DOI: 10.18721/JPM.14311
UDC 543.51+ 681.2–5

AN ALGORITHM FOR PROCESSING MASS SPECTROMETRIC ANALYSIS DATA FOR INITIAL DIAGNOSTICS OF DISEASES BY EXHALED GASES

V.V Manoilov, L.V. Novikov, A.G. Kuzmin, Yu.A. Titov, I.V. Zarutskii

Institute for Analytical Instrumentation of the Russian Academy of Sciences, St. Petersburg, Russian Federation

An algorithm for processing mass spectra of gases exhaled by patients has been proposed in the paper. The mass spectra are recorded on the MS7-200 quadrupole mass spectrometer, with electronic ionization and with direct capillary injection of the sample. The algorithm is based on transforming an array of spectra (not less than 10) in the space of principal components. The probability of a disease is determined through the Euclidean distance of the patient's coordinates from the centroid. Testing of the algorithm was carried out on the data of mass spectra of gases exhaled by cancer patients. The proposed procedure has several advantages over traditional laboratory methods. The algorithm uses the multidimensional probability density of the distribution of the parameters of the exhaled gases of control groups and the patient being tested and allows to compile an overall picture of the patient's probable diseases in a short time.

Keywords: diagnostics, principal component analysis, multivariate probability density, multivariate data processing

Citation: Manoilov V.V., Novikov L.V., Kuzmin A.G., Titov Yu.A., Zarutskii I.V., An algorithm for processing mass spectrometric analysis data for initial diagnostics of diseases by exhaled gases, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 14 (3) (2021) 139–149. DOI: 10.18721/JPM.14311

АЛГОРИТМ ОБРАБОТКИ ДАННЫХ МАСС-СПЕКТРОМЕТРИЧЕСКОГО АНАЛИЗА ДЛЯ ПЕРВИЧНОЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ПО ВЫДЫХАЕМЫМ ГАЗАМ

В.В. Манойлов, Л.В. Новиков, А.Г. Кузьмин, Ю.А. Титов, И.В. Заруцкий

Институт аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация

В работе предложен алгоритм обработки масс-спектров газов, выдыхаемых пациентами. Масс-спектры регистрируются на квадрупольном масс-спектрометре МС7-200 с электронной ионизацией и прямым капиллярным вводом пробы. Алгоритм основан на преобразовании массива спектров (не менее десяти) в пространство главных компонент. Вероятность заболевания определяется по евклидову расстоянию координат пациента от центроиды. Тестирование алгоритма проведено на данных масс-спектров газов, выдыхаемых больными с онкологическими заболеваниями. Предлагаемая процедура имеет ряд преимуществ перед традиционными лабораторными методами; алгоритм использует многомерную плотность вероятности распределения параметров выдыхаемых газов контрольных групп и тестируемого пациента; позволяет за короткое время составить общую картину вероятных заболеваний.

Ключевые слова: диагностика, метод главных компонент, многомерная плотность вероятности, обработка многомерных данных

Ссылка при цитировании: Манойлов В.В., Новиков Л.В., Кузьмин А.Г., Титов Ю.А., Заруцкий И.В. Алгоритм обработки данных масс-спектрометрического анализа для первичной диагностики заболеваний по выдыхаемым газам // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2021. Т. 14. № 3. С. 139–149. DOI: 10.18721/JPM.14311

Introduction

Exhaled breath analysis is becoming increasingly popular in diagnostics [1, 2]. This technique allows identifying a number of serious diseases, including Alzheimer's disease [3], epilepsy [4], kidney disease [5], cancer [6] and others. The algorithms supporting the devices used for such analysis are an integral component in this type of diagnostics. Software incorporating the given algorithm processes the mass spectra of exhaled gases; the composition of gases is measured using a spectrometer with a quadrupole analyzer. The algorithm is tested by measuring the exhaled gases in cancer patients.

