве первой главной компоненты.

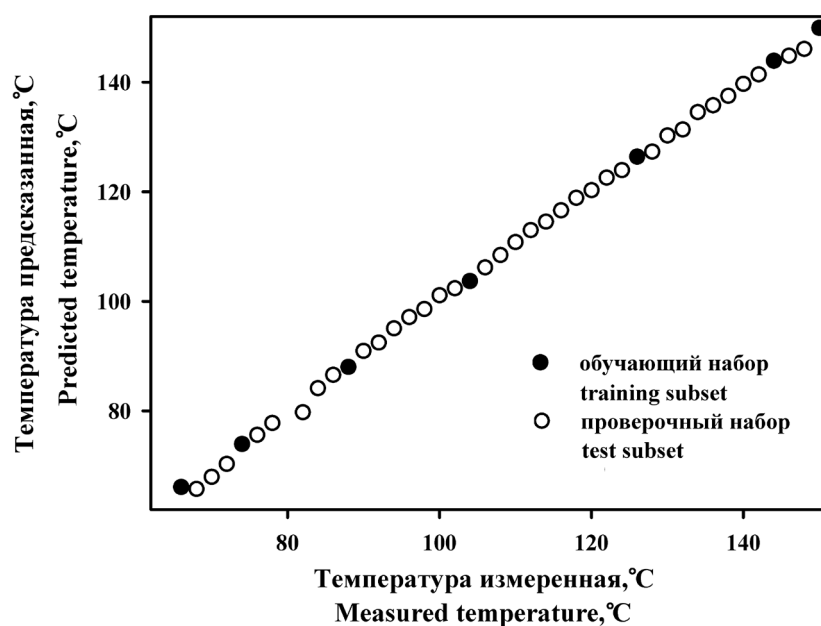


Рис. 2. Предсказанная величина температуры в зависимости от измеренной для построения обучающего набора с помощью иерархического кластерного анализа

Fig. 2. Temperature predicted by PLS vs. the measured temperature for selection of a training subset by the hierarchical cluster analysis

Изменившееся по сравнению с результатами, представленными на рис. 1, распределение образцов обучающего набора привело к уменьшению величины RMSEP до 0,98 °С. Данные, полученные при применении остальных рассмотренных методов выбора образцов обучающего набора, представлены в таблице. Их сравнение показывает, что интуитивно предпочитаемый равномерный выбор образцов обучающего набора для калибровки по спектрам флуоресценции оказывается наихудшим с точки зрения среднеквадратичной ошибки предсказания величины температуры. Широко применяемый алгоритм Кеннарда – Стоуна позволяет существенно уменьшить ошибку моделирования. Среди других методов кластеризации агломеративный иерархический анализ дает возможность понизить среднеквадратичную ошибку до величины 0,98 °С.

**Среднеквадратичные ошибки калибровки температуры с помощью PLS
при различных алгоритмах построения обучающего набора спектров флуоресценции**

**Root-mean-square errors of temperature calibration with PLS
for various algorithms of selecting a training set of fluorescence spectra**

Метод Method	Равномерный выбор обучающего набора The uniform distribution of the training subset samples	Алгоритм Кеннарда – Стоуна The Kennard and Stone algorithm	Иерархический кластерный анализ Hierarchical cluster analysis	Метод разделения гауссовых смесей Clustering based on Gaussian mixture models	Метод <i>k</i> -средних <i>k</i> -means method
RMSEP, °С	3,98	1,07	0,98	1,40	1,14

Таким образом, наилучших результатов по величине среднеквадратичной ошибки при калибровке параметра по спектральным данным относительно небольших по количеству образцов выборок можно достичь построением обучающих наборов с помощью иерархического кластерного анализа пространства главных компонент исследуемых спектров.

