



УДК 535.37/535.31

## ПРИМЕНЕНИЕ МЕТОДОВ ХЕМОМЕТРИКИ ДЛЯ АНАЛИЗА БИОАЭРОЗОЛЕЙ ПРОТОЧНО-ОПТИЧЕСКИМ МЕТОДОМ

Е.С. Худяков<sup>a</sup>, Е.А. Кочелаев<sup>a</sup>, А.О. Волчек<sup>a</sup>, Д.О. Кирсанов<sup>b,c</sup>, И.Э. Джагапанян<sup>b</sup><sup>a</sup> АО «НПО «Прибор», Санкт-Петербург, 199034, Российская Федерация<sup>b</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация<sup>c</sup> Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация

Адрес для переписки: drjie@mail.ru

### Информация о статье

Поступила в редакцию 21.09.15, принята к печати 01.12.15

doi:10.17586/2226-1494-2016-16-1-30-38

Язык статьи – русский

**Ссылка для цитирования:** Худяков Е.С., Кочелаев Е.А., Волчек А.О., Кирсанов Д.О., Джагапанян И.Э. Применение методов хемометрики для анализа биоаэрозолей проточно-оптическим методом // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 1. С. 30–38.

### Аннотация

**Предмет исследования.** Исследована информативность измерительных каналов анализатора биоаэрозолей. Работа анализатора основана на проточно-оптическом методе. **Метод.** Проведены измерения сигналов флуоресценции и упругого рассеяния отдельных частиц биоаэрозолей в пяти и двух спектральных диапазонах соответственно. Выполнена регистрация сигналов частиц почвенной пыли, которая использовалась в качестве имитатора фоновых атмосферных веществ. Для возбуждения флуоресценции частиц аэрозолей использованы источники света – лазерный на длине волны 266 нм и светодиодный на длине волны 365 нм. **Основные результаты.** На основе математического аппарата хемометрики произведена классификация признаков по степени информативности и выбраны три наиболее существенных признака, которые в сумме описывают 72% общей дисперсии данных. Тестирование произведено с использованием алгоритмов SIMCA и метода ближайших соседей. Показано, что использование исходного и уменьшенного набора из трех признаков позволяет получать сравнимую точность при классификации биоаэрозолей. **Практическая значимость.** Показана возможность быстрого обнаружения и идентификации частиц биоаэрозоля респираторной (задерживаемой в органах дыхания человека) фракции 1–10 мкм проточно-оптическим методом на фоне частиц небиологической природы. Для разработки компактного и недорогого анализатора выбраны наиболее информативные оптические спектральные диапазоны регистрации.

### Ключевые слова

проточно-оптический метод, биоаэрозоли, метод главных компонент, хемометрика, SIMCA, метод ближайших соседей, информативные признаки

### Благодарности

Работа частично выполнена при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

## APPLICATION OF CHEMOMETRICS FOR ANALYSIS OF BIOAEROSOLS BY FLOW-OPTICAL METHOD

E.S. Khudyakov<sup>a</sup>, E.A. Kochelaev<sup>a</sup>, A.O. Volchek<sup>a</sup>, D.O. Kirsanov<sup>b,c</sup>, I.E. Jahatspanian<sup>b</sup><sup>a</sup> JSC "NPO "Pribor", Saint Petersburg, 199034, Russian Federation<sup>b</sup> ITMO University, Saint Petersburg, 197101, Russian Federation<sup>c</sup> Saint Petersburg State University, Saint Petersburg, 199034, Russian Federation

Corresponding author: drjie@mail.ru

### Article info

Received 21.09.15, accepted 01.12.15

doi:10.17586/2226-1494-2016-16-1-30-38

Article in Russian

**For citation:** Khudyakov E.S., Kochelaev E.A., Volchek A.O., Kirsanov D.O., Jahatspanian I.E. Application of chemometrics for analysis of bioaerosols by flow-optical method. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 1, pp. 30–38.

### Abstract

**Subject of Research.** The informativity of detection channels for bioaerosol analyzer is investigated. Analyzer operation is based on flow-optical method. **Method.** Measurements of fluorescence and the light scattering of separate bioaerosol

particles were performed in five and two spectral ranges, correspondingly. The signals of soil dust particles were registered and used as an imitation of background atmospheric particles. For fluorescence induction of bioaerosol particles we used light sources: a laser one with a wavelength equal to 266 nm and 365 nm LED source. **Main Results.** Using chemometric data processing the classification of informative parameters has been performed and three most significant parameters have been chosen which account for 72% of total data variance. Testing has been done using SIMCA and k-NN methods. It has been proved that the use of the original and the reduced sets of three parameters produces comparable accuracy for classification of bioaerosols. **Practical Relevance.** The possibility of rapid detection and identification of bioaerosol particles of 1-10 microns respirable fraction (hindering in the human respiratory system) by flow-optical method on a background of non-biological particles is demonstrated. The most informative optical spectral ranges for development of compact and inexpensive analyzer are chosen.

#### Keywords

flow-optical method, bioaerosols, principal component analysis, chemometric, SIMCA, k-NN, informative characteristics

#### Acknowledgements

This work was partially financially supported by the Government of the Russian Federation (Grant 074-U01).

