Optics and Spectroscopy, 2022, Vol. 130, No. 10

23

Spectral variables selection in multivariate calibration of concentrations of C, Mn, Si, Cr, Ni and Cu in low-allow steels by laser induced breakdown spectroscopy

© *M.V. Belkov, D.A. Borisevich, K.Y. Catsalap, M.A. Khodasevich*¶

Stepanov Institute of Physics, Belarusian Academy of Sciences,
220072 Minsk, Belarus

¶ e-mail: m.khodasevich@ifanbel.bas-net.by

Received July 5, 2022
Revised July 05, 2022
Accepted July 20, 2022

Multivariate calibrations of concentrations of C, Mn, Si, Cr, Ni and Cu have been developed by the partial least squares method for 31 to 39 standard samples of low-alloy steels using low-resolution emission spectra (190–440 nm, resolution 0.4 nm, step 0.1 nm). Three methods of spectral variables selection are considered: a method of ranking spectral variables by their correlation coefficient with the value of the calibrated parameter, a successive projection algorithm and an original modification of searching combination moving windows. The partial least squares model with the spectral variables selection by the method of the searching combination moving windows for C is quantitative: the root mean square error is 0.004%, the residual predictive deviation in the test dataset is 23.4 in the concentration range 0.13 to 0.43%. Calibrations of Mn (0.04% and 5.2 in the range of 0.47–1.15%), Si (0.003% and 20.7 in the range of 0.15–0.33%), Cr (0.04% and 3.1 in the range of 0.09–0.43%) and Ni (0.01% and 4.8 in the range of 0.05–0.25%) are also quantitative. For Cu in the concentration range of 0.06–0.26%, calibration is qualitative (0.04% and 1.4).

Keywords: laser induced breakdown spectroscopy, multivariate calibration, partial least squares, low-alloy steels

DOI: 10.21883/EOS.2022.10.54875.3895-22

Introduction

In recent years, the use of laser induced breakdown spectroscopy (LIBS) has made it possible to achieve a significant improvement in the characteristics of stationary and portable instruments that perform qualitative or quantitative analysis of the samples under study [1–3]. With the help of LIBS, an express analysis of the composition of various objects is carried out with minimal or no sample preparation, which is a significant advantage over reference chemical methods. For example, in [1] LIBS was used to quantitatively measure the chemical composition of stainless steel welds with temporal and spatial resolution directly during welding. In [2] the detection of various types of adulteration of milk powder using LIBS and machine learning methods is considered. Such falsification can cause serious digestive disorders in consumers. The work [3] is devoted to the search by LIBS for ways to reduce the carbon calibration deviation caused by trace quantities of surface contaminations for low-carbon steels.

Iron-based steels and alloys, which are used in almost all areas of human activity, occupy a special place among the objects of application of the LIBS method. The content of technological impurities in steels, along with alloying with chromium, manganese and other chemical elements, determines their physical, chemical and technological properties. Quantitative analysis methods for determining the concentration of alloying additives and process impurities are important for the classification or sorting of steels. Usually it performs by mass spectrometry [4], optical emission spectroscopy using spark discharge [5], inductively coupled plasma [6] and glow discharge [7]. The advantages of the LIBS method are express multi-element analysis in the open air and the relatively low cost of instrumental implementation with an accuracy acceptable for qualitative analysis. As a major disadvantage of LIBS, it is possible to point out the lack of accuracy when making quantitative measurements [8]. Nevertheless, a large number of works are devoted not only to qualitative, but also to quantitative applications of LIBS [9–15]. Due to differences in the choice of implementation conditions (wavelength, pulse duration, laser beam energy and focusing spectral range and resolution of the spectrometer, time delay and spectrum measurement interval, number of preliminary and measuring laser pulses, number of accumulations, blowing the object with gas), which significantly affect the measured spectra, LIBS is considered a semi-quantitative method [16]. The construction of one- or multivariate quantitative models with various types of spectrum preprocessing also leads to significantly different results in accuracy.

When implementing the LIBS method in portable and mobile instruments, spectrometers with low spectral resolution are usually used. Therefore, in view of the strong overlap of the wings of the emission lines, the classical one-variate approach to constructing a calibration dependence from the intensity of isolated analytical line is of little use for such spectrometers. In this case, multivariate calibration models [1,17] are widely used, which process the entire recorded spectrum. According [18] calibration is the process used to create a model that relates two types of measured data. In the present work, we calibrate the concentration of the main elements in low-alloy steels, i.e. a mathematical model is created that relates the sought concentrations of chemical elements in a set of known reference samples of low alloy steels with spectral data obtained using the LIBS method.

