

Алгоритм идентификации веществ по набору спектров вторичного излучения

© Н.С. Васильев, А.Н. Морозов

МГТУ им. Н.Э. Баумана, Москва, 105005, Россия

Предложено для увеличения селективности и надежности химического экспресс-анализа веществ одновременно использовать несколько спектров схожей или различной физической природы, расширить применение метода наименьших квадратов для многодиапазонной методики, для выравнивания вкладов от всех используемых спектров в целевой функции нормировать их либо на величину измеренного спектра, либо на величину эталонного спектра. На основе целевых функций невязки измеренных спектров с эталонными получен классификатор, позволяющий проводить экспресс-анализ веществ. Приведена методика использования для статического фурье-спектрометра видимого и ближнего ультрафиолетового диапазонов для регистрации веществ. Проанализирована эффективность классификаторов путем построения ROC-кривых и выбран оптимальный алгоритм. Полученные результаты могут быть применены для увеличения надежности работы приборов экспресс-анализаторов, а также мобильных лабораторий и комплексов химического мониторинга окружающей среды.

Ключевые слова: спектроскопия, идентификация вещества, химический мониторинг, ультрафиолетовый спектр, метод наименьших квадратов, задача классификации, ROC-анализ.

Благодаря развитию средств регистрации, возбуждения и обработки спектров стало возможным создание экспресс-анализаторов химического состава веществ. Эти приборы применяются в службах мониторинга экологии окружающей среды, в правоохранительных органах и др. Наибольшее применение получили приборы на основе эффекта комбинационного рассеяния света (КРС) ([1–3], фотолюминесценции [4], рентгенолюминесценции [5], хроматографии и масс спектроскопии [6–9], инфракрасной (ИК) абсорбционной спектроскопии [10–13]. Каждый метод имеет свои особенности, которые определяют область его применения. Так, спектрометры КРС, обладая высокой селективностью (информативностью) спектров, имеют малую интенсивность и, как следствие, потребность в длительных экспозициях; кроме того, не все вещества активны для этого вида спектроскопии. Фотолюминесценция имеет высокий квантовый выход, но получение селективных спектров сопряжено с техническими трудностями возбуждения и регистрации электронных переходов в средней и дальней ультрафиолетовой (УФ) области спектра. Рентгенолюминесценция, возникая при «внутренних» переходах электронов, позволяет эффективно исследовать присутствие химических элементов, но не связи, которые они образуют. Хроматография, как и

масс-спектрокопия, являются контактными и пробоотборными методами изучения веществ, что сильно ограничивает их область применения и увеличивает время анализа. Наиболее распространенный физический метод для анализа химического строения молекул газов – абсорбционная ИК-спектрокопия – неприменима для регистрации одноатомных газов.

В данной работе для увеличения надежности и расширения области применения создаваемых экспресс-анализаторов веществ предлагается использовать в одном приборе или в измерительном комплексе несколько видов спектров одновременно. В рамках решаемой задачи использован статический фурье-спектрометр [14], позволяющий регистрировать вещества по их спектрам люминесценции. Физический принцип, лежащий в основе работы этого прибора, – фотолюминесценция веществ при возбуждении электронных переходов коротковолновым УФ-излучением. В качестве разнородных спектров рассматривали спектры от двух различных источников возбуждающего излучения с пиками на длинах волн 280 и 310 нм. Квантовый выход спектров люминесценции веществ зависит от положения их зоны поглощения относительно частоты возбуждающей линии. Наличие или отсутствие спектра люминесценции исследуемого вещества при подсветке излучениями с различными длинами волн предоставляет дополнительную информацию о его структуре. Используются источники излучения с длинами волн 280 и 310 нм, так как в этой области лежат границы зоны поглощения многих ароматических соединений и белков.

В работе [15] предлагается для идентификации веществ применять метод наименьших квадратов для определения количественного критерия соответствия измеренного спектра набору спектров из базы данных. Функционал невязки рассчитывается по следующей формуле:

$$\Phi = \int_{\omega_1}^{\omega_2} \left(a(\omega) - \sum_{i=1}^N x_i b_i(\omega) \right)^2 d\omega. \quad (1)$$

Здесь ω – циклическая частота; $a(\omega)$ – регистрируемый спектр; $b_i(\omega)$ – i -й спектр вещества из базы данных. Аппроксимирующей функцией служит сумма спектров с весовыми коэффициентами x_i . Зная концентрации веществ, для которых были получены эталонные спектры b_i , по значениям x_i в точке $\Phi(x_{\min})$ можно оценить концентрацию их примесей в исследуемом веществе. По найденному значению x_{\min} рассчитывается критерий соответствия двух спектров

$$\rho = \sqrt{1 - \frac{\Phi(x_{\min})}{\Phi(0)}}. \quad (2)$$

Каждому измеренному спектру ставится в соответствие столбец найденных коэффициентов соответствия. При превышении некоторого порога, определяемого эмпирически, образец идентифицируется набором веществ с превышенным коэффициентом соответствия.