### Введение

Биологическое загрязнение воздуха может произойти в результате утечек и аварий на предприятиях фармакологического профиля, эпидемий, стихийных бедствий, а также при применении биологического оружия. При этом аэрозольный путь распространения биологических агентов в настоящее время принято считать наиболее вероятным, а патогенные бактерии, входящие в состав биоаэрозоля, могут вызывать такие инфекционные заболевания, как бруцеллез, сибирская язва, сальмонеллез, столбняк, туляремия и др. [1].

Эффективность противодействия последствиям биологического загрязнения воздуха во многом определяется скоростью установления факта присутствия патогенных частиц в составе атмосферы, поэтому определение факта биологической угрозы должно происходить в условиях реального времени. Дополнительная информация об отнесении биологического агента к какому-либо классу, например, спорам, бактериям или белкам [2], способствует повышению эффективности мер медицинской помощи пострадавшим.

Существующие в настоящее время методики и технологии прямого микробиологического контроля, основанные на пробоотборе из воздуха микробных клеток с последующим осаждением на поверхность плотной питательной среды, культивирования и подсчета выросших колоний, имеют такие минусы, как длительность, громоздкость, трудозатратность. Помимо этого, в процессе отбора пробы воздуха и посева на питательную среду клетки микроорганизмов могут травмироваться, что приводит к искажению результатов [3–5]. В связи с этим актуальна задача разработки высокочувствительных экспресс-методов обнаружения патогенных биоаэрозолей в общей массе фоновых атмосферных аэрозолей как биологической, так и небιологической природы<sup>1</sup>.

Для быстрого обнаружения и идентификации патогенных биологических частиц респираторной (задерживаемой в органах дыхания человека) фракции 1–10 мкм может быть применен проточно-оптический метод (ПОМ) анализа, использующий измерения флуоресценции и рассеяния света от отдельных частиц аэрозоля [6–8]. В ПОМ отбор пробы и освещение частиц происходит в потоке воздуха, что позволяет анализировать до  $10^4$  отдельных частиц аэрозоля в секунду и выдавать результат анализа за время, не превышающее нескольких минут.

В общем случае в различных приборных реализациях ПОМ разделение биологического и небιологического аэрозоля, как правило, происходит на основе анализа флуоресценции триптофана, присутствующего в частицах биоаэрозоля в составе белков и остатков аминокислот и возбуждаемого в области длин волн 250–290 нм [6, 7]. Кроме того, для биоаэрозолей, относящихся к различным таксономическим группам [2], проявляются различия в спектральных характеристиках: форме, положению максимума, величине флуоресценции [9–11], что позволяет проводить их идентификацию [12].

При анализе ПОМ наличие в окружающем воздухе определенного уровня естественного (не патогенного) биологического фона, частицы которого также содержат триптофан, может приводить к ложноположительным результатам обнаружения патогенных биоаэрозолей. Задача осложняется тем, что биоаэрозольный фон атмосферы – динамичная система, флуоресцентные свойства которой могут меняться в естественных условиях при воздействии солнечного света, температуры, влажности и других факторов [8]. Таким образом, задача правильной идентификации патогенных частиц в составе аэрозоля с помощью ПОМ требует повышения селективности метода. В существующих устройствах ПОМ для повышения селективности включают в оптическую систему дополнительные источники излучения и (или) фотодетекторы [2, 7, 8]. Например, применяют дополнительный источник возбуждения в области длин волн 330–380 нм, что позволяет, помимо спектра флуоресценции триптофана, регистрировать спектры флуо-

<sup>1</sup> Ответные меры системы общественного здравоохранения на угрозу применения биологического и химического оружия. Руководство ВОЗ. 2001. <http://www.who.int/mc/pdfs/PublicHealthR.pdf>

ресценции NADH и флавинов в границах спектрального диапазона 400–600 нм [7]. Еще одним способом повышения селективности ПОМ является детектирование рассеянного света. Нормировка сигналов флуоресценции на сигнал упругого рассеяния позволяет значительно сократить разброс регистрируемых данных, связанный с нестабильностью источников излучения, неравномерностью распределения интенсивности излучения в области анализа, а также уменьшить разброс данных, связанный с широким распределением частиц аэрозоля по размерам [7]. Вместе с тем увеличение числа оптических элементов (источников и детекторов света) приводит к усложнению конструкции прибора, росту энергопотребления и массогабаритных характеристик, к значительному увеличению стоимости, усложняет его наладку и ухудшает надежность. Это вступает в противоречие с требованиями, предъявляемыми к «полевым» устройствам ПОМ – компактность, надежность, малое энергопотребление.

Целью настоящей работы является оценка информативности различных измерительных каналов ПОМ на основе данных по измерениям сигналов флуоресценции и упругого рассеяния от отдельных частиц биоаэрозолей и выбор признаков (информативные признаки), которые описывают наиболее существенную информацию при разделении четырех классов аэрозолей.

Для решения данной задачи были получены экспериментальные данные на проточно-оптическом анализаторе биоаэрозолей, описанном ранее в [13]. Анализ данных выполнен с применением математического аппарата хемометрики. Данный аппарат обеспечивает эффективное и корректное использование методов современной математической статистики для извлечения полезной информации из многомерных данных и получил широкое распространение для решения задач количественного анализа результатов инструментальных измерений (спектроскопических, электрохимических и т.д.) [14, 15].