The classical approach to mass spectral analysis involves the so-called initial, or preliminary processing of mass spectra; in this case, mass spectral peaks are detected and their parameters (position on the scale of mass numbers and amplitude, or sometimes area) are assessed. In some cases, initial processing of mass spectra means that the resolution of the device is improved mathematically by separating the superimposed spectral lines [7].

Initial data processing is followed by secondary processing. The algorithm for processing mass spectra described below is run at this stage in offline mode based on the data recorded in computer memory as line spectra. The software that runs the algorithm yields a preliminary estimate of how likely it is that the subjects tested belong to the class of patients with a specific disease or to the class of practically healthy people.

Thus, the algorithm complements the options for classification offered by discriminant and cluster analysis based on multivariate statistical data. Classification of mass spectra of exhaled gases by the methods of discriminant and cluster analysis is described in [8].

The main difference between the mass spectra of exhaled gases in patients with diseases from those in practically healthy people are additional components, for example, with masses of 71, 64, 59 and 55 Da. The presence of these lines in the mass spectrum of the gases exhaled by the examined patient points towards a high probability of disease. Similar to fingerprints, each mass spectrum of exhaled air has its own unique profile (an array of peaks with different intensities), so it can be assumed that different profiles of exhalation ‘fingerprints’ can serve as indicators of a certain disease. After a sufficient amount of such data has been collected, it will be possible to diagnose specific diseases with a very high efficiency, and either prescribe additional tests or immediately start treating the patient.

Basic assumptions

An array of mass spectral data called the ‘healthy’ group forms a matrix. Peak intensities of the mass spectra are recorded for all columns of the matrix, except for the first one, at the given mass values. The first column of the matrix contains the mass spectrum number of exhaled gases of a certain patient, and each of its rows contains data about the intensity of individual mass spectral lines.

We denote such a matrix as $\mathbf{XS} = [x_{ij}]$, where i, j are the numbers of subjects (patients) and attributes (mass numbers), respectively, $i = 1,2,..., I$, $j = 1, 2,..., J$; x_{ij} is the intensity of the j^{th} spectral component of the i^{th} healthy subject.

Similarly, we produce the matrix $I_0 \times J$ for the mass spectra of exhaled gases in diseased subjects. Each row of this matrix contains the intensities of individual lines of the corresponding mass spectrum. We denote this matrix $\mathbf{XI_1} = \left[x_{ij}^0 \right]$, where i, j, J have the same meaning but $i = 1, 2, ..., I_0$; x_{ij}^0 is the intensity of the jth spectral component of the ith diseased subject.

Let us combine the matrices of healthy and diseased subjects, obtaining a common data matrix **X** of size $(I_0 + I) \times J$:

$$\mathbf{X} = \left(\left[\mathbf{XS}; \mathbf{XI}_1 \right] \right).$$

We transform this matrix by the algorithm for constructing the space of principal components (PC) and calculate the score matrix **T**.

Based on clinical examinations of a large number of diseased and healthy subjects, the obtained mass spectra are divided into two sets: control and test groups.

Control groups of mass spectra are the data for subjects with a known diagnosis for their disease and the data for healthy subjects. The situation is different for processing the mass spectra array for the tested group: it is not known whether they belong to healthy or diseased subjects.

The number of mass spectra the gases exhaled by patients in each of the two groups should be at least 30. Each mass spectrum obtained for the breath exhaled by a patient from the control group is represented in the multidimensional space as a point. Such points, collected together from all patients in the group, form a 'cloud' of initial data. The closer to the center (centroid) of the cloud is the point describing the mass spectrum of the breath exhaled by the test patient, the higher the probability that the patient suffers from the disease corresponding to the control group. If the patient is healthy, the probability of their disease is estimated by the control group of healthy subjects; the farther from the center (centroid) of the cloud is the point corresponding to the mass spectrum of the breath exhaled by the test patient, the higher the probability that the patient suffers from a disease

From a mathematical standpoint, the mass spectra of the gases exhaled by subjects from one group make up a data space with some known disease. The probability that a tested subject belongs to one of these groups is estimated sequentially using the formula for the multivariate probability density.