В многодиапазонной методике можно использовать несколько спектров различной физической природы. Одним из подходов является объединение их в сумму при построении функционала от невязки. Минимизируемый функционал при этом примет вид

$$\Phi(c) = \frac{1}{N} \sum_{j=1}^N \int_{\omega_{j1}}^{\omega_{j2}} (a_j(\omega) - x b_j(\omega))^2 d\omega. \quad (3)$$

Здесь индекс j снизу означает номер спектра из серии для каждого вещества; всего идентификация проводится по N спектрам. Не ограничивая общности, для простоты в формуле (3) рассматривается функционал для одного вещества в базе данных. Предложенный алгоритм применим для анализа спектров с соизмеримыми амплитудами.

В случае если спектры имеют различную интегральную интенсивность, проводится нормировка. Вопрос о способе нормировки становится не очевидным, если учитывать соотношение между интенсивностями спектров в различных диапазонах. Поэтому нормировку измеренного и сопоставляемого спектров из базы данных следует проводить на одну и ту же величину. Предлагается рассмотреть

два случая: нормировка на величину $A_j = \int_{\omega_{j1}}^{\omega_{j2}} a_j^2 d\omega$ и на величину

$B_j = \int_{\omega_{j1}}^{\omega_{j2}} b_j^2 d\omega$, т. е. нормировка осуществляется на величину измеренного спектра или спектра из базы данных соответственно. Приходим к двум модифицированным функционалам вида

$$\Phi(c) = \frac{1}{N} \sum_{j=1}^N \frac{1}{A_j} \int_{\omega_{j1}}^{\omega_{j2}} [a_j(\omega) - x b_j(\omega)]^2 d\omega; \quad (4)$$

$$\Phi(c) = \frac{1}{N} \sum_{j=1}^N \frac{1}{B_j} \int_{\omega_{j1}}^{\omega_{j2}} [a_j(\omega) - x b_j(\omega)]^2 d\omega. \quad (5)$$

Задача идентификации веществ сводится к экстремальной задаче поиска минимума функционалов вида (1)–(5). Затем по формуле (2) определяется числовой критерий соответствия исследуемого спектра эталонному.

Для экспериментальной апробации методик применяли статический фурье-спектрометр видимого и ближнего УФ-диапазонов [14]. Эксперимент состоял в регистрации набора спектров с помощью статического фурье-спектрометра при возбуждении образцов УФ-излучением с длинами волн 280 и 310 нм. Для проверки эффективности предложенных алгоритмов использовали тестовые вещества: антрацен (№ 1), РОРОР (№ 2), РРО (№ 3), стильбен (№ 4) и триптофан (№ 5). Далее вместо названий веществ будут приводиться их указанные номера.

Для получения эталонных спектров использовалось большое количество исходного вещества (> 10 г), длительные экспозиции (≈ 1 мин) и усреднение по большому числу зарегистрированных спектров (>10). Таким образом, получили спектры исследуемых веществ с высоким отношением сигнал/шум. Для анализа эффективности методик требуется большое число измеренных спектров веществ, моделирующих слабый сигнал при экспресс-анализе. С целью упрощения эксперимента увеличения однотипности результатов и сокращения времени полученные двумерные интерферограммы эталонных спектров искусственно аддитивно зашумлялись гауссовым шумом.

На рис. 1, а представлен эталонный и зашумленные спектры вещества № 1, полученные при подсветке возбуждающим излучением с длиной волны 280 нм; на рис. 1, б – спектры вещества № 4, полученные при подсветке источником возбуждающего излучения с длиной волны 310 нм. Среднеквадратичное отклонение выбирали эмпирически равным 10 000 разрядов 14-битного АЦП. По 10 эталонным спектрам были получены 20 спектров, имитирующих измеренные сильнозашумленные спектры.