### Описание эксперимента

Для создания аэрозоля использовалась аэрозольная камера (рис. 1), рабочий объем которой составляет  $0,7 \text{ м}^3$ . В камере подача частиц внутрь бокса осуществлялась с использованием вакуумного насоса (эжектора). Порошки проб веществ по отдельности наносились на фильтр АФА ВП-10 и при создании давления около 2 атм воздушным насосом 1 на входе эжектора уносились возникшим потоком с поверхности фильтра в бокс аэрозольной камеры. В процессе измерений воздух бокса перемешивался вентиляторами, что обеспечивало однородность дисперсного состава создаваемого аэрозольного облака по всей высоте бокса.

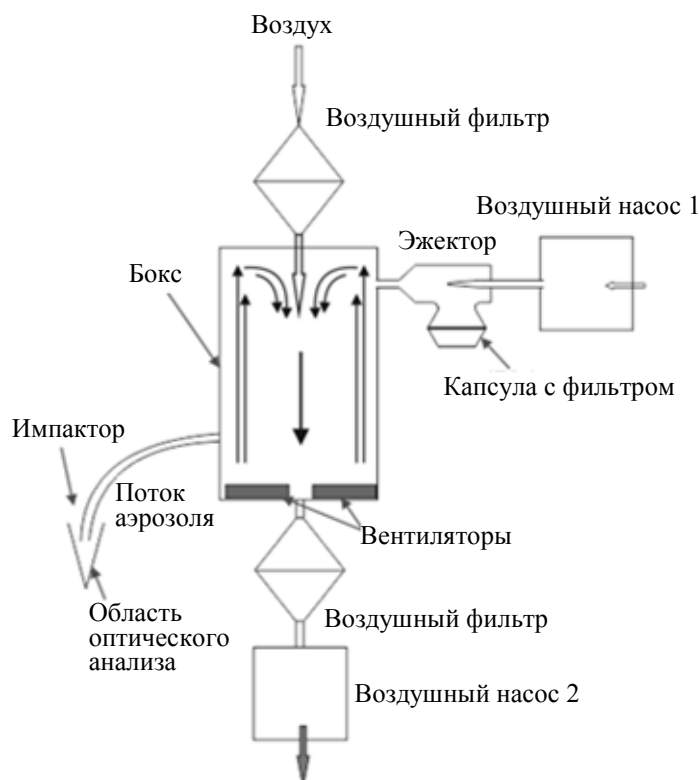


Рис. 1. Блок-схема аэрозольной камеры

После создания аэрозольного облака включался анализатор аэрозолей, воздушный вход которого соединялся с боксом аэрозольной камеры. После включения поток воздуха из бокса камеры поступал на вход виртуального импактора [16] анализатора.

При работе анализатора виртуальный импактор обеспечивает удаление из анализа частиц размером менее 1 мкм и формирует поток воздуха, содержащий частицы с размерами более 1 мкм. На выходе импактора поток аэрозольных частиц с помощью аэродинамической фокусировки [17] сжимается до диаметра 0,2 мм и поступает в область оптического анализа.

Оптическая схема (рис. 2), подробно описанная в [13], обеспечивает измерение сигналов флуоресценции и упруго рассеянного света частицами аэрозолей. Для возбуждения флуоресценции частиц аэрозоля взаимно ортогональные световые потоки ультрафиолетового (УФ) излучения лазера и светодиода фокусируются в точке О (область анализа), через которую проходит воздушный поток частиц аэрозоля. Ось воздушного потока (на рисунке не показана) ортогональна плоскости световых потоков. Пройдя через область анализа, излучение от лазера и от светодиода попадает на оптические ловушки и практически полностью поглощается. В качестве детекторов использованы фотоэлектронные умножители (ФЭУ).

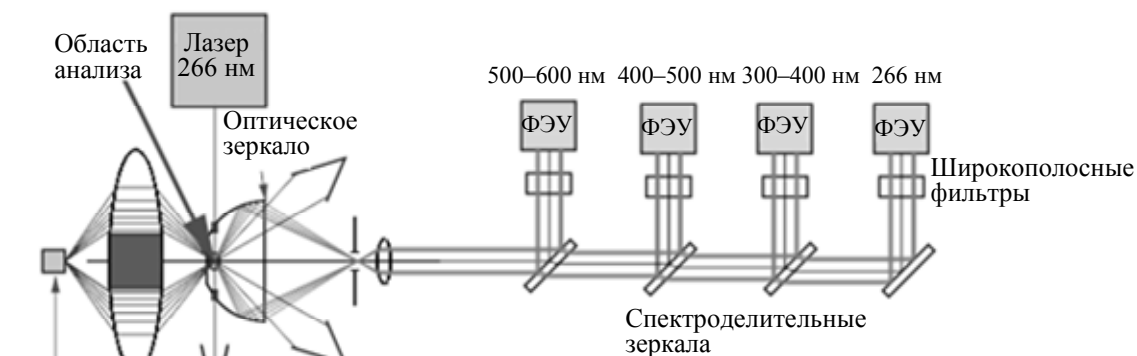


Рис. 2. Оптическая схема измерительной системы

В качестве источников излучения в анализаторе использованы:

1. опытный образец импульсно-периодического лазера с длиной волны  $\lambda=266$  нм, частотой импульсов  $n=22$  кГц и средней мощностью излучения 4 мВт. Плотность энергии возбуждающего излучения в пятне засветки 70–150 мкДж/см<sup>2</sup>;
2. светодиод NSU-033A (Nichia corporation) с длиной волны излучения  $\lambda=365$  нм.