The examination described yields a complete probabilistic picture of the disease in the tested subject. This allows to quickly develop the approaches to further, more thorough examination and subsequent treatment.

Theoretical description of the algorithm

Let x_{ij} be the j^{th} parameter, in the specific case, the intensity of the spectral line for the data of the i^{th} patient from one of the groups, where $i = 1, 2, ..., I$ and $j = 1, 2, ..., J$, and an array of I patients with J recorded spectral lines of exhaled gases is considered. Let us form a matrix of size $I \times J$ from these data. We call this a training matrix and denote it as **X**. The columns of this matrix are denoted as X_j: $\mathbf{X} = [X_1, ..., X_j, ..., X_J]$ [9].

The vector X_j is a random variable. Suppose that this random variable has a normal distribution with the expected value $\overline{X}_j$ and the variance σ_j^2. Matrix elements **X** form a 'cloud' in J dimensional space. We denote this cloud as G, consisting of $I \times J$ points, with the centroid at $\overline{X}$. Its coordinates are expressed as as

$$\overline{X} = \left[\overline{X}_1, \overline{X}_2, ..., \overline{X}_j \right].$$

The subject considered is diseased (or healthy), if the measured vector of parameters of the air they exhale takes the form

$$X_d = \left[X_{d,1}, X_{d,2}, ..., X_{d,j} \right].$$

$X_{i,j}$ is the parameter in J dimensional space G, i.e., $X_d \in G$. This condition is satisfied if the probability P that the point X_d deviates from the centroid X does not exceed a certain threshold α. This probability is calculated by constructing a multivariate probability distribution of the event belonging to the space G, assuming that this distribution obeys the normal law [10]:

$$P\left(\mathbf{X}_d\right) = W * \exp\left\{-\frac{1}{2}\left(\mathbf{X}_d - \overline{\mathbf{X}}\right)^T \mathbf{K}_X^{-1}\left(\mathbf{X}_d - \overline{\mathbf{X}}\right)\right\}, \mathbf{X}_d \in G, \quad (1)$$

where $\mathbf{K}_X$ is the covariance matrix, $\mathbf{K}_X = \mathrm{E}\left[\left(\mathbf{X} - \overline{\mathbf{X}}\right) * \left(\mathbf{X} - \overline{\mathbf{X}}\right)^T\right]$ (E is the symbol for the expected value); W is the normalizing factor; $(...)^T$ is the symbol for matrix transposition.

Obviously, the probability P of this event is equal to unity, if it turns out for the next patient examined that $\mathbf{X}_d \equiv \overline{\mathbf{X}}$. This condition is fulfilled if we assume that $W = 1$ in Eq. (1). Then the condition under which the tested patient belongs to the control group takes the form $P(\mathbf{X}_d) < \alpha$, where the quantity α is selected by the method of expert assessment.

Notably, calculating the required probability P by Eq. (1) carries considerable computational difficulties due to a large number of parameters J and correlations between the columns of the matrix $\mathbf{X}$. To compress data and reduce the dimension of the space G, we introduce the orthogonal transformation of data into a space of principal components (the principal components analysis, abbreviated as PCA) [11].

We make a transition to the PC space by generating a new matrix consisting of all rows of the matrix $\mathbf{X}$ and the row $\mathbf{X}_d$. We denote this matrix as $\mathbf{XI}$. In the new coordinate system, we obtain

$$\mathbf{XI} = \mathbf{TP}^T + \mathbf{e} = \sum_{j=1}^{A} t_j \mathbf{p}_j^T + \mathbf{e}, \quad (2)$$

where $\mathbf{p}_j^T$ are the eigenfunctions of the covariance matrix $\mathbf{K}_X$; $\mathbf{T}$ is called the score matrix and is expressed as $\mathbf{T} = [T_1, T_2, ..., T_A]$, its dimension is $I \times A$; $\mathbf{P}^T$ is called the load matrix, its dimension is also $I \times A$; $\mathbf{e}$ is the matrix of residuals (noises), its dimension is $(I \times J)$; the column vectors T_j ($j = 1, 2, ..., A$) are called the principal components (A is the number of principal components).