Список использованных источников

1. Kalivas, J. H. Calibration Methodologies / J. H. Kalivas // *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis* / editors-in-chief: R. Tauler, B. Walczak, S. D. Brown. – Elsevier, 2009. – P. 1–32.
2. Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy / Tao Chen [et al.] // *Environ. Pollut.* – 2015. – Vol. 206. – P. 217–226.
3. Noninvasive and fast measurement of blood glucose in vivo by near infrared (NIR) spectroscopy / Xue Jintao [et al.] // *Spectrochim. Acta Part A: Molecular and Biomolecular Spectroscopy*. – 2017. – Vol. 179. – P. 250–254.
4. Joint analyses model for total cholesterol and triglyceride in human serum with near-infrared spectroscopy / Lijun Yao [et al.] // *Spectrochim. Acta Part A: Molecular and Biomolecular Spectroscopy*. – 2016. – Vol. 159. – P. 53–59.
5. Da Silva, V. H. Portable near-infrared instruments: Application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer / V. H. da Silva, J. J. da Silva, C. F. Pereira // *J. Pharm. Biomed. Anal.* – 2017. – Vol. 134. – P. 287–294.
6. Identification of anisodamine tablets by Raman and near-infrared spectroscopy with chemometrics / Lian Li [et al.] // *Spectrochim. Acta Part A: Molecular and Biomolecular Spectroscopy*. – 2014. – Vol. 127. – P. 91–97.
7. Sulub, Y. Determination of polymer blends composed of polycarbonate and rubber entities using near-infrared (NIR) spectroscopy and multivariate calibration / Y. Sulub, J. DeRudder // *Polym. Test.* – 2013. – Vol. 32, № 4. – P. 802–809.
8. Application of benchtop and portable near-infrared spectrometers for predicting the optimum harvest time of *Verbena officinalis* / C. K. Pezzei [et al.] // *Talanta*. – 2017. – Vol. 169. – P. 70–76.
9. Comparing predictive ability of laser-induced breakdown spectroscopy to visible near-infrared spectroscopy for soil property determination / M. Knadel [et al.] // *Biosystems Engineering*. – 2017. – Vol. 156. – P. 157–172.
10. Eating quality evaluation of Khao Dawk Mali 105 rice using near-infrared spectroscopy / P. Siriphollakul [et al.] // *LWT – Food Science and Technology*. – 2017. – Vol. 79. – P. 70–77.
11. Monitoring of alcohol strength and titratable acidity of apple wine during fermentation using near-infrared spectroscopy / B. Peng [et al.] // *LWT – Food Science and Technology*. – 2016. – Vol. 66. – P. 86–92.
12. Определение химических характеристик и производителя дивинов по их широкополосным спектрам пропускания / М. А. Ходасевич [и др.] // *Оптика и спектроскопия*. – 2016. – Т. 120, № 6. – С. 1046–1051.
13. Применение многопараметрического анализа широкополосных спектров пропускания для идентификации вин с географическим наименованием происхождения / М. А. Ходасевич [и др.] // *Приборы и методы измерений*. – 2016. – Т. 7, № 1. – С. 104–113.
14. Alves, J. C. L. Quantification of conventional and advanced biofuels contents in diesel fuel blends using near-infrared spectroscopy and multivariate calibration / J. C. L. Alves, R. J. Poppi // *Fuel*. – 2016. – Vol. 165. – P. 379–388.
15. Clustering and training set selection methods for improving the accuracy of quantitative laser induced breakdown spectroscopy / R. B. Anderson [et al.] // *Spectrochim. Acta Part B*. – 2012. – Vol. 70. – P. 24–32.

16. Abdi, H. Partial least squares regression and projection on latent structure regression / H. Abdi // *Wiley Interdisciplinary Reviews: Computational Statistics*. – 2010. – Vol. 2, № 1. – P. 97–106.
17. Determining the Stark structure of Yb^{3+} energy levels in $\text{Y}_3\text{Al}_5\text{O}_{12}$ and CaF_2 using principal component analysis of temperature dependences of fluorescence spectra / M. Khodasevich [et al.] // *J. Lumin.* – 2017. – Vol. 187. – P. 295–297.
18. Esbensen, K. H. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice / K. H. Esbensen, P. Geladi // *Comprehensive Chemometrics*. – 2009. – Vol. 2. – P. 211–226.
19. Применение регрессии на латентные структуры для определения температуры активированной ионами эрбия свинцово-фторидной наностеклокерамики по спектрам апконверсионной флуоресценции / В. А. Асеев [и др.] // *Оптика и спектроскопия*. – 2015. – Т. 118, № 2. – С. 47–49.
20. Kennard, R. W. Computer aided design of experiments / R. W. Kennard, L. A. Stone // *Technometrics*. – 1969. – Vol. 11, № 1. – P. 137–148.
21. Daszykowski, M. Representative subset selection / M. Daszykowski, B. Walczak, D. L. Massart // *Anal. Chim. Acta*. – 2002. – Vol. 468. – P. 91–103.
22. Мандель, И. Д. Кластерный анализ / И. Д. Мандель. – М.: Финансы и статистика, 1988. – 176 с.