Скорость движения частицы аэрозоля в области анализа  $v$ , частота повторения импульса  $n$  и размер области анализа вдоль струи аэрозоля  $L$  согласованы и удовлетворяют соотношению  $n=v/L$ . Выполнение данного условия обеспечивает облучение импульсом лазера каждой частицы, поступающей в анализируемый объем [6]. Импульс светодиода запускается после импульса лазера только при условии, если пришедший от частицы сигнал рассеяния лазерного луча превышает величину порогового значения – минимального уровня сигнала, соответствующего сигналу от частицы размером 1 мкм.

При освещении частицы аэрозоля УФ излучением она переизлучает в виде флуоресценции часть светового потока, падающего на ее поверхность, а также частично упруго рассеивает его. Так как центр освещенной зоны совпадает с передним фокусом объектива регистрации – эллиптического зеркала, то световые потоки флуоресценции и рассеяния отражаются зеркалом (телесный угол светосбора  $\Omega = 2\pi$  стер) в область второго фокуса и далее направляются системой линз и спектроделительных зеркал на приемные площадки ФЭУ. Комбинация характеристик отражения и пропускания спектроделительных зеркал и установленных перед ФЭУ широкополосных оптических светофильтров обеспечивает разделение излучения в четырех каналах регистрации по четырем спектральным диапазонам 260–290, 300–400, 400–500 и 500–600 нм.

В качестве фотодетекторов использованы ФЭУ-175. Разделение во времени моментов освещения частицы импульсами лазера и светодиода позволяет использовать одни и те же каналы ФЭУ для регистрации сигналов флуоресценции, возбуждаемых обоими источниками.

В процессе измерений от каждой отдельной частицы аэрозоля регистрируются сигналы флуоресценции в диапазонах длин волн 300–400 нм ( $F_{11}$ ), 400–500 нм ( $F_{21}$ ) и 500–600 нм ( $F_{31}$ ), возбуждаемые импульсом лазера; сигналы флуоресценции в диапазонах длин волн 400–500 нм ( $F_{22}$ ) и 500–600 нм ( $F_{32}$ ), возбуждаемые импульсом светодиода, сигналы упругого рассеяния на частицах аэрозоля световых потоков лазера и светодиода – 266 нм ( $S_1$ ) и 300–400 нм ( $S_2$ ) соответственно. В приведенных обозначениях подстрочный индекс 1 соответствует освещению частиц УФ лазером, 2 – светодиодом. Всего измеряются 7 независимых показателей:  $F_{11}$ ,  $F_{21}$ ,  $F_{31}$ ,  $S_1$ ,  $F_{22}$ ,  $F_{32}$  и  $S_2$ .

В качестве образцов проб биоаэрозолей в работе использовались:

- порошки двух типов сухого яичного белка (Ovalbumin), отличающиеся методами пробоподготовки, аналогичные использованным ранее в [6];

– споры бактерий *Bacillus subtilis*.

Спектральные характеристики интенсивности флуоресценции двух типов биоаэрозолей на основе яичного белка были измерены на люминесцентном спектрометре LS-50B (Perkin Elmer, США).

Спектральные характеристики спор *Bacillus subtilis* по форме подобны кривым на рис. 3 и приведены в работе [12].

В качестве имитатора, показывающего влияние фоновых атмосферных веществ на измерения и классификацию биоаэрозолей, использовалась естественная примесь – почвенная пыль. Для этого отбирались образцы грунта с поверхности земли, из которых просеиванием удалялись крупные частицы. Затем образцы высушивали при температуре выше 100 °С.

Таким образом, в аэрозольной камере создавались порошки для трех типов биоаэрозолей и порошок почвенной пыли. Аэрозоли этих порошков составили 4 класса для исследования ПОМ. Количество отдельных зарегистрированных частиц аэрозоля за время создания аэрозольного облака каждого класса составило не менее 500, при этом в каждом измерении от отдельной частицы регистрировали 7 сигналов.

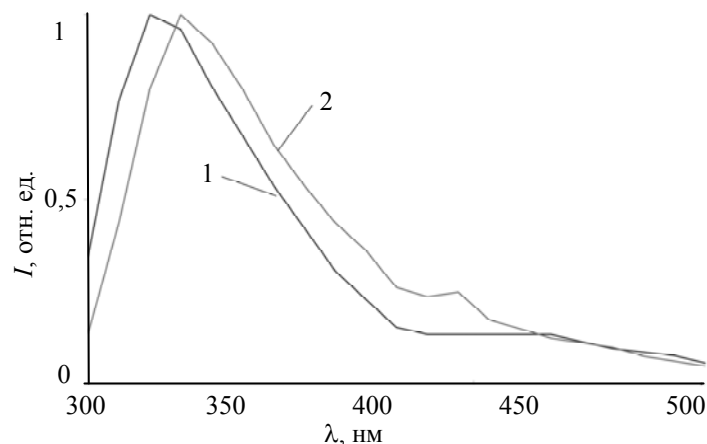


Рис. 3. Интенсивность флуоресценции 2 типов белков *Ovalbumin* в зависимости от длины волны: 1 – первый тип; 2 – второй тип

#### Обработка экспериментальных данных

Для многомерного статистического анализа данных использовался метод главных компонент, (МГК) [14, 15].

