

DETERMINING OPTIMAL AMBIENT IONIZATION MASS SPECTROMETRY DATA PRE-PROCESSING PARAMETERS IN NEUROSURGERY

Zavorotnyuk DS¹✉, Sorokin AA¹, Bormotov DS¹, Eliferov VA¹, Bocharov KV², Pekov SI^{1,3,4}, Popov IA^{1,4}

¹ Moscow Institute of Physics and Technology, Moscow, Russia

² Semenov Federal Research Center for Chemical Physics of the Russian Academy of Sciences, Moscow, Russia

³ Skolkovo Institute of Science and Technology, Moscow, Russia

⁴ Siberian State Medical University, Tomsk, Russia

Radical tumor resection is still the most effective treatment method for brain tumors. The problems of intraoperative monitoring are currently solved using positron emission tomography, magnetic resonance imaging, and histochemical analysis, however, these require using expensive equipment by highly qualified personnel and are therefore still not widely available. As an alternative, it is possible to use mass spectrometry methods without sample preparation and then the analysis of mass spectrometry data involving the use of machine learning methods. The spectra that are more rich and diverse in terms of peak number are typical for mass spectrometry without sample preparation, therefore the use of this method requires specific pre-processing of experimental data. The study was aimed to develop the methods to determine the optimal parameter values for pre-processing of the data acquired by ambient ionization mass spectrometry. The paper presents two such methods and provides specific parameter values for the data acquired using the Thermo LTQ XL Orbitrap ETD mass spectrometer.

Keywords: mass spectrometry, ambient ionization, data analysis, data preprocessing

Funding: the study was performed within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation (agreement № 075-03-2022-107, project № 0714-2020-0006). The study involved the use of equipment of the Semenov Federal Research Center for Chemical Physics RAS.

Author contribution: Zavorotnyuk DS — data acquisition and interpretation, software development, manuscript writing and editing; Sorokin AA — study planning, data analysis and interpretation, manuscript editing; Bormotov DS — data acquisition and interpretation, manuscript writing; Eliferov VA — financial support of the experiment; Bocharov KV — data acquisition; Pekov SI — study planning, data analysis and interpretation, manuscript draft writing and manuscript text finalization; Popov IA — project management, financial support.

Compliance with ethical standards: the study was approved by the Ethics Committee of the Burdenko Research Institute of Neurosurgery (protocols № 40 dated 12 April 2016 and № 131 dated 17 July 2018) and conducted in accordance with the principles of the Declaration of Helsinki (2000) and its subsequent revisions. All patients submitted the informed consent to study participation and the use of biomaterial for scientific purposes.

✉ **Correspondence should be addressed:** Denis S. Zavorotnyuk
Institutskiy per., 9, str. 7, Dolgoprudny, Moscow Region, 141701; denis.zavorotnyuk@gmail.com

Received: 19.12.2023 **Accepted:** 03.03.2024 **Published online:** 27.04.2024

DOI: 10.24075/brsmu.2024.013

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНЫХ ПАРАМЕТРОВ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ МАСС-СПЕКТРОМЕТРИИ С ПРЯМОЙ ИОНИЗАЦИЕЙ В НЕЙРОХИРУРГИИ

Д. С. Заворотнюк¹✉, А. А. Сорокин¹, Д. С. Бормотов¹, В. А. Елиферов¹, К. В. Бочаров², С. И. Пеков^{1,3,4}, И. А. Попов^{1,4}

¹ Московский физико-технический институт, Москва, Россия

² Федеральный исследовательский центр химической физики имени Н. Н. Семенова Российской академии наук, Москва, Россия

³ Сколковский институт науки и технологий, Москва, Россия

⁴ Сибирский государственный медицинский университет, Томск, Россия

Радикальное удаление опухоли до сих пор остается наиболее эффективным методом лечения онкологических заболеваний головного мозга. Задачи интраоперационного мониторинга на сегодняшний день решают с помощью позитронно-эмиссионной томографии, магнитно-резонансной томографии и гистохимического анализа, однако они требуют применения дорогостоящего оборудования высококвалифицированным персоналом, поэтому до сих пор не получили широкого распространения. В качестве альтернативы возможно применение методов масс-спектрометрии без пробоподготовки с последующим анализом масс-спектрометрических данных методами машинного обучения. Так как для масс-спектрометрии без пробоподготовки характерны более богатые и разнообразные по количеству пиков спектры, ее применение требует специальной предварительной обработки экспериментальных данных. Целью исследования было разработать методы определения оптимальных значений параметров предварительной обработки данных масс-спектрометрии без пробоподготовки. В работе представлены два таких метода, а также приведены конкретные значения параметров для данных, полученных с помощью масс-спектрометра Thermo LTQ XL Orbitrap ETD.

Ключевые слова: масс-спектрометрия, прямая ионизация, анализ данных, предварительная обработка данных

Финансирование: работа выполнена в рамках государственного задания Министерства науки и высшего образования (соглашение № 075-03-2022-107, проект № 0714-2020-0006). Исследование выполнено с использованием оборудования ЦКП ФИЦ ХФ им. Н. Н. Семенова РАН.

Вклад авторов: Д. С. Заворотнюк — анализ и интерпретация данных, создание программного обеспечения, написание и редактирование рукописи; А. А. Сорокин — планирование исследования, анализ и интерпретация данных, редактирование рукописи; Д. С. Бормотов — сбор и интерпретация данных, написание рукописи; В. А. Елиферов — финансовое обеспечение эксперимента; К. В. Бочаров — сбор данных; С. И. Пеков — планирование исследования, анализ и интерпретация данных, редактирование рукописи. И. А. Попов — руководство проектом, обеспечение финансирования.

Соблюдение этических стандартов: исследование одобрено этическим комитетом НМИЦН имени Н. Н. Бурденко (протоколы № 40 от 12 апреля 2016 г. и № 131 от 17 июля 2018 г.), проведено в соответствии с принципами Хельсинкской декларации (2000 г.) и ее последующих пересмотров. Все пациенты подписали добровольное информированное согласие на участие в исследовании и использование биоматериалов в исследовательских целях.