Experiment

Previously, we solved the problem of calibration over the entire spectral range of measuring low-resolution emission spectra (190−440 nm, resolution 0.4 nm, spectral step 0.1 nm). The experimental setup and measurement conditions are given in [19]. Reference samples of low-alloy steels UG0d-UG7d, UG9d (Russia) and 51/1-58/1, 72-76, 101-103, 110-125 (Poland) were studied, a total of 44 standards, of which from 31 to 39 samples were used for calibration with non-recurring concentrations of C (in the range of no more than 0.8%), Mn (2.0%), Si (1.2%), Cr (1.0%), Ni (0.8%) and Cu (0.5%).

Methods and results

Approximately the same number of standards training and test samplings were formed in accordance with the standard Kennard-Stone algorithm [20]: the first sample in the training sampling is selected with a concentration closest to the center of the studied range, and each subsequent one in concentration should be the most remote from those already selected. In our case, this algorithm, in comparison with the uniform or random distribution of the training sampling, makes it possible to obtain more stable models by narrowing the intervals for estimating the concentrations of chemical elements in the test sampling. After the spectra were normalized to the intensity at the characteristic iron emission wavelength of 252.0609 nm, calibration models were constructed using the partial least squares method with standard deviations from the corresponding reference values in the test sampling $RMSEP = 0.06\%$ for C, 0.12% for Mn, 0.09% for Si, 0.13% for Cr, 0.07% for Ni and 0.08% for Cu.

To improve the calibration accuracy in this work, when forming the training and test samplins, we partially took into account the requirements formulated in [18]. Since the conditions for the minimum number of samples in the training (24 samples) and test (20 samples) samplings are not met in total, with the calibration under consideration, approximately 60% of the samples form a training sampling, and the remaining 40% form a test one. The proportion is due to the requirement [18] to train multivariate models by the number of samples, which is equal to the number of used latent variables increased by 6 times. For the test sampling, the corresponding ratio is 4. This problem was solved at the first stage of work. Next, three methods for selecting spectral variables were applied to the broadband multivariate model: the RSV method (ranking of spectral variables) [21], based on the ranking of spectral variables by their correlation coefficient with the value of the desired parameter; successive projections algorithm (SPA) [22] and [23] original modification of scmwiPLS i.e. searching combination moving window iPLS [24]. Let us characterize each of the applied methods for choosing spectral variables and the results obtained.

When using the RSV method, all spectral variables are ranked in order of decreasing correlation coefficient with the concentration being calibrated. Then, at each step, one variable with minimal correlation is removed and multivariate modeling is performed using the partial least squares method [25] with the determination of the optimal number of latent variables. Selected spectral variables are determined by the minimum standard deviation of estimates of the concentration of the sought element in the samples of the test sampling from the reference values. With the introduction of a restriction on the variables selection by a given correlation coefficient, this modification of the RSV method is called as method of filtering by the correlation value, i. e. the SMC (significance multivariate correlation) [26]. The application of the SMC method is characterized by the arbitrariness of the researcher in choosing the indicated restriction, from which the RSV method is spared.

Comparison of the characteristics of the broadband model PLS and PLS + RSV is presented in Table 1. It can be seen that the regrouping of the training and test samplings from the ratio of the number of samples in them equal to 1 : 1 to 3 : 2 led to slight changes in the $RMSEP$ values only for the broadband calibration of Si and Ni concentrations (by 0.02% and 0.01% respectively). This confirms the robustness of PLS multivariate models to changes in samplings. Comparison of the RMSEP values of the PLS and PLS + RSV models shows that the use of the correlation method for spectral variables selection is not efficient enough in the cases under consideration. The quality of the calibration models for Cr and Cu did not change, since out of 3630 spectral variables 3629 and 3625 were selected, respectively. For the other four elements, the quality of the calibration improved, but only slightly. Let us illustrate the obtained results of the multivariate calibration of the C concentration by the PLS + RSV method. Fig. 1 shows the dependence of $RMSEP$ on the number of spectral variables removed from the model, ranked by the correlation coefficient with the C concentration in the samples. The minimum RMSEP is achieved when 757 spectral variables are included in the model, which allow using 4 latent variables in the PLS method to obtain the calibration dependence shown in Fig. 2. The value of $RMSEP$ is 0.04%, and the residual deviation RPD in the test sampling (ratio of performance to deviation is the ratio of the standard deviation of the parameter in the sampling and the standard deviation from the reference values) —

Table 1. Characteristics of PLS and PLS + RSV multivariate models for calibration of the concentration of six chemical elements in low-alloy steels by emission spectra