References

1. Kalivas, J. H. Calibration Methodologies. Tauler R., Walczak B., Brown S. D. (eds.). *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Elsevier, 2009, pp. 1–32. Doi: 10.1016/b978-044452701-1.00072-7
2. Tao Chen, Qingrui Chang, Clevers J.G.P.W., Kooistra L. Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy. *Environmental Pollution*, 2015, vol. 206, pp. 217–226. Doi: 10.1016/j.envpol.2015.07.009
3. Xue Jintao, Ye Liming, Liu Yufei, Li Chunyan, Chen Han. Noninvasive and fast measurement of blood glucose in vivo by near infrared (NIR) spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2017, vol. 179, pp. 250–254. Doi: 10.1016/j.saa.2017.02.032
4. Lijun Yao, Ning Lyu, Jiemei Chen, Tao Pan, Jing Yu. Joint analyses model for total cholesterol and triglyceride in human serum with near-infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2016, vol. 159, pp. 53–59. Doi: 10.1016/j.saa.2016.01.022
5. Da Silva V. H., da Silva J. J., Pereira C. F. Portable near-infrared instruments: Application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer. *Journal of Pharmaceutical and Biomedical Analysis*, 2017, vol. 134, pp. 287–294. Doi: 10.1016/j.jpba.2016.11.036
6. Lian Li, Hengchang Zang, Jun Li, Dejun Chen, Tao Li, Fengshan Wang. Identification of anisodamine tablets by Raman and near-infrared spectroscopy with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2014, vol. 127, pp. 91–97. Doi: 10.1016/j.saa.2014.02.022
7. Sulub Y., DeRudder J. Determination of polymer blends composed of polycarbonate and rubber entities using near-infrared (NIR) spectroscopy and multivariate calibration. *Polymer Testing*, 2013, vol. 32, no. 4, pp. 802–809. Doi: 10.1016/j.polymertesting.2013.03.008
8. Pezzeri C. K., Schonbichler S. A., Kirchler C. G., Schmelzer J., Hussain S., Huck-Pezzeri V. A., Popp M., Krolitzek J., Bonn G. K. Application of benchtop and portable near-infrared spectrometers for predicting the optimum harvest time of *Verbena officinalis*. *Talanta*, 2017, vol. 169, pp. 70–76. Doi: 10.1016/j.talanta.2017.03.067
9. Knadel M., Gislum R., Hermansen C., Peng Y., Moldrup P., de Jonge L. W., Greve M. H. Comparing predictive ability of laser-induced breakdown spectroscopy to visible near-infrared spectroscopy for soil property determination. *Biosystems Engineering*, 2017, vol. 156, pp. 157–172. Doi: 10.1016/j.biosystemseng.2017.01.007
10. Siriphollakul P., Nakano K., Kanlayanarat S., Ohashi S., Sakai R., Rittiron R., Maniwaru P. Eating quality evaluation of Khao Dawk Mali 105 rice using near-infrared spectroscopy. *LWT – Food Science and Technology*, 2017, vol. 79, pp. 70–77. Doi: 10.1016/j.lwt.2017.01.014
11. Peng B., Ge N., Cui L., Zhao H. Monitoring of alcohol strength and titratable acidity of apple wine during fermentation using near-infrared spectroscopy. *LWT – Food Science and Technology*, 2016, vol. 66, pp. 86–92. Doi: 10.1016/j.lwt.2015.10.018
12. Khodasevich M. A., Sinitsyn G. V., Skorbanova E. A., Rogovaya M. V., Kambur E. I., and Aseev V. A. Determination of the Chemical Parameters and Manufacturer of Divins from Their Broadband Transmission Spectra. *Optics and Spectroscopy*, 2016, vol. 120, pp. 978–982. Doi: 10.1134/s0030400x16050155
13. Khodasevich M. A., Skorbanova E. A., Obade L. I., Degtyar N. F., Kambur E. I., Rogovaya M. V. Application of multivariate analysis of transmission spectra to identify wines with protected geographical indication (IGP). *Pribory i metody izmerenii = Devices and Methods of Measurements*, 2016, vol. 7, no. 1, pp. 104–113 (in Russian). Doi: 10.21122/2220-9506-2016-7-1-104-113
14. Alves J. C. L., Poppi R. J. Quantification of conventional and advanced biofuels contents in diesel fuel blends using near-infrared spectroscopy and multivariate calibration. *Fuel*, 2016, vol. 165, pp. 379–388. Doi: 10.1016/j.fuel.2015.10.079
15. Anderson R. B., Bell J. F., Wiens R. C., Morris R. V., Clegg S. M. Clustering and training set selection methods for improving the accuracy of quantitative laser induced breakdown spectroscopy. *Spectrochimica Acta Part B*, 2012, vol. 70, pp. 24–32. Doi: 10.1016/j.sab.2012.04.004
16. Abdi H. Partial least squares regression and projection on latent structure regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, vol. 2, pp. 97–106. Doi: 10.1002/wics.51