✉ **Для корреспонденции:** Денис Сергеевич Заворотнюк
Институтский переулок, д. 9, 141701, г. Долгoprудный, Московская область; denis.zavorotnyuk@gmail.com

Статья получена: 19.12.2023 **Статья принята к печати:** 03.03.2024 **Опубликована онлайн:** 27.04.2024

DOI: 10.24075/vrgmu.2024.013

Ambient ionization mass spectrometry represents one of the promising methods to improve accuracy and completeness of the glial tumor resection, since radical tumor removal is currently the most effective treatment method for brain tumors [1]. However, there is a problem of identifying the tumor margins in order to ensure resection completeness for relapse prevention on the one hand and prevention of excessive resection and development of neuropathological sequelae on the other hand [2]. The main universal methods to ensure intraoperative control of the resected tumor margins still include positron emission tomography-computed tomography (PET-CT), magnetic resonance imaging (MRI), and histochemical analysis, since other methods, such as fluorescence staining, can turn out to be non-specific for certain diagnoses. However, these methods are time-consuming, and tomography is also expensive due to the need to equip the specialized surgical units [3].

Ambient ionization mass spectrometry (MS) makes it possible to quickly acquire the data on the molecular structure of the sample [4–6]. However, today, the vast majority of computational tools to deal with mass spectrometry data involve working with the spectra acquired by tandem MS coupled with gas/liquid chromatography. These data are distinguished by the fact that the number of peaks per scan of such a spectrum is much less than the number per scan obtained by ambient ionization MS [7, 8]. When using ambient ionization MS, the sample preparation simplicity and analysis speed make it possible to acquire far more complex mass spectra, i.e., large amounts of data within minutes. At the same time, the analysis of such data requires the use of automated processing methods and complex analysis algorithms [9–11], therefore, great attention should be paid to the data quality control and pre-processing [12].

Mass spectrometry data are the time-ordered sets of scans. Each scan represents the profile of the ion current intensities accumulated by the instrument over a certain time that is ordered on the mass-to-charge ratio (m/z) scale. In the pre-processing phase, it is necessary to transform this scan into the set consisting of intensities and m/z values of the detected peaks. Usually, this is achieved through implementation of such steps as normalization of intensity values, noise determination and elimination, peak position determination and alignment [13–15]. The great diversity of approaches to MS data processing suggests that the above steps should be implemented with various parameters depending on the nature of samples used in the study, mass spectrometer construction, ion acquisition mode, and the type of further analysis.

The paper describes the method to determine the mass spectra pre-processing parameters in order to ensure unification of mass spectrometry data for further automated analysis on the example of the experimental data obtained by mass spectrometry without sample preparation when assessing human brain tumor tissue samples.

METHODS

The study involved mass spectrometry data acquired when processing brain tissue samples of the individual diagnosed with glioblastoma and grade IV astrocytoma (according to the 2021 WHO classification [16]) and non-neoplastic samples obtained during surgical treatment of drug-resistant epilepsy. A total of 307 tissue samples obtained from 74 patients were assessed. The data were acquired using the Thermo LTQ XL Orbitrap ETD mass spectrometer (Thermo Fisher Scientific; USA) with an inline cartridge extraction [3, 17]. Each sample was separated into two parts. The first part was sent for

standard histochemical analysis to obtain a medical record on the sample, while the remaining part was used to extract three fragments, about 1 mm³ each, to be subjected to mass spectrometry analysis. The mass spectrometry protocol involved the analysis and detection of ions in eight different modes, each of which was characterized by the ions' polarity, detector resolution and bandwidth of the registered ions' m/z . Ion acquisition was performed twice in each mode.

The experimental data acquired were pre-processed using different values of the parameters described in the Results section. The pre-processing procedure involved peak intensity calibration, peak alignment relative to the scan showing maximum total ion current (TIC), reciprocal alignment of peaks among scans performed in the same mode of ion detection and filtration of rare and low-intensity peaks. Distinct scan sets were obtained for each ion detection mode. Each set of scans was transformed into the matrix of peak intensities used to train a classification model. When training the models, the matrix columns containing distributions of peak intensities across all scans of the appropriate mode were used as predictors, while the patients' histological diagnoses were used as response. The mass spectrometry data acquired for brain tissue samples of 33 patients diagnosed with glioblastoma and seven patients diagnosed with non-neoplastic disorders were used to train and validate the models. The dataset available for each mode was divided into the training and validating groups in a ratio of 3 : 1, respectively; division was implemented in such a way that different scans of the same sample were present in both groups, to reduce model overfitting.

The data were analyzed using the computer running Ubuntu 16.04 with the installed R package v. 3.4.4 and R packages MALDIquant [18], caret [19], glmnet [20], ggplot2 [21]. For that the data received from the mass spectrometer were converted from the source Thermo Finnigan format to the open NetCDF format [22] using the in-lab developed software tool [23].

RESULTS

In 2012, it was shown that the differences between mass spectra of tumors and non-neoplastic brain tissues could be used for construction of the classifiers for automatic recognition of cancerous tissues in biopsy samples [24]. Fig. 1 demonstrates peaks of two mass scans of the tissue samples obtained from the patients diagnosed with glioblastoma and non-neoplastic disorders.

The mass spectrometry data pre-processing procedure consists of several phases. In the first phase, noise is assessed and the signal-to-noise ratios are determined for all scans:

$$SNR = \frac{I_s}{I_n},$$

where I_s is signal intensity, I_n is noise intensity. There are several methods to determine the digital data noise intensity, for example, using mean absolute deviation (MAD) or regression with adaptive bandwidths (Super Smoother) [25]. In the subsequent phases, the low-intensity peaks with the signal-to-noise ratios lower than the specified SNR value are excluded from the spectrum. Positions of maxima within the scan may vary slightly under exposure to variable environmental factors and occasional fluctuation. In the next phase, alignment of profiles in different scans is performed to compensate for such changes. The scan showing maximum TIC is used as a reference one, since it is assumed that this scan has the largest number of reported ions, and its profile comprises the largest number of various ion peaks. Here every profile is