Основная идея МГК заключается в снижении размерности исходного набора многомерных данных за счет построения линейной комбинации переменных, наилучшим образом объясняющей их разброс. При этом новые координаты находятся путем построения проекций точек в пространстве исходных данных на оси новой системы координат – системы главных компонент (ГК). Направление оси первой ГК соответствует направлению наибольшего разброса данных. Каждая последующая ось ГК ортогональна предыдущим и описывает максимальный разброс данных в пространстве ортогональном предыдущим ГК. Положение осей новой системы координат определяется направляющими косинусами, или нагрузками (в терминологии МГК). В силу независимости новых переменных, измеряемых вдоль осей ГК, общая дисперсия данных  $D_0$  может быть представлена как

$$D_0 = \sum_{i=1}^S D_i + O(D) = \sum_{i=1}^S D_i \sum_{j=1}^N \alpha_{ij}^2 + O(D) \approx \sum_{j=1}^N \sum_{i=1}^S \alpha_{ij}^2 D_i = \sum_{j=1}^N R_j,$$

где  $D_i$  – дисперсия вдоль  $i$ -ой главной компоненты;  $S$  – число значимых главных компонент, сумма дисперсий которых близка к  $D_0$ ,  $O(D)$  – остаточная дисперсия, описываемая ГК с номерами выше  $S$  ( $O(D) \ll D_0$ ),  $N$  – число исходных переменных ( $N=5$ ),  $\alpha_{ij}$  – направляющий косинус между осями  $i$ -ой ГК и  $j$ -ой переменной.

Коэффициенты  $R_j = \sum_{i=1}^S \alpha_{ij}^2 D_i$  определяют вклад  $j$ -ой переменной в общую дисперсию данных  $D_0$ .

В настоящей работе значимость исходных переменных определялась на основе вычислений коэффициентов  $R_j/D_0$ , описывающих вклад отдельных переменных в общую объясненную дисперсию.

Обработка исходных данных происходила в несколько этапов:

- построение графиков счетов и нагрузок и определение числа  $S$  существенных ГК;
- определение значимых переменных на основе величины коэффициентов  $R_j/D_0$ ;
- тестирование и сравнение способности классификации аэрозолей по исходному и уменьшенному набору признаков.

Построение графиков счетов исходных данных (рис. 4, а) показало большой разброс показаний внутри каждого класса аэрозолей и отсутствие выраженных различий между классами. Такой «шум» данных может быть связан с большим разбросом размеров частиц аэрозолей, временной нестабильностью источников излучения, неоднородностью пятна засветки. Для уменьшения разброса показаний дальнейшие расчеты проводились после нормировки сигналов флюоресценции на сигнал рассеяния. Таким образом, количество обрабатываемых признаков было уменьшено до 5: первый –  $F1_1/S_1$ , второй –  $F2_1/S_1$ , третий –  $F3_1/S_1$ , четвертый –  $F2_2/S_2$ , пятый –  $F3_2/S_2$ .

На рис. 4, б, в, показаны двумерные графики счетов по первым трем главным компонентам для 5 признаков и 4 классов. Так как ГК 1, 2 и 3 описывают 58, 21 и 10% общей объясненной дисперсии, было принято, что число значимых главных компонент  $S=3$ . Как видно из рис. 4, после нормировки разброс показаний внутри каждого класса значительно уменьшился, появилось визуально различимое разделение по классам.

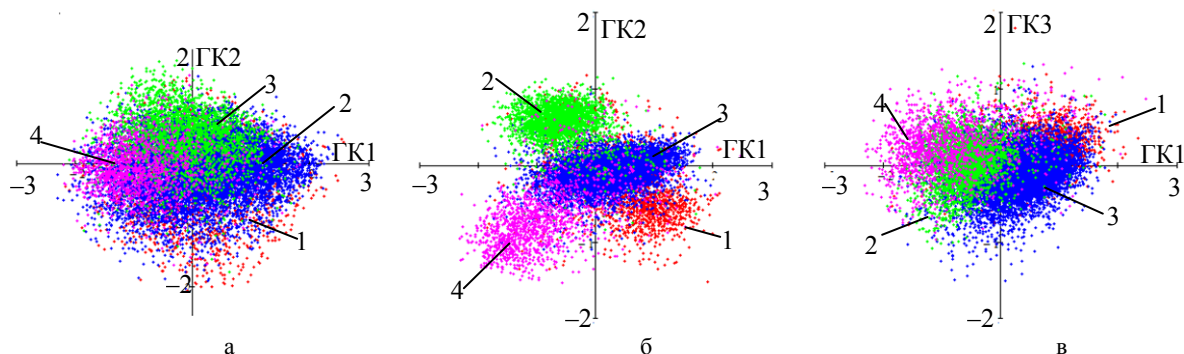


Рис. 4. Графики счетов для 7 (а) и 5 (б, в) признаков по первым главным компонентам для аэрозолей: 1 – спор *Bacillus subtilis*; 2 – *Ovalbumin* тип 1; 3 – *Ovalbumin* тип 2; 4 – пыль

Для пяти новых переменных были определены коэффициенты  $R_j/D_0$  (табл. 1). Как следует из табл. 1, вклад признаков 1, 3 и 4 в общую объясненную дисперсию составляет 72%, что существенно превышает вклад признаков 2 и 5 (28%). При этом доля дисперсии, объясненной признаками 1, 3 и 4, оказывается примерно одинаковой. Следовательно, признаки 1, 3 и 4 приняты как наиболее важные. График нагрузок измеренных признаков приведен на рис. 5.