Обработка экспериментальных данных заключалась в усреднении двумерных интерферограмм, фурье-преобразовании полученных одномерных интерферограмм, минимизации функционалов вида (1)–(5), получении коэффициентов соответствия по формуле (2) для всех эталонных спектров отдельно. Затем по набору из 20 зашумленных двумерных интерферограмм строили матрицу коэффициентов соответствия размером 10×5 . Номер по горизонтали соответствовал номеру

вещества из базы данных, а по вертикали – номеру веществу из набора измеренных спектров (каждый эталонный спектр зашумлялся дважды).

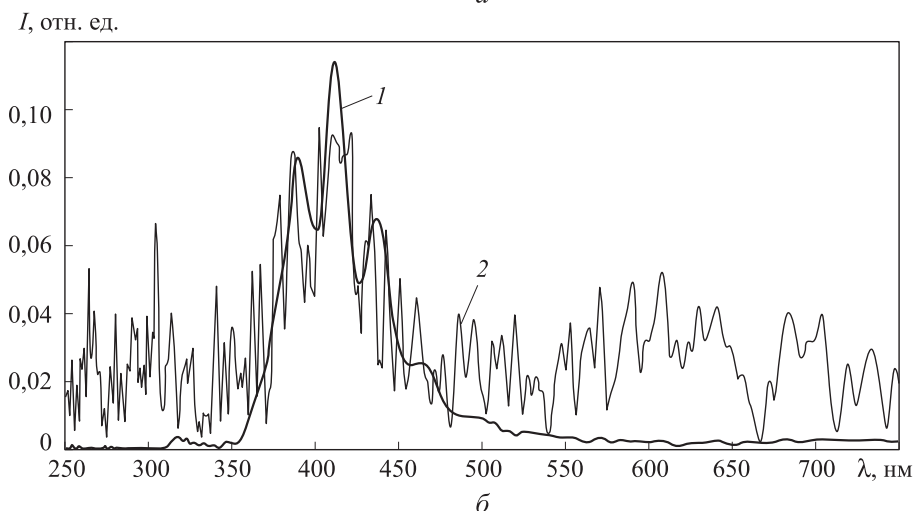
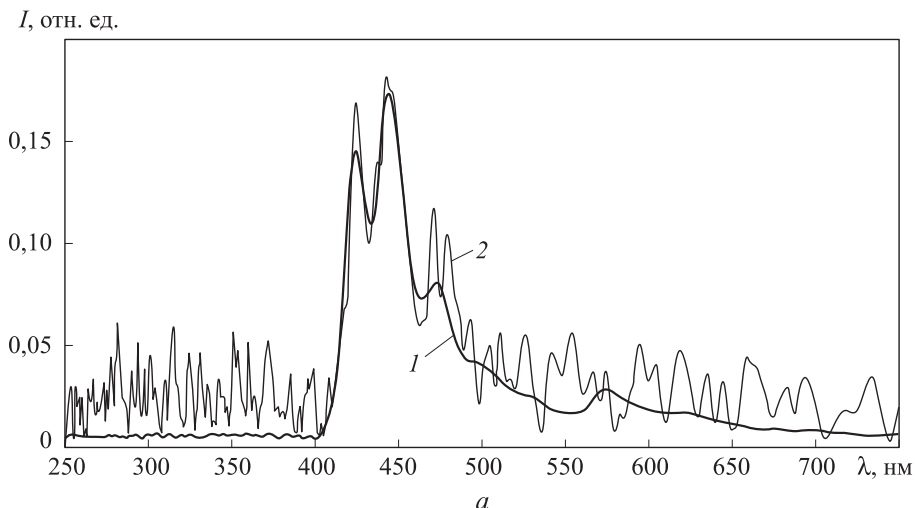


Рис. 1. Эталонный спектр и спектр, имитирующий слабый рассеянный сигнал от образца:

a – тестовое вещество № 1, $\lambda = 280$ нм; *б* – тестовое вещество № 4, $\lambda = 310$ нм

Общий вид такой матрицы (для N эталонных и M измеренных спектров)

$$K_{N \times M} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} & \cdots & \mathbf{K}_{1N} & \mathbf{K}_{1N+1} & \mathbf{K}_{1N+2} & \cdots & \mathbf{K}_{1M} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} & \cdots & \mathbf{K}_{2N} & \mathbf{K}_{2N+1} & \mathbf{K}_{2N+2} & \cdots & \mathbf{K}_{2M} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{K}_{N1} & \mathbf{K}_{N2} & \mathbf{K}_{N3} & \cdots & \mathbf{K}_{NN} & \mathbf{K}_{NN+1} & \mathbf{K}_{NN+2} & \cdots & \mathbf{K}_{NM} \end{pmatrix} \quad (6)$$

Если это максимальный элемент в строке блока N на N и он превышает порог срабатывания, то вещество распознается верно. Жирным шрифтом в матрице отмечены коэффициенты, соответствующие правильному распознаванию. Если порог превышен для недиагонального элемента в блоке N на N , то это приводит к ложным срабатываниям.