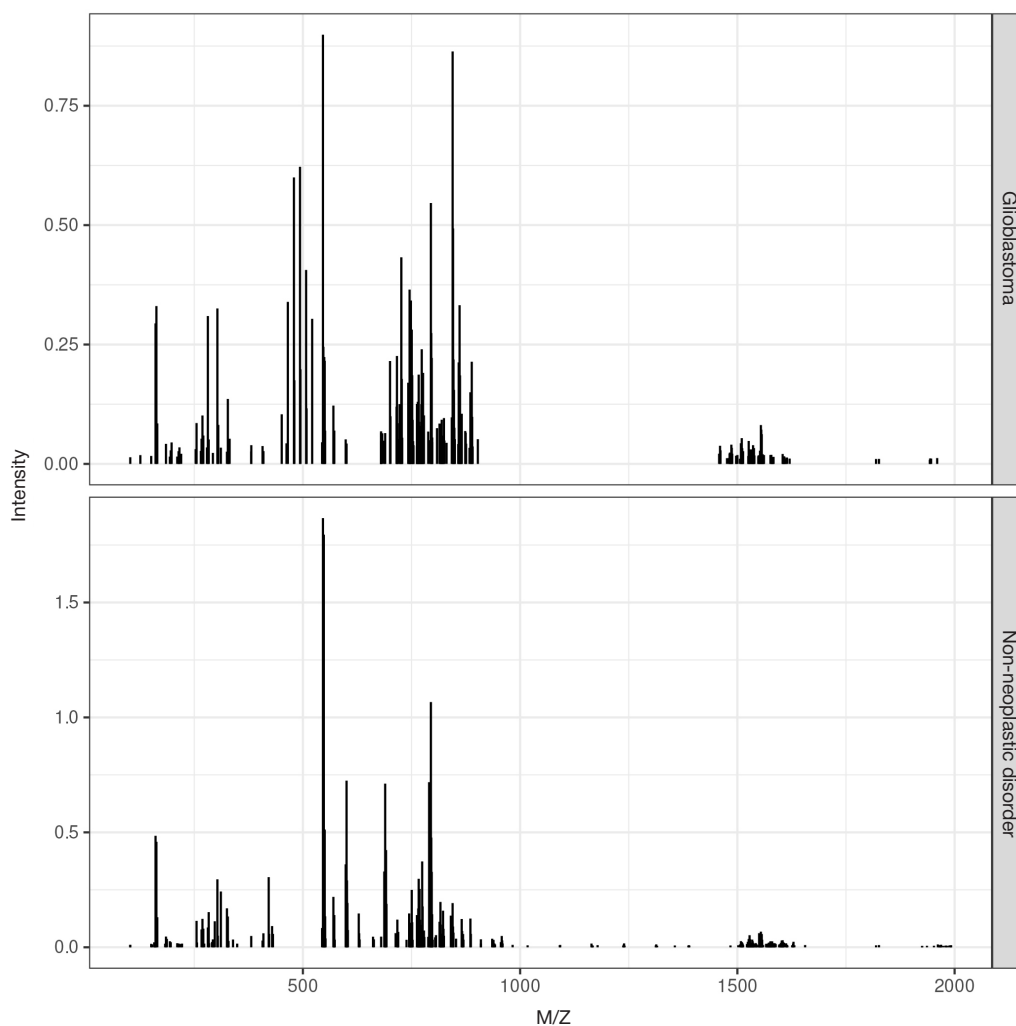


Fig. 1. Comparison of peaks in mass scans of neoplastic tumors and non-neoplastic specimens samples

subjected to alignment along the m/z axis to become as similar to the reference profile as possible. The maximum permissible value of such alignment is specified using the alignment tolerance (TA). Then peaks are detected: the scan profile is converted into the set of individual peaks. For that the entire profile is divided into several parts. The size of each part is determined by the half window size (HWS) representing the range of m/z points, within which the search for a point with the maximum intensity value is carried out. This point is designated as a peak in this part of the profile. Then positions of identical peaks are aligned across the entire set of scans. Here, peaks, the differences in m/z between which do not exceed the tolerance specified when detecting peaks (TBP), are considered to be identical. In the final phase, rare peaks are removed, and peaks of all scans are combined into the common matrix of intensities.

Thus, as a result of mass spectrometry data pre-processing, the matrix is produced [26], the number of rows in which is determined by the number of scans obtained during the experiment, while the number of rows represents the combined number of peaks from all scans. It is clear that the above parameters (SNR, TA, HWS, and TBP) have a significant impact on the number of peaks in the matrix of intensities and the question, which values these parameters should take in each particular ion acquisition mode, is not trivial.

In the classic tasks to determine the model that best describes experimental data [27, 28], the information criteria [29] are used and the extreme values of this criteria correspond to optimal values of the set of model construction criteria obtained with the regularization method. In our study, the minimum

value of the classic Akaike information criterion (AIC) [30] was used to determine the optimal SNR value. Optimality of other parameters (HWS, TA and TBP) was determined based on the manual evaluation of spectra processing quality.

SNR parameter

The optimal SNR value was determined using the Akaike criterion of the LASSO classification models. For that we made a combination of SNR, TA and TBP values, pre-processed the mass spectra, constructed the matrix of intensities, and then trained the LASSO model using the matrix and the patient's diagnosis as the training data. Training of models involved 5/10-fold cross-validation, and the best model was selected based on the Accuracy metric. The parameter combinations were made of value sets:

$$\begin{aligned} \text{SNR} &= \{1.5, 2\} \\ \text{TA} = \text{TBP} &= \{20, 200, 2000\} \end{aligned}$$

The combination of parameters, with which the resulting model had the lowest AIC value, was named optimal. The optimal parameter values are provided in Table 1.

To prevent the emergence of negative noise intensities in the scan, 100 nulls were added to the set of points (M/Z , Intensity) on the left and on the right. As a result, the noise signal was evaluated in the broader range of M/Z values with a constant number of significant peaks in the spectrum.