The magnitude of A is much less than the number of variables J. This circumstance means that almost all primary information about the state of the tested patient is concentrated in the first few PCs. The last row of the matrix (2), the vector T_d, are the coordinates of the patient's parameters in the PC space:

$$\mathbf{T}_d = \left[t_{d,1}, t_{d,2}, ..., t_{d,A} \right].$$

The mean values of the columns in the matrix $\mathbf{T}$ are equal to zero, and the variance is the vector σ^2 with the elements is equal to $\sigma_j^2 = \lambda_j$, i.e., the eigenvalues of the covariance matrix. The expansion in PC is characterized by the variance rapidly decreasing by the fourth PC, while the columns of the matrix $\mathbf{T}$ are not correlated, i.e.,

$$\mathbf{T}_m \mathbf{T}_n' = \begin{cases} 0, \text{with } n \neq m, \\ \lambda_j, \text{with } n = m. \end{cases}$$

In view of this, Eq. (1) takes the following form in the new coordinate system:

$$P(\mathbf{T}_d) = \exp\left\{-\frac{1}{2}\mathbf{T}_d\sigma^{-2}\mathbf{T}_d^{'}\right\} = \exp\left\{-\frac{1}{2}\sum_{j=1}^{A}\frac{t_{d,j}^2}{\sigma_j^2}\right\} \quad (3)$$

and the Euclidean distance from the patient with index d to the centroid of the control group is equal to

$$D(d) = \sqrt{\sum_{j=1}^{A} t_{d,j}^2}. \quad (4)$$

Main steps of the algorithm

The algorithm considered can be divided into two stages, training and diagnostics, executed one after the other.

K matrices $\mathbf{X}^k$ ($k = 1, 2, ..., K$) are formed at the first stage (training) based on the results of clinical examination of a large number of patients, where K is the number of subgroups with the information about the intensity of spectral lines for the exhaled gases corresponding to each type of disease. Notice that the number of rows I^k of each matrix must be greater than the number of columns J^k. Training is performed once, based on the data accumulated on the spectra of gases exhaled by patients (examined in-clinic).

The spectral components of gases exhaled by the tested patients are measured at the second stage of the algorithm (diagnostics), and the probability whether they belong to one of the control groups is calculated in several steps.

Step 1. Use the data on the intensities of the spectral components at the given masses to construct a row vector $\mathbf{X}_d = \{x_{d,1}, x_{d,2}, ..., x_{d,J}\}$.

Step 2. Calculate the matrix $\mathbf{XI}^k = \left[\mathbf{X}^k; \mathbf{X}_d\right]$ with the size $\left(I^k + 1 \times J^k\right)$; the last row in this matrix is the vector X_d.

Step 3. Calculate the normalized values of this matrix. The normalization factor is the maximum value of its elements.

Step 4. Calculate the score matrix $\mathbf{T}^k$ and the eigenvalue vector λ_j^k by the algorithm of the PC method [12].

Step 5. Find the principal components and determine the number A (number of principal components) of the matrix $\mathbf{T}^k$ in such a way that $\sigma_A^k = \sqrt{\lambda_A^k} < \varepsilon$. We can assume that $\varepsilon = (0.001, ..., 0.01)$ without loss of reliability. The last row of the matrix $\sigma_A^k = \sqrt{\lambda_A^k} < \varepsilon$ are the principal components of the parameters characterizing the gases exhaled by the examined patient.

Step 6. Calculate the probability $P\left(T_d^k\right), k = 1, 2, ..., K,$ by Eq. (3).

Step 7. Go back to step 2 ($k = k + 1$) to determine the probability that the examined patient belongs to another group of possible diseases.

Step 8. Analyze the calculated probabilities $P\left(\mathbf{T}_d^k\right), k = 1, 2, ..., K,$ for all K diseases in order to determine the disease with the highest probability.