$j$	1	2	3	4	5
$R_j/D_0$	0,26	0,12	0,20	0,26	0,16

Таблица 1. Коэффициенты  $R_j/D_0$  для пяти новых переменных

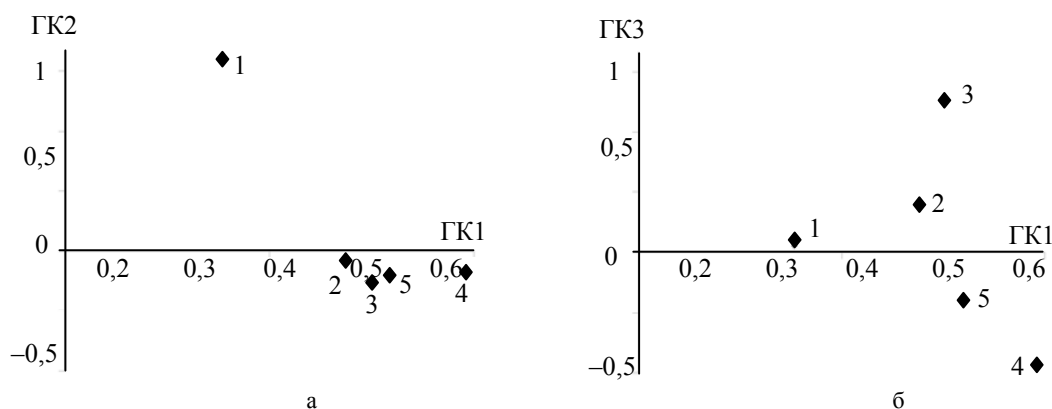


Рис. 5. Графики нагрузок по первым трем главным компонентам по пяти признакам для 4 классов аэрозолей (а–б)

Тестирование и сравнение возможности классификации отдельных частиц аэрозолей по исходному и уменьшенному набору признаков происходило с применением методов SIMCA (Soft independent modeling of class analogy) [15, 18] и метода ближайших соседей ( $k$ -Nearest Neighbors,  $k$ -NN) [19].

Для устройств ПОМ приборного исполнения принятие решения об обнаружении в составе воздуха облака аэрозоля включает два этапа:

- определение класса каждой отдельной зарегистрированной частицы;
- обнаружение аэрозоля определенного класса, для которого частота поступления частиц превышает заданный уровень порогового значения.

Точность принятия решения при классификации отдельных частиц во многом определяется тем, используется ли в алгоритме классификации часть информации или все известные данные о функции распределения характеристик аэрозолей в пространстве признаков. Преимуществом алгоритмов первого типа является высокое быстродействие и возможность реализации «в полевых» условиях; алгоритмов второго типа – более высокая точность анализа. Однако вследствие высоких требований к вычислительным ресурсам реализация последних алгоритмов затруднительна в приборном исполнении устройств ПОМ. Метод SIMCA является примером алгоритмов, использующих часть информации о функции распределения обучающей выборки, а метод  $k$ -NN использует всю информацию.

В SIMCA на этапе обучения каждый класс моделируется обособленно, независимо от остальных. В качестве критерия для классификации новых образцов используется модельное расстояние  $f_{crit}$ , определяемое расстоянием до центра класса и квантилем  $F$ -статистики Фишера. При этом каждая частица может быть одновременно отнесена к нескольким классам [15, 18].

В методе  $k$ -ближайших соседей при классификации рассчитывается евклидово расстояние до всех образцов обучающего набора и выделяется  $k$ -ближайших соседей, расстояние до которых минимально. Параметр  $k$  подбирается эмпирическим путем. Увеличение  $k$  приводит к уменьшению влияния погрешностей, но ухудшает разделение на классы. Объект присваивается тому классу, который является наиболее распространенным среди его  $k$ -соседей [19].

Для оценки качества классификации в обоих алгоритмах данные разделялись на обучающую (90%) и контрольную (10%) выборки. Результаты классификации по методу SIMCA представлены в табл. 2.

Классификация методом ближайших соседей была проведена со значениями  $k = 3-10$ . Лучшие результаты были получены с  $k = 5$  (табл. 3).