Table 1. Optimal SNR values, which correspond to the LASSO models with minimal AIC values

Scanning mode	SNR	TA = TBP, ppm
Negative, High, 120–2000	1.5	20
Negative, High, 500–1000	2	2000
Negative, Low, 120–2000	1.5	20
Negative, Low, 500–1000	2	2000
Positive, High, 120–2000	1.5	2000
Positive, High, 500–1000	2	2000
Positive, Low, 120–2000	1.5	20
Positive, Low, 500–1000	2	2000

HWS, TA, TBP parameters

Optimality of the HWS, TA and TBP parameters was determined by manual evaluation of spectra processing quality. The interactive Shiny application Mass-Spectrum Observer allowing one to explore, how the spectrum shape, peak positions and characteristics of the intensity matrix of certain mass scan change with changing values of these parameters, was developed for this purpose. The application source code is available from GitHub repository [31], and the application demo version is available from the open access library of Shiny applications [32]. The screen-captured images of the application are provided in Fig. 2 and 3.

The lists of possible HWS, TA and TBP values were determined, and the mass spectrometry data pre-processing procedures were applied to each combination of these values in order to obtain separate matrices of intensities for each ion acquisition mode. The TBP parameter was proportional to the TA parameter with three possible proportionality coefficient values. The lists of parameter values are provided in Table 2.

The number of columns corresponding to the total number of peaks obtained from the mass scan profiles was

determined for each matrix of intensities. Furthermore, when constructing the intensity matrix, we determined the number of peaks located close to each other in the resulting spectra. When the distance between peaks was smaller than two instrument resolutions during detection of ions in this mode, the peaks were considered as probably duplicate. Such peaks can emerge during conversion of the scan profiles into the sets of individual peaks, for example, within the same scan at too low HWS values, with the result that the intensity spike that is relatively broad on the m/z scale is represented by several spectral peaks, or in the scans of the same file at low TBP values, due to which the algorithm cannot compile the list of identical peaks from different scans. The duplicate peaks were determined within the same scan, in all scans of the same tissue specimen sample used for mass spectrometry analysis, and among all peaks of the intensity matrix. Peak duplication was defined based on the mass spectrometer resolution in this ion acquisition mode; the value of 800 at m/z = 400 was selected for the low-resolution mode, the value of 30,000 at m/z = 400 was selected for the high-resolution mode.

The reference HWS, TA and TBP values that were later subjected to manual evaluation performed using Mass-Spectrum Observer were determined based on the changes

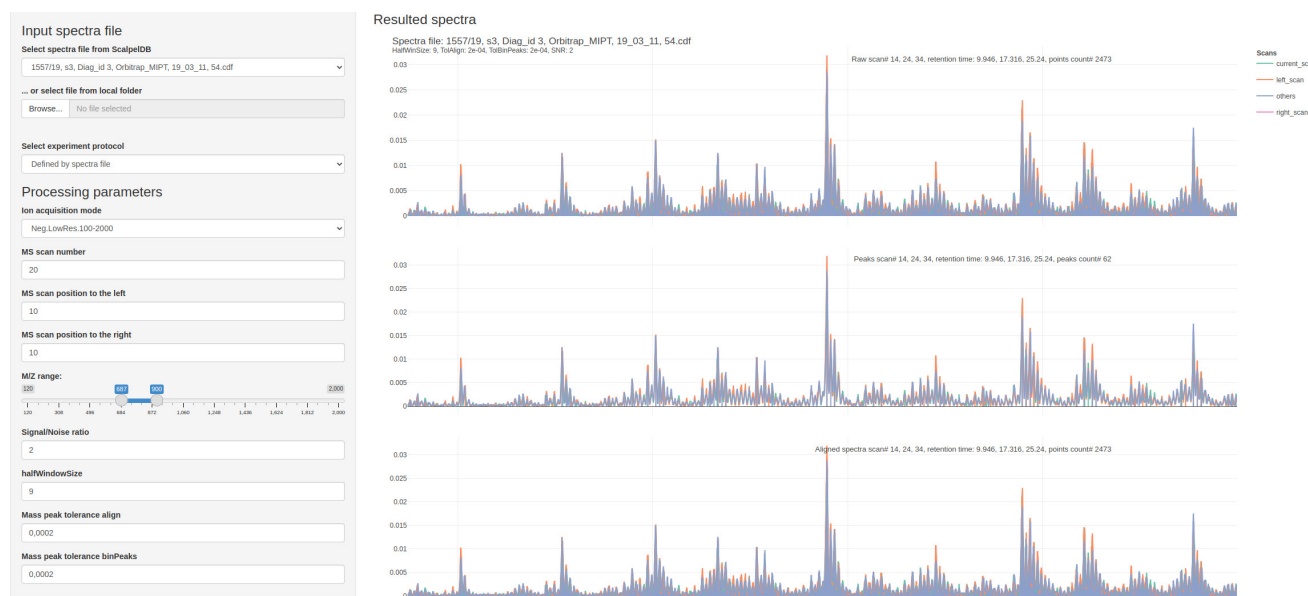


Fig. 2. Screen-captured image of the Mass-Spectrum Observer application window with the spectrum pre-processing parameter control panel

Table 2. Possible HWS, TA, and TBP values

Parameter	Values for high resolution	Values for low resolution
HWS	{3, 5, 7}	{7, 9, 11, 13, 15, 17, 19}
TA, ppm	{1, 20.8, 40.6, 60.4, 80.2, 100, 208, 406, 604, 802, 1·10 ³ }	{100, 325, 550, 775, 1·10 ³ }
TBP = m·TA	m := {0.1, 1, 10}	