Element	Number of samples			Number of latent variables in wideband PLS	*RMSE* in test in wideband variables PLS, %	Number of selected spectral variables for PLS + RSV	Number of latent variable for PLS + RSV	*RMSE* in test % for PLS+RSV
	total	in training	in test					
C	33	18	15	3	0.06	757	4	0.04
Mn	39	24	15	4	0.12	3458	3	0.09
Si	34	22	12	3	0.07	426	6	0.05
Cr	33	18	15	2	0.13	3629	2	0.13
Ni	31	18	13	3	0.08	3486	2	0.05
Cu	38	22	16	4	0.08	3625	4	0.08

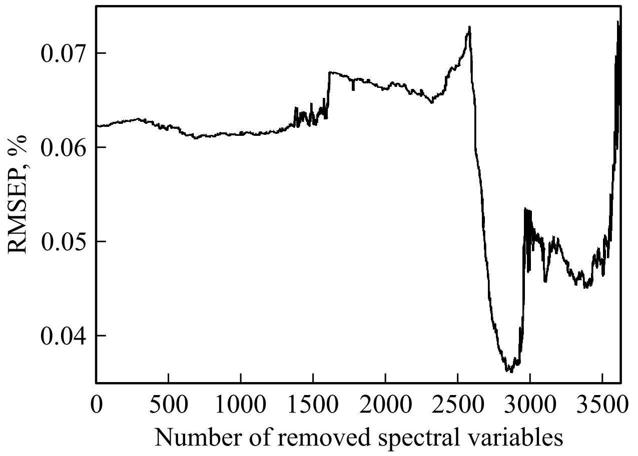

Figure 1. The standard deviation of the C concentration estimate from the reference values in the test sampling of low-alloy steel samples, depending on the number of spectral variables ranked by the correlation coefficient and removed from the partial least squares model.

2.7. This is the best of the models constructed for the 6 considered elements by the PLS + RSV method, but it is only semi-quantitative ($2.5 < RPD < 3$) [27]. Attention is drawn to the location of the selected spectral variables shown in Fig. 3 for the emission spectrum of the etalon 123. It can be seen that only one spectral variable (252.06 nm) lies in the region of intense emission lines, where we can expect the presence of useful information for calibration. The remaining selected spectral variables are at the edges of the measured spectra, which, as expected, are of little information.

The second method we used for selecting spectral variables for multivariate calibration is SPA [22]. At the first stage of the SPA implementation, for each of the available 3630 spectral variables, an ordered sequence is constructed from all other variables. In these sequences, the second variable will be chosen according to the maximum projection into the subspace perpendicular to the first one. And so on until all the measured spectral variables are included in the sequence. At the second stage of SPA, PLS models are constructed for increasing sets of elements of each sequence with the choice of the optimal number of latent variables. In the considered case of limiting the number of latent variables to 10, the number of such models is $3630^2 \cdot 10 \approx 1.3 \cdot 10^8$. At the third stage, the best multivariate model and, accordingly, the desired set of spectral variables, leading to the minimum calibration deviation, are determined from the minimum RMSEP value. Table 2 shows the parameters of the best multivariate models for the PLS + SPA method.

Comparison of the data in Tables 1 and 2 shows that the quality of the PLS+SPA calibration models improved compared to PLS + RSV for Mn, Si, Cr, and Cu, but deteriorated for C. For Ni, *RMSEP* has not changed.

The third method we used for selecting spectral variables is the scmwiPLS method. This method does not operate with individual spectral variables, but with spectral intervals of a certain width i.e. windows The original modification of

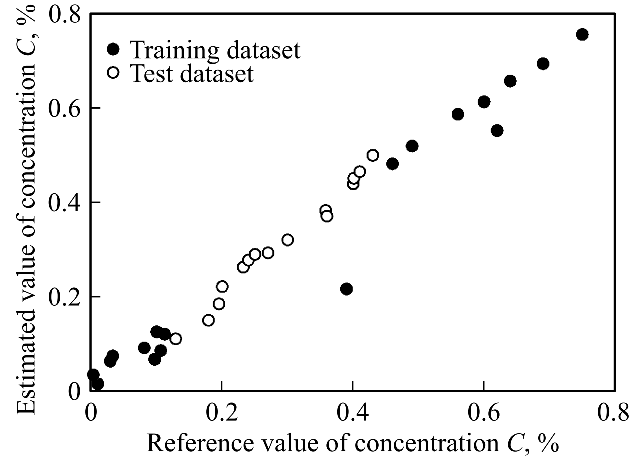

Figure 2. Correlation between estimates of C concentration and reference values in the construction of a multivariate partial least squares model for 757 spectral variables with the maximum correlation with the calibrated parameter.