Номер класса, из которого были отобраны тестируемые данные	Процент идентификации по классам по методу SIMCA								
	по пяти признакам				по трем признакам				
	1	2	3	4	1	2	3	4	
1	биоаэрозоли	<b>95</b>	73	34	70	<b>95</b>	88	36	55
2		82	<b>95</b>	24	17	80	<b>96</b>	12	14
3		35	21	<b>96</b>	4	19	15	<b>96</b>	2
4	пыль	44	54	13	<b>94</b>	30	63	11	<b>95</b>

Таблица 2. Тестирование модели МГК по пяти и трем признакам по методу SIMCA

Номер класса, из которого были отобраны тестируемые данные	Процент идентификации по классам по методу ближайших соседей ( $k$ -NN)								
	по пяти признакам				по трем признакам				
	1	2	3	4	1	2	3	4	
1	биоаэрозоли	<b>53</b>	30	10	7	<b>48</b>	30	12	10
2		2	<b>95</b>	2	1	2	<b>95</b>	1	2
3		1	5	<b>94</b>	0	1	4	<b>95</b>	0
4	пыль	5	9	4	<b>82</b>	6	11	5	<b>78</b>

Таблица 3. Проверка модели по пяти и трем признакам по методу ближайших соседей ( $k$ -NN)

Из данных табл. 2, 3 следует:

- для исходного набора из пяти признаков, также как и для уменьшенного (из трех), большинство определяемых частиц соответствуют правильному классу;
- использование всей накопленной статистики измерений позволяет проводить правильную классификацию обнаруживаемого биоаэрозоля, если в качестве критерия использовать максимальную частоту появления биоаэрозолей выше порогового значения. При этом допускается, что одновременное появление двух облаков патогенных биоаэрозолей маловероятно;
- ухудшение точности измерений при использовании уменьшенного числа признаков является незначительным и может считаться приемлемым для устройств ПОМ «полевого» исполнения.

Для приборного исполнения аппаратуры ПОМ практическим результатом проведенного анализа явилось бы уменьшение числа используемых в устройстве оптических приемников с одновременным использованием «быстрого» алгоритма типа SIMCA. При использовании одного, трех и четырех признаков уменьшить число приемников невозможно, так как для измерения трех данных признаков требуется регистрация излучения во всех исходных спектральных диапазонах всеми четырьмя приемниками (рис. 2). Однако более подробный анализ нагрузок показал, что вклад 2-го признака ( $F_{21}/S_1$ ) в дисперсию вдоль первой и второй главной компонент незначительно отличается от вклада третьего признака ( $F_{31}/S_1$ ), что позволяет ожидать незначительного снижения избирательности при замене третьего признака на второй. Результаты тестирования по методу SIMCA (табл. 4), на основе трех признаков ( $F_{11}/S_1$ ,  $F_{21}/S_1$ ,  $F_{22}/S_2$ ) подтверждают это. Следовательно, в устройстве ПОМ «полевого» исполнения целесообразно использо-

вать три оптических приемника с диапазонами регистрации 266 нм ( $S_1$ ), 300–400 нм ( $F_{11}$ ,  $S_2$ ), 400–500 нм ( $F_{21}$ ,  $F_{22}$ ).

Номер класса, из которого были отобраны тестируемые данные		Процент идентификации по классам по методу SIMCA			
		1	2	3	4
1	биоаэрозоли	<b>94</b>	88	36	55
2		81	<b>95</b>	12	14
3		19	15	<b>97</b>	2
4	пыль	30	83	11	<b>95</b>

Таблица 4. Проверка модели по трем признакам:  $F_{11}/S_1$ ,  $F_{21}/S_1$ ,  $F_{22}/S_2$  по методу SIMCA

#### Заключение

На основе математического аппарата хемометрики выбраны информативные признаки, которые описывают наиболее существенную информацию при разделении четырех классов аэрозолей. При этом исходное число из пяти нормированных признаков уменьшено до трех, которые в сумме описывают 72% общей дисперсии данных:  $F_{11}/S_1$ ,  $F_{31}/S_1$ ,  $F_{22}/S_2$ .

Тестирование алгоритма классификации по исходному и уменьшенному набору признаков показало сравнимую точность при классификации биоаэрозолей, что позволяет использовать уменьшенный набор параметров в измерениях проточно-оптическим методом без существенной потери избирательности. Результаты анализа показывают также, что число используемых приемников может быть уменьшено с четырех до трех без заметного снижения избирательности.

Использованный подход при отборе «важных» признаков может быть применен при разработке анализаторов проточно-оптическим методом. При определении таких признаков является существенным использование максимально полной базы данных биоаэрозолей, подлежащих обнаружению и идентификации.