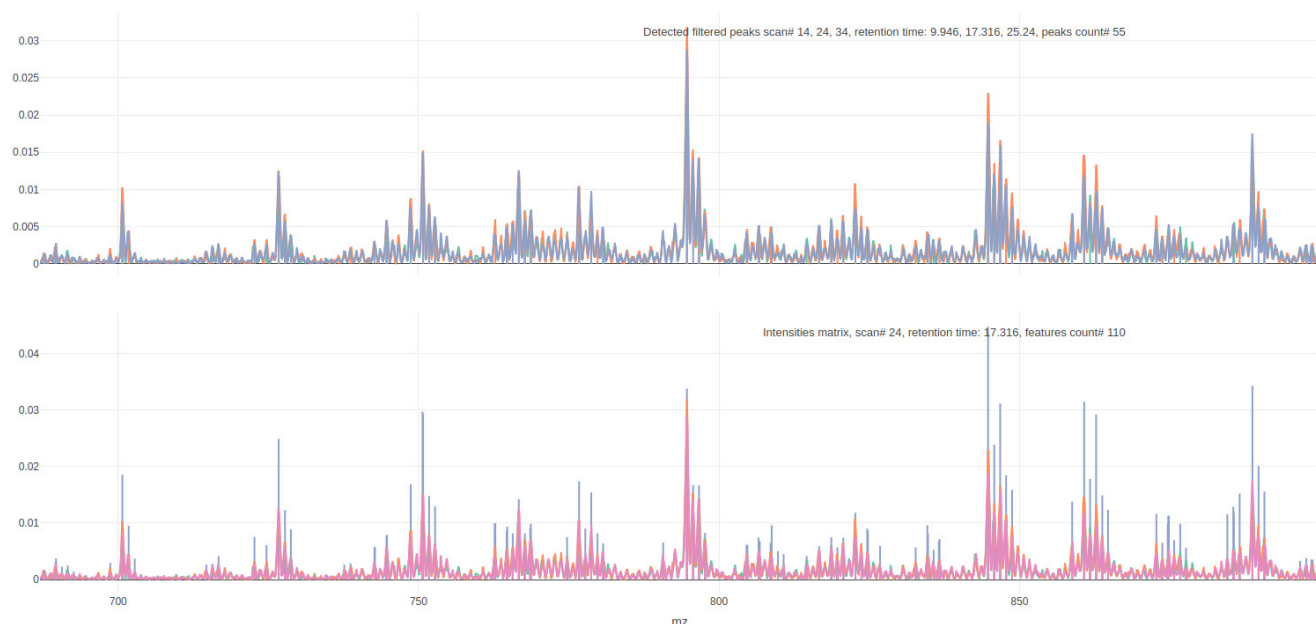


Fig. 3. Screen-captured of the Mass-Spectrum Observer application window with the plots corresponding to the spectra yielded after applying the pre-processing procedure

in these four indicators in accordance with the processing parameters. The manual evaluation results are provided in Table 3.

DISCUSSION

The findings show a close relationship between the ambient ionization mass spectrometry data pre-processing parameters and the quality of acquired spectra. The SNR parameter makes it possible to reduce the number of peaks in the resulting spectrum. However, attention should be paid to the presence of the negative estimate of noise signal values that may occur in the border spectral regions as an artifact. When detecting peaks in the profile, the noise estimate is used to determine peak intensity in this region of the profile, so negative noise can result in the emergence of the excessive number of peaks in the spectrum. This may not matter much in case of ion detection in the broad M/Z range (for example, 120–2000), but may be significant for the narrow range of 500–1000. In some cases, it is possible to eliminate such artifacts by fine-tuning the Super Smoother method (for example, by changing the smoothness degree during approximation or by narrowing the profile region, for which noise estimation is performed). However, these methods can yield different results for each particular mass scan, therefore the method of false dataset expansion was selected as a more sustainable method to eliminate negative values.

The HWS, TA and TBP values should be selected based primarily on the instrument resolution. The increase in half window size during the profile conversion into the intensity matrix enables

elimination of artifact and duplicate peaks on the one hand (Fig. 4), but on the other hand the too high values of this parameter lead to exclusion of significant peaks from the subsequent analysis (Fig. 5). The values of peak position tolerance at alignment and detection are also closely related to the half window size and, therefore, to resolution, as well as to other mass spectrometer features resulting from the mass drift and the signal digitization methods. Furthermore, the TBP value should not be less than the TA value, since such configuration of values always results in the increase in the average number of possible duplicate peaks. This is due to the fact that the algorithm does not have enough tolerance for shift of identical peaks in different scans to eliminate duplicate peaks even after alignment of all scans relative to the scan with the highest ion current. It should be also noted that changing the width of the range without changing resolution and polarity of the detected ions has no significant effect on the parameter values, which is considered the expected result.

CONCLUSIONS

We developed a universal approach to determining the optimal parameter values for pre-processing of the data acquired by ambient ionization MS. The use of this approach was demonstrated on the data acquired by assessing human brain tissue samples using the Thermo LTQ XL Orbitrap ETD mass spectrometer. The approach developed can be used to determine the optimal parameter values for pre-processing

Table 3. Optimal HWS, TA and TBP values acquired by manual evaluation

Ion acquisition mode	TA, ppm	TBP, ppm	HWS
Negative, High, 120–2000	40.6	40.6	3
Negative, High, 500–1000	60.4	60.4	3
Negative, Low, 120–2000	775	7.75×10^3	13
Negative, Low, 500–1000	1×10^3	1×10^3	13
Positive, High, 120–2000	60.4	60.4	3
Positive, High, 500–1000	60.4	60.4	3
Positive, Low, 120–2000	1×10^3	1×10^4	13
Positive, Low, 500–1000	1×10^3	1×10^4	13

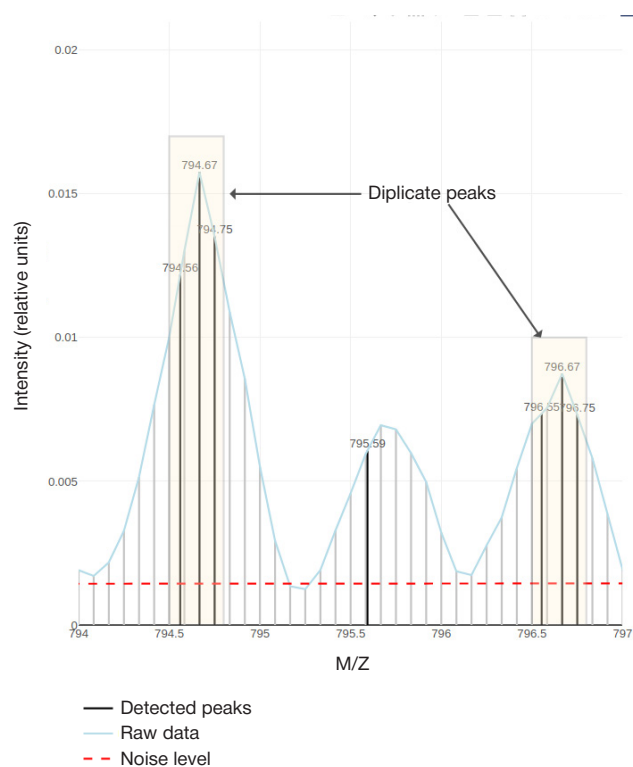


Fig. 4. Determining peak positions. The emergence of peaks, the distance between which in the wide-range low-resolution mass scan of negative ions obtained at suboptimal processing parameter values is less than two resolutions of the instrument in this ion detection mode (duplicate peaks)

of the data acquired when assessing samples of other types using other mass spectrometry equipment. The findings show that it is necessary to thoroughly adjust the mass spectrometry data processing parameters when using ambient ionization MS in the clinics as the faster and more affordable alternative to conventional intraoperative monitoring methods. Parameters have to be determined considering the mass spectrometer and research conditions. In particular, the SNR parameter determining the number of peaks in the resulting spectra should be selected based on the assessed tissue type and