Testing the developed algorithm

The gases exhaled by the examined patients were injected into a quadrupole mass spectrometer, which has a detection range of 1 – 200 Da, mass number resolution of 0.5 Da, and detection time of up to 15 s for one mass spectrum.

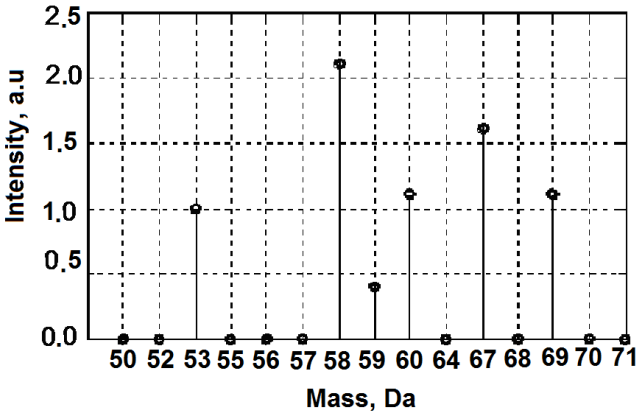

Fig. 1. Mass spectrum of gases exhaled by the patient with a disease

Exhaled gases were measured in 43 patients diagnosed with cancer[1]. These patients were divided into the control group (36 subjects) and the test group (7 subjects). The intensity of the exhaled gas components of the first test patient, for example, coincided with the centroid of the control group. The effectiveness of the algorithm was verified with 10 more tested patients[2] with other types of diseases. Patients from both test groups were assigned serial numbers: $d = (1, 2, ..., 7)$ and $d = (1, 2, ..., 10)$, respectively.

The analyzer recorded the mass spectrum of the gases exhaled by each patient in the range from 50 to 100 Da. Examining a large number of patients and subsequently analyzing the obtained data, we found that the exhaled gases contained spectrum components at masses of 53, ..., 69 Da (12 in total), which are markers of most diseases. A matrix was constructed from the initial data on the control group, containing 12 columns; the intensities of the spectral components serving as markers were located along its rows.

Fig. 1 shows an example for a mass spectrum of the exhaled gases from one of 7 test patients. The mass of singly charged ions in daltons (Da) is plotted along the horizontal axis. The data on exhaled gases of the tested patients were selected sequentially and Steps *1 – 6* of the algorithm were performed. Apparently, it is sufficient to assume that $A = 6$ in the PC space, since $\sigma^k \cong 0$. Eqs. (4) and (3) were used to calculate the Euclidean distance to the centroids of the control group and the probability of diagnosing cancer in each tested patient. These results are given in Table for all tested patients.

Fig. 2,*a* shows the space of the first two principal components of the matrix (PC1 and PC2), with the plotted coordinates of patients in control group 1 (see Table) (o) and seven tested (+) patients. It follows from the data in Table for this group consisting of 7 people that as the coordinates of the mass spectrum of exhaled gases move away from the center of the cloud (centroid), the probability of cancer in these patients decreases but remains high.

Fig. 2,*b* shows the coordinates of patients from control group 2 and 10 tested patients in the space of the first two PCs (PC1 and PC2). As it turned out, all patients (except for the first one) fell outside the cloud of the control group. It is evident from the data for group 2 given in Table that the probability of a different type of cancer is negligible for these patients compared to the type of cancer in group 1 patients.

[1] Data on exhaled gases for this group were provided by the Almazov National Medical Research Centre, St. Petersburg.

[2] Data on exhaled gases for this group were provided by the N.N. Petrov National Medical Research Center of Oncology, St. Petersburg.