#### References

1. Dutkiewicz J. Bacteria and fungi in organic dust as potential health hazard. *Annals of Agriculture and Environmental Medicine*, 1997, vol. 4, no. 1, pp. 11–16.
2. Sivaprakasam V., Huston A.L., Scotto C., Eversole J.D. Multiple UV wavelength excitation and fluorescence of bioaerosols. *Optics Express*, 2004, vol. 12, no. 19, pp. 4457–4466. doi: 10.1364/OPEX.12.004457
3. Pepper I.L., Gerba C.P., Gentry T., Raina M. *Environmental Microbiology*. 2<sup>nd</sup> ed. San Diego, Academic Press, 2011, 624 p. doi: 10.1016/B978-0-12-370519-8.50002-9
4. Kikikawa M., Iwasaka Y., Kobayashi F., Maki T. Molecular identification of microorganisms in bioaerosols. *Earozoru Kenkyu*, 2010, vol. 25, pp. 29–34.
5. Usachev E.V., Pankova A.V., Rafailova E.A., Pyankov O.V., Agranovski I.E. Portable automatic bioaerosol sampling system for rapid on-site detection of targeted airborne microorganisms. *Journal of Environmental Monitoring*, 2012, vol. 14, no. 10, pp. 2739–2745. doi: 10.1039/c2em30317e
6. Kochelaev E.A., Volchek A.O. Optical recording system for a flow-through optical method of analyzing bioaerosols. *Journal of Optical Technology*, 2011, vol. 78, no. 6, pp. 365–370.
7. Jeys T.H., Herzog W.D., Hybl J.D., Czerwinski R.N., Sanchez A. Advanced trigger development. *Lincoln Laboratory Journal*, 2007, vol. 17, no. 1, pp. 29–60.
8. Pan Y-L. Detection and characterization of biological and other organic-carbon aerosol particles in atmosphere using fluorescence. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2015, vol. 150, pp. 12–35. doi: 10.1016/j.jqsrt.2014.06.007
9. Burstein E.A., Abornev S.M., Reshetnyak Y.K. Decomposition of protein tryptophan fluorescence spectra into log-normal components. I. Decomposition algorithms. *Biophysical Journal*, 2001, vol. 81, no. 3, pp. 1699–1709.
10. Reshetnyak Y.K., Burstein E.A. Decomposition of protein tryptophan fluorescence spectra into log-normal components. II. The statistical proof of discreteness of tryptophan classes in proteins. *Biophysical Journal*, 2001, vol. 81, no. 3, pp. 1710–1734.
11. Reshetnyak Y.K., Koshevnik Y., Burstein E.A. Decomposition of protein tryptophan fluorescence spectra into log-normal components. III. Correlation between tryptophan and microenvironment parameters of individual tryptophan residues. *Biophysical Journal*, 2001, vol. 81, no. 3, pp. 1735–1758.
12. Pan Y-L., Hill S.C., Pinnick R.G., Huang H., Bottiger J.R., Chang R.K. Fluorescence spectra of atmospheric aerosol particles using one or two excitation wavelengths: comparison of classification schemes employing different emission and scattering results. *Optics Express*, 2010, vol. 18, no. 12, pp. 12436–12457. doi: 10.1364/OE.18.012436



13. Kochelaev E.A., Volchek A.O. *Method for Optical Detection of Fluorescence and Scattering signals of Aerosol Particles in Stream and Optical System for Realising Said Method*. Patent RU2448340, 2012.
14. Esbensen K.H. *Multivariate Data Analysis – In Practice*. 5<sup>th</sup> ed. Oslo, Norway, CAMO Process AS, 2002, 598 p.
15. Rodionova O.E., Pomerantsev A.L. *Khemometrika v Analiticheskoi Khimii* [Chemometrics in Analytical Chemistry]. Moscow, Institute of Chemical Physics RAS Publ., 2006, 61 p.
16. Loo B.W., Cork C.P. Development of high efficiency virtual impactors. *Aerosol Science and Technology*, 1988, vol. , no. 3, pp. 167–176.
17. Rao N.P., Navanscues J., Fernandez de la Mora J. Aerodynamic focusing of particles in viscous jets. *Journal of Aerosol Science*, 1993, vol. 24, no. 7, pp. 879–892. doi: 10.1016/0021-8502(93)90068-K
18. Beebe K.R., Pell R.J., Seasholtz M.B. *Chemometrics: A Practical Guide*. NY, John Wiley & Sons, 1998, 360 p.
19. Hastie T., Tibshirani R., Friedman J. *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2<sup>nd</sup> ed. NY, Springer, 2009, 745 p. doi: 10.1007/978-0-387-84858-7

- |                                       |  |
|---------------------------------------|--|
| <i>Худяков Евгений Сергеевич</i>      | – инженер, АО «НПО «Прибор», Санкт-Петербург, 199034, Российская Федерация, laanbird@gmail.com   |
| <i>Кочелаев Евгений Александрович</i> | – кандидат технических наук, руководитель группы, АО «НПО «Прибор», Санкт-Петербург, 199034, Российская Федерация, kochelaev@bk.ru   |
| <i>Волчек Андрей Олегович</i>         | – кандидат физико-математических наук, главный конструктор, АО «НПО «Прибор», Санкт-Петербург, 199034, Российская Федерация, volchek@npo-pribor.ru   |
| <i>Курсанов Дмитрий Олегович</i>      | – кандидат химических наук, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; доцент, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, d.kirsanov@gmail.com |
| <i>Джагацпнян Игорь Эдуардович</i>    | – кандидат технических наук, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, drjie@mail.ru   |
| <i>Evgeniy S. Khudyakov</i>           | – engineer, JSC "NPO "Pribor", Saint Petersburg, 199034, Russian Federation, laanbird@gmail.com  |
| <i>Evgeniy A. Kochelaev</i>           | – PhD, Head of group, JSC "NPO "Pribor", Saint Petersburg, 199034, Russian Federation, kochelaev@bk.ru   |
| <i>Andrey O. Volchek</i>              | – PhD, Chief Designer, JSC "NPO "Pribor", Saint Petersburg, 199034, Russian Federation, volchek@npo-pribor.ru  |
| <i>Dmitry O. Kirsanov</i>             | – PhD, engineer, ITMO University, Saint Petersburg, 197101, Russian Federation; Associate Professor, Saint Petersburg State University, Saint Petersburg, 199034, Russian Federation, d.kirsanov@gmail.com                         |
| <i>Igor E. Jahatspanian</i>           | – PhD, engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, drjie@mail.ru  |