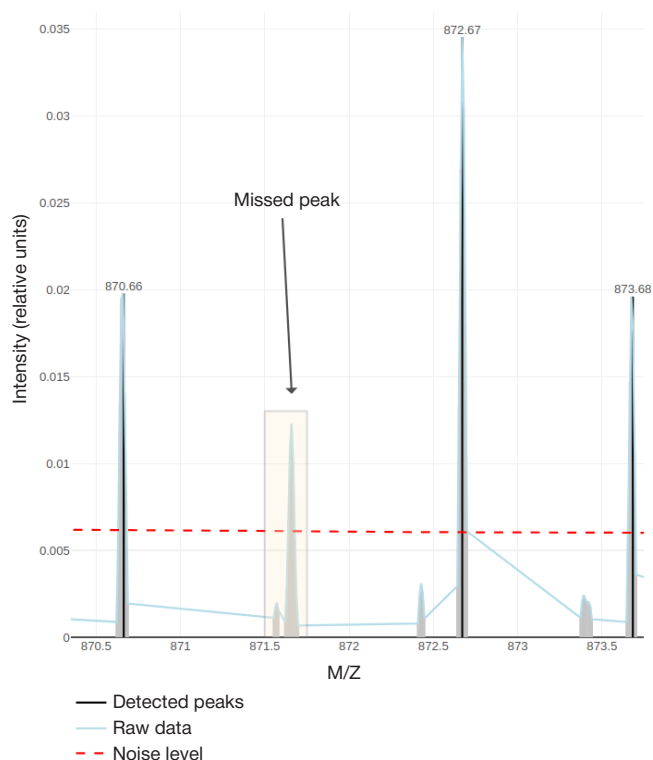


Fig. 5. Determining peak positions. A missed significant peak in the narrow-range high-resolution mass scan of negative ions obtained at suboptimal processing parameter values

ionization method, while the value of 1.5–2 can be considered the lower limit. When performing scan profile alignment and peak detection, the half window size (HWS) and scan modification tolerance (TA) should be selected in accordance with the resolution of the mass spectrometer used, and the tolerance for spectra peak alignment (TBP) should not be lower than the TA value. Both machine learning methods and manual evaluation of the quality of acquired spectra can be used to choose optimal values of these parameters from several options.

References

- Young RM, Jamshidi A, Davis G, Sherman JH. Current trends in the surgical management and treatment of adult glioblastoma. *Ann Transl Med* 2015; 1–15. <https://doi.org/10.3978/j.issn.2305-5839.2015.05.10>.
- Chanbour H, Chotali S. Review of intraoperative adjuncts for maximal safe resection of gliomas and its impact on outcomes. *Cancers*. 2022; 14: 5705. Available from: <https://doi.org/10.3390/cancers14225705>.
- Pekov SI, Bormotov DS, Nikitin PV, Sorokin AA, Shurkhay VA, Eliferov VA, et al. Rapid estimation of tumor cell percentage in brain tissue biopsy samples using inline cartridge extraction mass spectrometry. *Anal Bioanal Chem*. 2021; 413: 2913–22. Available from: <https://doi.org/10.1007/s00216-021-03220-y>.
- Eberlin LS, Norton I, Orringer D, Dunn IF, Liu X, Ide JL, et al. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc Natl Acad Sci*. 2013; 110: 1611–6. Available from: <https://doi.org/10.1073/pnas.1215687110>.
- Hänel L, Kwiatkowski M, Heikaus L, Schlüter H. Mass spectrometry-based intraoperative tumor diagnostics. *Future Sci OA*. 2019; 5: FSO373. Available from: <https://doi.org/10.4155/fsoa-2018-0087>.
- Li L-H, Hsieh H-Y, Hsu C-C. Clinical application of ambient ionization mass spectrometry. *Mass Spectrom*. 2017; 6: S0060–S0060. Available from: <https://doi.org/10.5702/massspectrometry.S0060>.
- Huang M-Z, Yuan C-H, Cheng S-C, Cho Y-T, Shiea J. Ambient ionization mass spectrometry. *Annu Rev Anal Chem*. 2010; 3: 43–65. Available from: <https://doi.org/10.1146/annurev.anchem.111808.073702>.
- Shi L, Habib A, Bi L, Hong H, Begum R, Wen L. Ambient Ionization Mass Spectrometry: Application and Prospective. *Crit Rev Anal Chem*. 2022; 1–50. Available from: <https://doi.org/10.1080/10408347.2022.2124840>.
- Boiko DA, Kozlov KS, Burykina JV, Ilyushenkova VV, Ananikov VP. Fully automated unconstrained analysis of high-resolution mass spectrometry data with machine learning. *J Am Chem Soc*. 2022; 144: 14590–606. Available from: <https://doi.org/10.1021/jacs.2c03631>.
- Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites*. 2020; 10: 1–23. Available from: <https://doi.org/10.3390/metabo10060243>.
- Piras C, Hale OJ, Reynolds CK, Jones AK (Barney), Taylor N, Morris M, et al. LAP-MALDI MS coupled with machine learning: an ambient mass spectrometry approach for high-throughput diagnostics. *Chem Sci*. 2022; 13: 1746–58. Available from: <https://doi.org/10.1039/D1SC05171G>.
- Seddiki K, Saudemont P, Precioso F, Ogrinc N, Wisztorski M, Salzet M, et al. Cumulative learning enables convolutional neural network representations for small mass spectrometry