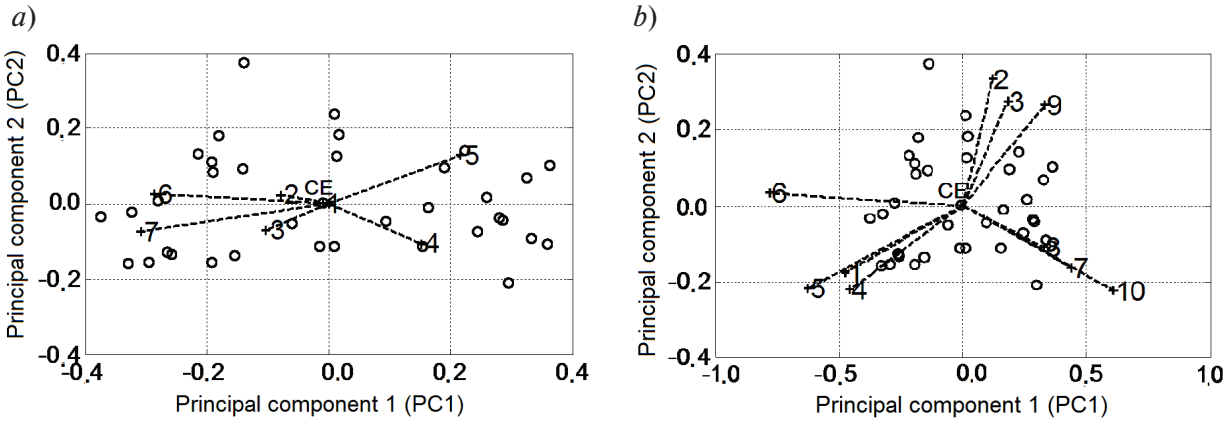

Fig. 2. Examination results in the space of principal components (PC1 and PC2) of the matrix for control groups 1 (*a*) and 2 (*b*) (see Table).
The central points (CE) correspond to the Euclidean center, the rest of the points (o) to the coordinates (mass spectral data) of patients from the corresponding groups. Dashed lines show Euclidean distances to patients (+) who were assigned numbers

Table

Estimated probabilities of cancer in patients from two control groups

Patient no.	Euclidean distance	Probability of disease
Group 1		
1	0.0116	0.9854
2	0.1076	0.1888
3	0.1499	0.2055
4	0.2044	0.1164
5	0.2554	0.2864
6	0.2892	0.3458
7	0.3247	0.1197
Group 2		
1	2.8791	0.0159
4	4.5739	0.0000
3	4.7337	0.0000
2	5.5927	0.0000
5	5.6321	0.0000
8	5.7129	0.0000
9	5.7322	0.0000
6	5.8520	0.0000
10	5.8635	0.0000
7	5.9168	0.0000

Notes. 1. Data for patients from *groups 1 and 2* were obtained from different medical institutions (see footnotes 1 and 2 on page ..., respectively).
2. Patients from *group 2* suffered from cancer of a different type compared with patients from *group 1*.

Conclusion

The medical diagnostics algorithm that we have developed is based on multivariate probability density characterizing the parameter distribution of exhaled gases in the control groups and in patients tested. The algorithm allows to diagnose the most likely diseases for the patient and, if necessary, recommend further tests or treatment. The proposed algorithm and method can also be useful for screening patients, for example, during annual checkups.

Whether the algorithm is applied effectively depends on the size and reliability of the mass spectral data array for healthy controls and the diseased group. The data array for the healthy group should include the data for the detected mass spectra of gases exhaled by the people previously examined by various medical specialists (with high accuracy). These specialists must confirm that no pathologies were found at the time of the experiment on mass spectrometric analysis of exhaled gases. For example, the data array for the healthy group can be compiled from the mass spectra of exhaled gases taken from volunteers in military service, or cadets of military schools and academies who have been thoroughly examined by medical specialists and were found fit for service or training. We have recently carried out experiments on mass spectrometric analysis of exhaled gases in more than 30 cadets of S.M. Kirov Military Medical Academy (St. Petersburg), who were thoroughly examined by medical specialists, making it possible to construct a data array for the healthy group and to refine the test results given in Table.

The data array for the group with diseases was constructed from a large number of mass spectra collected from patients with a well-established diagnosis, for example, a certain type of cancer. This data array was compiled in the course of the study but should be expanded by at least two or three times to improve the reliability of the health decisions made for the patients examined.