Задача идентификации сводится к классификации 10 зашумленных спектров по пяти эталонным. Для оценки качества классификатора строили ROC-кривые [16]. Этот метод применяется в теории обработки сигналов (впервые в США для повышения качества распознавания объектов по радиолокационному сигналу). В данной работе для определения качества распознавания веществ по спектрам использовался ROC-анализ. Входными данными для построения кривой являлись значения в матрице коэффициентов соответствия. Сначала порог срабатывания для нее устанавливался равным единице. При таком значении порога число ложных и верных срабатываний почти всегда равняется нулю. Затем порог уменьшался и в результате появлялись правильные распознавания, а кроме них – ложные. Доли ложных и правильных срабатываний отмечались точкой на соответствующих осях. Если классификатор случайным образом выдает результат, то ROC-кривая такого классификатора будет близка к прямой линии.

Для анализа эффективности методик построены ROC-кривые для одно- и двухдиапазонных алгоритмов при подсветке источниками возбуждающего излучения с длинами волн 280 и 310 нм. Для минимизации использовались функционалы, рассчитанные по формулам (1)–(5). Результаты представлены на рис. 2.

Видно, что при подсветке возбуждающим излучением с длиной волны 280 нм конечный спектр оказывается менее зашумленным, чем при длине волны 310 нм. Это позволяет оценить эффективность алгоритмов при использовании спектров в различных диапазонах длин волн с различным отношением сигнал/шум. Согласно рис. 2, *a*, использование диапазона длин волн 310 нм приводит к менее надежным результатам, площадь под ROC-кривой составила 0,72 против 0,92 для длины волны 280 нм. При использовании совместных алгоритмов площадь под ROC-кривой в большинстве точек оказалась между аналогичной для одного диапазона длиной волны 280 нм и длиной волны 310 нм. Площади под ROC-кривой для алгоритмов, использующих функционалы, рассчитываемые по формулам (3)–(5) составили соответственно 0,87, 0,89, и 0,84.

Таким образом, предложенный алгоритм позволяет проводить идентификацию веществ по набору различных спектров одного и того же вещества. Из предложенных функционалов от невязок в методе