- data classification. *Nat Commun.* 2020; 11. Available from: <https://doi.org/10.1038/s41467-020-19354-z>.
13. Huang YC, Chung HH, Dutkiewicz EP, Chen CL, Hsieh HY, Chen BR, et al. Predicting breast cancer by paper spray ion mobility spectrometry mass spectrometry and machine learning. *Anal Chem.* 2020; 92: 1653–7. Available from: <https://doi.org/10.1021/acs.analchem.9b03966>.
 14. Iwano T, Yoshimura K, Inoue S, Odate T, Ogata K, Funatsu S, et al. Breast cancer diagnosis based on lipid profiling by probe electrospray ionization mass spectrometry. *Br J Surg.* 2020; 107: 632–5. Available from: <https://doi.org/10.1002/bjs.11613>.
 15. Zhou M, Guan W, Walker LDE, Mezencev R, Benigno BB, Gray A, et al. Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. *Cancer Epidemiol Biomarkers Prev.* 2010; 19: 2262–71. Available from: <https://doi.org/10.1158/1055-9965.EPI-10-0126>.
 16. Torp SH, Solheim O, Skjulsvik AJ. The WHO 2021 Classification of central nervous system tumours: a practical update on what neurosurgeons need to know — a minireview. *Acta Neurochir (Wien).* 2022; 164: 2453–64. Available from: <https://doi.org/10.1007/s00701-022-05301-y>.
 17. Bormotov DS, Eliferov VA, Peregudova OV, Zavorotnyuk DS, Bocharov KV, Pekov SI, et al. Incorporation of a disposable ESI emitter into inline cartridge extraction mass spectrometry improves throughput and spectra stability. *J Am Soc Mass Spectrom.* 2023; 34: 119–22. Available from: <https://doi.org/10.1021/jasms.2c00207>.
 18. Gibb S, Strimmer K. Maldiquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics.* 2012; 28. Available from: <https://doi.org/10.1093/bioinformatics/bts447>.
 19. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* 2008; 28: 1–26. Available from: <https://doi.org/10.18637/jss.v028.i05>.
 20. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010; 33: 1–22. Available from: <https://doi.org/10.18637/jss.v033.i01>.
 21. Wickham H. *ggplot2: Elegant graphics for data analysis.* Springer-Verlag, New York; 2016.
 22. Rew R, Davis G, Emmerson S, Cormack C, Caron J, Pincus R, et al. *Unidata NetCDF* 1989. Available from: <https://doi.org/10.5065/D6H70CW6>.
 23. Zavorotnyuk DS, Pekov SI, Sorokin AA, Bormotov DS, Levin N, Zhvansky E, et al. Lipid profiles of human brain tumors obtained by high-resolution negative mode ambient mass spectrometry. *Data.* 2021; 6: 1–7. Available from: <https://doi.org/10.3390/data6120132>.
 24. Eberlin LS, Norton I, Dill AL, Golby AJ, Ligon KL, Santagata S, et al. Classifying human brain tumors by lipid imaging with mass spectrometry. *Cancer Research.* 2012; 72 (3): 645–54. Available from: <https://doi.org/10.1158/0008-5472.can-11-2465>.
 25. Friedman JH. *Smart user's guide.* Stanford Univ CA, Laboratory for Computational Statistics; 1984.
 26. Morris JS, Coombes KR, Koomen J, Baggerly KA, Kobayashi R. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics.* 2005; 21: 1764–75. Available from: <https://doi.org/10.1093/bioinformatics/bti254>.
 27. Burnham KP, Anderson DR, editors. *Model Selection and Multimodel Inference.* New York, NY: Springer New York, 2004. Available from: <https://doi.org/10.1007/b97636>.
 28. Gustafsson F, Hjalmarsson H. Twenty-one ML estimators for model selection. *Automatica.* 1995; 31: 1377–92. Available from: [https://doi.org/10.1016/0005-1098\(95\)00058-5](https://doi.org/10.1016/0005-1098(95)00058-5).
 29. Shitikov VK, Mastitsky SE. *Klassifikacija, regressija i drugie algoritmy Data Mining s ispol'zovaniem R.* 2017. Dostupna po ssylke: <https://github.com/ranalytics/data-mining>. Russian.
 30. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control.* 1974; 19: 716–23. Available from: <https://doi.org/10.1109/TAC.1974.1100705>.
 31. Zavorotnyuk DS. *MS Spectrum observer repository.* Available from: <https://github.com/zdens/MS-Spectrum-Observer/releases/tag/1.0> (data obrashhenija: 29 fevralja 2024 g.).
 32. Zavorotnyuk DS. *MS spectrum observer Demo.* Available from: <https://zdens.shinyapps.io/ms-spectrum-observer> (data obrashhenija: 29 fevralja 2024 g.).