The range of capabilities provided by the algorithm can be further extended for processing the data of the mass spectra collected from the exhaled gases not only for devices with quadrupole analyzers but also for devices with other types of analyzers: static, time-of-flight, etc. Studies with such devices can use a system for injecting the sample into the analyzer, similar to the one in our experiments.

The study was performed as part of State Assignment 075-00780-19-02 (registration ID AAAA-A19-11905319-0069-2)

REFERENCES

1. **Li X., Huang D., Zeng J., et al.,** Positive matrix factorization: A data preprocessing strategy for direct mass spectrometry-based breath analysis, Talanta. 192 (15 January) (2019) 32–39.

2. **Li X., Huang L., Zhu H., Zhou Z.,** Direct human breath analysis by secondary nano electrospray ionization ultrahigh resolution mass spectrometry: Importance of high mass resolution and mass accuracy, Rapid Commun. Mass Spectrom. 31 (3) (2017) 301–308.

3. **Tiele A., Wicaksono A., Daulton E., et al.,** Breath-based non-invasive diagnosis of Alzheimer's disease: A pilot study, J. Breath Res. 14 (2) (2020) 026003.

4. **Van Dartel D., Schelhaas H.J., Colon A.J., et al.,** Breath analysis in detecting epilepsy, J. Breath Res. 14 (3) (2020) 031001.

5. **Chen Ch.-Ch., Hsieh J.-Ch., Chao Ch.-H., et al.,** Correlation between breath ammonia and blood ureanitrogen levels in chronic kidney disease and dialysis patients, J. Breath Res. 14 (3) (2020) 036002.

6. **Kononov A., Korotetsky B., Jahatspanian I., et al.,** Online breath analysis using metal oxide semiconductor sensors (electronic nose) for diagnosis of lung cancer, J. Breath Res. 14 (1) (2020) 016004.

7. **Manoilov V.V., Novikov L.V., Belozertsev A.I., et al.,** Processing of mass spectra of exhaled gases based on correlation algorithms, J. Analytical Chem. 75 (13) (2020) 1678–1684.

8. **Manoilov V.V., Novikov L.V., Zarutskii I.V., et al.,** Methods for processing mass spectrometry signals from exhaled gases for medical diagnosis, Biomed. Eng. 53 (5) (2020) 355–359.

9. **Bolshakov A.A., Karimov R.N.,** Metody obrabotki mnogomernykh dannykh i vremennykh ryadov [Methods for processing multivariate data and time series], Goryachaya Liniya-Telekom Publishing, Moscow, 2007 (in Russian).

10. **Shiryayev A.N.,** Probability, Springer-Verlag, New-York, 1996.

11. **Ayvazyan S.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D.,** Prikladnaya statistika. Klassifikatsiya i snizheniye razmernosti [Applied statistics. Classification and dimension reduction], Finance and Statistics Publishing, Moscow, 1989 (in Russian).

12. **Esbensen K.H.,** Multivariate data analysis – in Practice, CAMO, Oslo, 1994 – 2002.

Received 21.06.2021, accepted 19.07.2021.

THE AUTHORS

MANOILOV Vladimir V.
Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
manoilov_vv@mail.ru

NOVIKOV Lev V.
Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
novik38@mail.ru

KUZMIN Aleksey G.
Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
agqz55@rambler.ru

TITOV Yuriy A.
Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
toplm@mail.ru

ZARUTSKII Igor V.
Institute for Analytical Instrumentation of the Russian Academy of Sciences
26, Rizhsky Ave., St. Petersburg, 190103, Russian Federation
igorzv@yandex.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Li X., Huang D., Zeng J., Chan Ch.K., Zhou Z.** Positive matrix factorization: A data preprocessing strategy for direct mass spectrometry-based breath analysis // Talanta. 2019. Vol. 192. 15 January. Pp. 32–39.