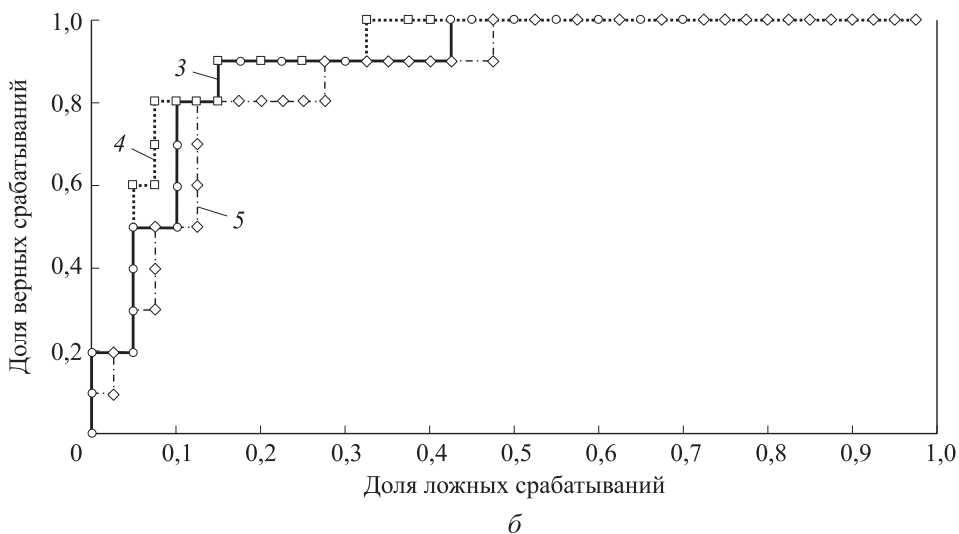
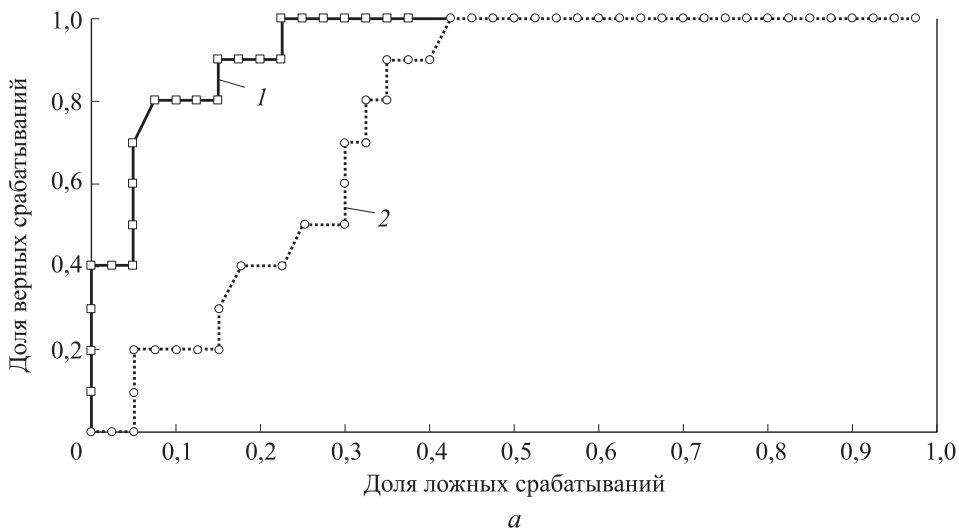


Рис. 2. ROC-кривые при использовании возбуждающего излучения с длинами волн 280 (1) и 310 (2) нм:

а – однодиапазонный алгоритм, минимизация по функционалу (1) для отдельных спектров; б – двухдиапазонный алгоритм, минимизация по функционалам (3–5) для наборов спектров

наименьших квадратов наиболее подходящим оказался функционал (4). ROC-анализ показал, что классификатор, основанный на функционале такого вида оптимально определяет вещества по наборам экспериментальных спектров. Предложенный алгоритм позволяет определять химический состав исследуемых веществ даже в условиях слабого сигнала вторичного излучения, что может быть эффективно для использования в портативных приборах и экспресс-лабораториях мониторинга окружающей среды.