Литература

1. Young RM, Jamshidi A, Davis G, Sherman JH. Current trends in the surgical management and treatment of adult glioblastoma. *Ann Transl Med* 2015; 1–15. <https://doi.org/10.3978/j.issn.2305-5839.2015.05.10>.
2. Chanbour H, Chotai S. Review of intraoperative adjuncts for maximal safe resection of gliomas and its impact on outcomes. *Cancers.* 2022; 14: 5705. Available from: <https://doi.org/10.3390/cancers14225705>.
3. Pekov SI, Bormotov DS, Nikitin PV, Sorokin AA, Shurkhay VA, Eliferov VA, et al. Rapid estimation of tumor cell percentage in brain tissue biopsy samples using inline cartridge extraction mass spectrometry. *Anal Bioanal Chem.* 2021; 413: 2913–22. Available from: <https://doi.org/10.1007/s00216-021-03220-y>.
4. Eberlin LS, Norton I, Orringer D, Dunn IF, Liu X, Ide JL, et al. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc Natl Acad Sci.* 2013; 110: 1611–6. Available from: <https://doi.org/10.1073/pnas.1215687110>.
5. Hänel L, Kwiatkowski M, Heikaus L, Schlüter H. Mass spectrometry-based intraoperative tumor diagnostics. *Future Sci OA.* 2019; 5: FSO373. Available from: <https://doi.org/10.4155/fsoa-2018-0087>.
6. Li L-H, Hsieh H-Y, Hsu C-C. Clinical application of ambient ionization mass spectrometry. *Mass Spectrom.* 2017; 6: S0060–S0060. Available from: <https://doi.org/10.5702/massspectrometry.S0060>.
7. Huang M-Z, Yuan C-H, Cheng S-C, Cho Y-T, Shiea J. Ambient ionization mass spectrometry. *Annu Rev Anal Chem.* 2010; 3: 43–65. Available from: <https://doi.org/10.1146/annurev.anchem.111808.073702>.
8. Shi L, Habib A, Bi L, Hong H, Begum R, Wen L. Ambient Ionization Mass Spectrometry: Application and Prospective. *Crit Rev Anal Chem.* 2022; 1–50. Available from: <https://doi.org/10.1080/10408347.2022.2124840>.
9. Boiko DA, Kozlov KS, Burykina JV, Ilyushenkova VV, Ananikov VP. Fully automated unconstrained analysis of high-resolution mass spectrometry data with machine learning. *J Am Chem Soc.* 2022; 144: 14590–606. Available from: <https://doi.org/10.1021/jacs.2c03631>.
10. Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites.* 2020; 10: 1–23. Available from: <https://doi.org/10.3390/metabo10060243>.
11. Piras C, Hale OJ, Reynolds CK, Jones AK (Barney), Taylor N, Morris M, et al. LAP-MALDI MS coupled with machine learning: an ambient mass spectrometry approach for high-throughput diagnostics. *Chem Sci.* 2022; 13: 1746–58. Available from: <https://doi.org/10.1039/D1SC05171G>.
12. Seddiki K, Saudemont P, Precioso F, Ogrinc N, Wisztorski M, Salzet M, et al. Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nat Commun.* 2020; 11. Available from: <https://doi.org/10.1038/s41467-020-19354-z>.
13. Huang YC, Chung HH, Dutkiewicz EP, Chen CL, Hsieh HY, Chen BR, et al. Predicting breast cancer by paper spray ion mobility spectrometry mass spectrometry and machine learning. *Anal Chem.* 2020; 92: 1653–7. Available from: <https://doi.org/10.1021/acs.analchem.9b03966>.
14. Iwano T, Yoshimura K, Inoue S, Odate T, Ogata K, Funatsu S, et al. Breast cancer diagnosis based on lipid profiling by probe electrospray ionization mass spectrometry. *Br J Surg.* 2020; 107: 632–5. Available from: <https://doi.org/10.1002/bjs.11613>.
15. Zhou M, Guan W, Walker LDE, Mezencev R, Benigno BB, Gray A, et al. Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. *Cancer Epidemiol Biomarkers Prev.* 2010; 19: 2262–71. Available from: <https://doi.org/10.1158/1055-9965.EPI-10-0126>.
16. Torp SH, Solheim O, Skjulsvik AJ. The WHO 2021 Classification of central nervous system tumours: a practical update on what neurosurgeons need to know — a minireview. *Acta Neurochir (Wien).* 2022; 164: 2453–64. Available from: <https://doi.org/10.1007/s00701-022-05301-y>.
17. Bormotov DS, Eliferov VA, Peregudova OV, Zavorotnyuk DS, Bocharov KV, Pekov SI, et al. Incorporation of a disposable ESI emitter into inline cartridge extraction mass spectrometry improves throughput and spectra stability. *J Am Soc Mass Spectrom.* 2023; 34:

- 119–22. Available from: <https://doi.org/10.1021/jasms.2c00207>.
18. Gibb S, Strimmer K. Maldiquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics*. 2012; 28. Available from: <https://doi.org/10.1093/bioinformatics/bts447>.
 19. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008; 28: 1–26. Available from: <https://doi.org/10.18637/jss.v028.i05>.
 20. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010; 33: 1–22. Available from: <https://doi.org/10.18637/jss.v033.i01>.
 21. Wickham H. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York; 2016.
 22. Rew R, Davis G, Emmerson S, Comack C, Caron J, Pincus R, et al. *Unidata NetCDF* 1989. Available from: <https://doi.org/10.5065/D6H70CW6>.
 23. Zavorotnyuk DS, Pekov SI, Sorokin AA, Bormotov DS, Levin N, Zhvansky E, et al. Lipid profiles of human brain tumors obtained by high-resolution negative mode ambient mass spectrometry. *Data*. 2021; 6: 1–7. Available from: <https://doi.org/10.3390/data6120132>.
 24. Eberlin LS, Norton I, Dill AL, Golby AJ, Ligon KL, Santagata S, et al. Classifying human brain tumors by lipid imaging with mass spectrometry. *Cancer Research*. 2012; 72 (3): 645–54. Available from: <https://doi.org/10.1158/0008-5472.can-11-2465>.
 25. Friedman JH. *Smart user's guide*. Stanford Univ CA, Laboratory for Computational Statistics; 1984.
 26. Morris JS, Coombes KR, Koomen J, Baggerly KA, Kobayashi R. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*. 2005; 21: 1764–75. Available from: <https://doi.org/10.1093/bioinformatics/bti254>.
 27. Burnham KP, Anderson DR, editors. *Model Selection and Multimodel Inference*. New York, NY: Springer New York, 2004. Available from: <https://doi.org/10.1007/b97636>.
 28. Gustafsson F, Hjalmarsen H. Twenty-one ML estimators for model selection. *Automatica*. 1995; 31: 1377–92. Available from: [https://doi.org/10.1016/0005-1098\(95\)00058-5](https://doi.org/10.1016/0005-1098(95)00058-5).
 29. Шитиков В. К., Мастицкий С. Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 2017. Доступна по ссылке: <https://github.com/ranalytics/data-mining>.
 30. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974; 19: 716–23. Available from: <https://doi.org/10.1109/TAC.1974.1100705>.
 31. Zavorotnyuk DS. *MS Spectrum observer repository*. Available from: <https://github.com/zdens/MS-Spectrum-Observer/releases/tag/1.0> (дата обращения: 29 февраля 2024 г.).
 32. Zavorotnyuk DS. *MS spectrum observer Demo*. Available from: <https://zdens.shinyapps.io/ms-spectrum-observer> (дата обращения: 29 февраля 2024 г.).