17. Khodasevich M., Varaksa Y., Sinitsyn G., Aseev V., Demesh M., Yasukevich A. Determining the Stark structure of Yb^{3+} energy levels in $\text{Y}_3\text{Al}_5\text{O}_{12}$ and CaF_2 using principal component analysis of temperature dependences of fluorescence spectra. *Journal of Luminescence*, 2017, vol. 187, pp. 295–297. Doi: 10.1016/j.jlumin.2017.03.014
18. Esbensen K. H., Geladi P. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice. *Comprehensive Chemometrics*, 2009, vol. 2, pp. 211–226. Doi: 10.1016/b978-044452701-1.00043-0
19. Aseev V. A., Varaksa Yu. A., Kolobkova E. V., Sinitsyn G. V., Khodasevich M. A. Application of Projection on Latent Structures for Determining Temperature of Erbium-Doped Lead Fluoride Nano-Glass-Ceramics from Upconversion Fluorescence Spectra. *Optics and Spectroscopy*, 2015, vol. 118, pp. 727–728. Doi: 10.1134/s0030400x15050033
20. Kennard R. W., Stone L. A. Computer aided design of experiments. *Technometrics*, 1969, vol. 11, no. 1, pp. 137–148. Doi: 10.2307/1266770
21. Daszykowski M., Walczak B., Massart D. L. Representative subset selection. *Analytica Chimica Acta*, 2002, vol. 468, pp. 91–103. Doi: 10.1016/s0003-2670(02)00651-7
22. Mandel I. D. *Cluster analysis*. Moscow, Finansy i statistika Publ., 1988, 176 p. (in Russian).

Информация об авторах

Ходасевич Михаил Александрович – кандидат физико-математических наук, заместитель заведующего лабораторией, Институт физики им. Б. И. Степанова Национальной академии наук Беларуси (пр. Независимости, 68, 220072, г. Минск, Республика Беларусь). E-mail: m.khodasevich@ifanbel.bas-net.by

Саскевич Николай Алексеевич – кандидат физико-математических наук, старший научный сотрудник, Институт физики им. Б. И. Степанова Национальной академии наук Беларуси (пр. Независимости, 68, 220072, г. Минск, Республика Беларусь). E-mail: n.saskevich@ifanbel.bas-net.by

Information about the authors

Mikhail A. Khodasevich – Ph. D. (Physics and Mathematics), Deputy Head of the Laboratory of the B. I. Stepanov Institute of Physics, National Academy of Sciences of Belarus (68, Nezavisimosti Ave., 220072, Minsk, Republic of Belarus). E-mail: m.khodasevich@ifanbel.bas-net.by

Nikolai A. Saskevich – Ph. D. (Physics and Mathematics), Senior Researcher, B. I. Stepanov Institute of Physics, National Academy of Sciences of Belarus (68, Nezavisimosti Ave., 220072, Minsk, Republic of Belarus). E-mail: n.saskevich@ifanbel.bas-net.by