ЛИТЕРАТУРА

- [1] Amme M., Renker B., Schmid B., Feth M.P., Bertagnolli H., Döbelin W. Raman Microspectrometric Identification of Corrosion products Formed on uo2 Nuclear Fuel During Leaching Experiments. *Journal of Nuclear Materials*, 306(2–3): 202–212, 2002.
- [2] Clark R.J.H., Curri M.L. The Identification by Raman Microscopy and x-ray Diffraction of Iron-Oxide Pigments and of the Red Pigments found on Italian Pottery Fragments. *Journal of Molecular Structure*, 440(1–3): 105–111, 1998.
- [3] Hwang J., Choi N., Park A., Park J., Chung J.H., Baek S., Cho S.G., Baek S., Choo J. Fast and Sensitive Recognition of Various Explosive compounds using Raman Spectroscopy and Principal Component Analysis. *Journal of Molecular Structure*, 1039: 130–136, 2013.
- [4] Lipatov E.I., Avdeev S.M., Tarasenko V.F. Photoluminescence and Optical Transmission of Diamond and its Imitators. *Journal of Luminescence*, 130(11): 2106 – 2112, 2010.
- [5] Mendoza Cuevas A., Perez Gravie H. Portable Energy Dispersive x-ray Fluorescence and x-ray Diffraction and Radiography System for Archaeometry. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 633(1): 72 – 78, 2011.
- [6] Kusch P., Obst V., Schroeder-Obst D., Fink W., Knupp G. and Steinhaus J. Application of Pyrolysis – Gas Chromatography/Mass Spectrometry for the Identification of Polymeric Materials in Failure Analysis in the Automotive Industry. *Engineering Failure Analysis*, 2013, issn: 1350–6307, doi: <http://dx.doi.org/10.1016/j.engfailanal.2012.12.005>.
- [7] Ferarya S., Augera J., Toucheb A. Trace Identification Of Plant Substances by Combining Gas Chromatography-mass Spectrometry and Direct Deposition Gas Chromatography-Fourier Transform Infrared Spectrometry. *Talanta*, 43(3): 349–357, 1996.
- [8] Muhammad A. H., Guenther S., Stefan M., Guenther K.B. Identification of Amino Acids by Material Enhanced Laser Desorption/Ionisation Mass Spectrometry (Meldi-ms) in Positive- and Negative-ion Mode. *International Journal of Mass Spectrometry*, 279(1): 15–24, 2009.
- [9] Cai Sheng Wua, Zhi Xin Jiaa, Bao Ming Ningc, Jin Lan Zhanga, Song Wub. Separation and Identification Of Moxifloxacin Impurities in Drug Substance by High-Performance Liquid Chromatography Coupled with Ultraviolet Detection and Fourier Transform Ion Cyclotron Resonance mass Spectrometry. *Chinese Chemical Letters*, 23(10): 1185–1188, 2012.
- [10] Gorchakovskii S.N., Ivlev O.A., Kochikov I.V., Maznichko A.A., Morozov A.N., Palatov Yu.A., Svetlichnyi S.I., Tabalin S.E. Compact Fourier Spectrometer for Remote Analysis of gaseous Media. *Journal of Optical Technology (A Translation of Opticheskii Zhurnal)*, 65(6): 510–512, 1998.
- [11] Бойко А.Ю., Григорьев А.А., Дворук С.К., Корниенко В.Н., Кочиков И.В., Лельков М.В., Мацюк Г.В., Морозов А.Н., Павлов А.Ю., Светличный С.И., Табалин С.Е., Шишкин Г.В., Шлыгин П.Е. Проблема идентификации и определения концентраций загрязняющих веществ с помощью фурье-спектрорадиометра. *Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки*, (1): 26, 2004.
- [12] Кочиков И.В., Морозов А.Н., Светличный С.И., Фуфурин И.Л. Распознавание веществ в открытой атмосфере по единичной интерферограмме фурье-спектрорадиометра. *Оптика и спектроскопия*, 106(5): 743–749, 2009.

- [13] Kochikov I.V., Morozov A.N., Svetlichnyi S.I., Fufurin I.L. Substance Recognition in the Open Atmosphere from a single Fourier Transform Spectroradiometer Interferogram. *Optics and Spectroscopy (English translation of Optika i Spektroskopiya)*, 106(5): 666–671, 2009.
- [14] Бойко А.Ю., Голяк И.С., Голяк И.С., Дворук С.К., Доровских А.М., Есаков А.А., Корниенко В.Н., Косенко Д.В., Кочиков И.В., Морозов А.Н., Светличный С.И., Табалин С.Е. Статический фурье-спектрометр видимого и ближнего ультрафиолетового диапазонов спектра. *Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки*, (3): 10–28, 2009.
- [15] Глаголев К.В., Голяк И.С., Голяк И.С., Есаков А.А., Корниенко В.Н., Кочиков И.В., Морозов А.Н., Светличный С.И., Табалин С.Е. Методика получения и обработки спектральной информации с помощью статического фурье-спектрометра. *Оптика и спектроскопия*, 110(3): 486–492, 2011.
- [16] Tan N.C., Fisher W.G., Rosenblatt K.P., Garner H.R. Application of Multiple Statistical Tests to Enhance Mass Spectrometrybased Biomarker Discovery. *BMC Bioinformatics*, 10, 2009.

Статья поступила в редакцию 05.06.2013

Ссылку на эту статью просим оформлять следующим образом:

Васильев Н.С., Морозов А.Н. Алгоритм идентификации веществ по набору спектров вторичного излучения. *Инженерный журнал: наука и инновации*, 2013, вып. 8. URL: <http://engjournal.ru/catalog/fundamentals/physics/1109.html>

Васильев Николай Сергеевич – ассистент кафедры «Физика» МГТУ им. Н.Э. Баумана. Область научных интересов: спектроскопия. e-mail: nickliamg@gmail.com

Морозов Андрей Николаевич родился в 1959 г., окончил МВТУ им. Н.Э. Баумана в 1982 г. Д-р физ.-мат. наук, профессор, заведующий кафедрой «Физика» МГТУ им. Н.Э. Баумана. Автор более 200 научных работ в области прецизионных измерений и физической кинетики. e-mail: amor@mx.bmstu.ru