2. **Li X., Huang L., Zhu H., Zhou Z.** Direct human breath analysis by secondary nano electrospray ionization ultrahigh resolution mass spectrometry: Importance of high mass resolution and mass accuracy // Rapid Communications in Mass Spectrometry. 2017. Vol. 31. No. 3. Pp. 301–308.

3. **Tiele A., Wicaksono A., Daulton E., Ifeachor E., Eyre V., Clarke S., Timings L., Pearson S., Covington J.A., Li X.** Breath-based non-invasive diagnosis of Alzheimer's disease: A pilot study // Journal of Breath Research. 2020. Vol. 14. No. 2. P. 026003.

4. **Van Dartel D., Schelhaas H.J., Colon A.J., Kho K.H., de Vos C.C.** Breath analysis in detecting epilepsy // Journal of Breath Research. 2020. Vol. 14. No. 3. P. 031001.

5. **Chen Ch.-Ch., Hsieh J.-Ch., Chao Ch.-H., Yang W.-Sh., Li X., Cheng H.-T., Chan Ch.-K., Lu Ch.-J., Meng H.-F., Zan H.-W.** Correlation between breath ammonia and blood ureanitrogen levels in chronic kidney disease and dialysis patients // Journal of Breath Research. 2020. Vol. 14. No. 3. P. 036002.

6. **Kononov A., Korotetsky B., Jahatspanian I., et al.** Online breath analysis using metal oxide semiconductor sensors (electronic nose) for diagnosis of lung cancer // Journal of Breath Research. 2020. Vol. 14. No. 1. P. 016004.

7. **Manoilov V.V., Novikov L.V., Belozertsev A.I., Zarutskii I.V., Titov Yu.A., Kuzmin A.G., El-Salim S.Z.** Processing of mass spectra of exhaled gases based on correlation algorithms // Journal of Analytical Chemistry. 2020. Vol. 75. No. 13. Pp. 1678–1684.

8. **Manoilov V.V., Novikov L.V., Zarutskii I.V., Kuzmin A.G., Titov Yu.A.** Methods for processing mass spectrometry signals from exhaled gases for medical diagnosis // Biomedical Engineering. 2020. Vol. 53. No. 5. Pp. 355–359.

9. **Большаков А.А., Каримов Р.Н.** Методы обработки многомерных данных и временных рядов. М.: Горячая линия-Телеком, 2007. 522 с.

10. **Ширяев А.Н.** Вероятность. В 2 кн. 4-е изд., перераб. и доп. Т. 1. Вероятность–1. М.: МЦНМО, 2007. 552 с.

11. **Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.** Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.

12. **Эсбенсен К.** Анализ многомерных данных: избранные главы. Пер. с англ. Под ред. О.Е. Родионовой. Черноголовка.: ИПХФ РАН, 2005. 161 с.

Статья поступила в редакцию 21.06.2021, принята к публикации 19.07.2021.

СВЕДЕНИЯ ОБ АВТОРАХ

МАНОЙЛОВ Владимир Владимирович – *доктор технических наук, заведующий лабораторией автоматизации измерений и цифровой обработки сигналов Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.*

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26

manoilov_vv@mail.ru

НОВИКОВ Лев Васильевич – *доктор физико-математических наук, ведущий научный сотрудник лаборатории оптики заряженных частиц и математического моделирования Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.*

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26

novik38@mail.ru

КУЗЬМИН Алексей Георгиевич – *кандидат физико-математических наук, старший научный сотрудник лаборатории экологической масс-спектрометрии Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.*

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26

agqz55@rambler.ru

ТИТОВ Юрий Алексеевич – *научный сотрудник лаборатории автоматизации измерений и цифровой обработки сигналов Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.*

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26

toplm@mail.ru

ЗАРУЦКИЙ Игорь Вячеславович – *кандидат технических наук, старший научный сотрудник лаборатории автоматизации измерений и цифровой обработки сигналов Института аналитического приборостроения Российской академии наук, Санкт-Петербург, Российская Федерация.*

190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26

igorzv@yandex.ru