Table 2. Characteristics of PLS + SPA multivariate models for calibration of the concentration of six chemical elements in low-alloy steels by emission spectra

Element	Number of selected spectral variables for PLS + SPA	Number of latent variables for PLS + SPA	*RMSE* in test, % for PLS + SPA
C	1229	2	0.05
Mn	278	3	0.08
Si	11	10	0.03
Cr	2628	2	0.12
Ni	2696	2	0.05
Cu	11	8	0.07

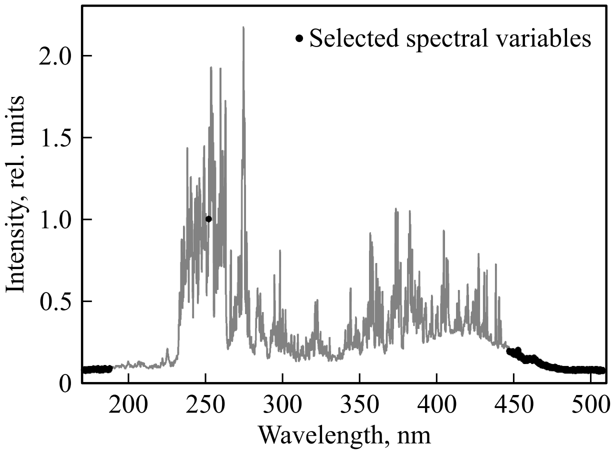

Figure 3. Emission spectrum of standard 123, which indicates the spectral variables used to calibrate the C concentration by the PLS + RSV method.

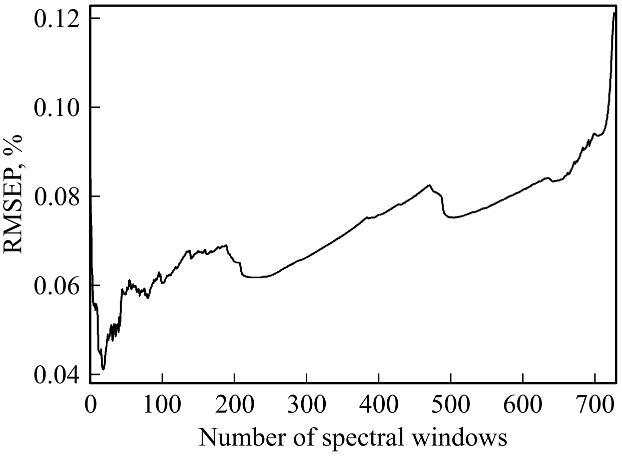

Figure 4. Dependence of the root mean square of prediction of Mn concentration in the test sampling of low alloy steels from emission spectra using a partial least squares model with the spectral variables selection by searching combination moving windows.

the scmwiPLS [23] method contains three stages. The first one is the construction of a broadband multivariate PLS model and the determination of the optimal number of n latent variables from the minimum value of $RMSEP$. At the second stage, the width of the spectral windows is fixed, the number of spectral variables in which exceeds n per unity. This condition minimizes the width of the window while maintaining the ability to select latent variables even in one window. Next, the first window is shifted by one spectral variable per step, and at each step a multivariate model is built by the PLS method, also characterized by $RMSEP$. After the first window reaches the edge of the measured spectral range, the optimal position of this window is determined by the minimum value of $RMSEP$ and is fixed. The procedure is repeated with the second spectral window. In this case, the simulation is performed by combining the spectral variables belonging to the fixed first and moving second spectral windows. Each subsequent window adds $n + 1$ spectral variables to the model until all the measured variables are taken into account in the calibration. For correct operation of scmwiPLS, it is first necessary to reduce the total number of spectral variables to a multiple of $n + 1$. The total number of multivariate scmwiPLS models in the case under consideration for, for example, four latent variables is $3630^2/10 \approx 1.3 \cdot 10^6$, which is two orders of magnitude less than for PLS + SPA. The third stage consists in selecting the spectral variables corresponding to the combination of windows, constructing on which the scmwiPLS model is characterized by the minimum value of $RMSEP$.

Let's consider in detail the application of scmwiPLS for Mn concentration calibration. Broadband PLS shows a minimum of $RMSEP = 0.12\%$ for four latent variables. The corresponding training error is $RMSEC = 0.14\%$. At the same time, $RPDC = 4.7$ for the training sampling, and $RPDP = 1.8$ for the test sampling, which shows the deterioration of this calibration quality indicator due to the narrowing of the considered range of Mn concentration in the test sampling.

In the scmwiPLS method, windows with five variables are used to select spectral variables in emission spectra when calibrating with four latent variables. The initial number of variables 3630 is a multiple of 5 and does not need to be reduced. Fig. 4 shows the dependence of RMSEP on the number of spectral windows taken into account in the PLS model. The minimum root mean square error corresponds to 19 windows or 95 variables. The position of the selected spectral variables in the emission spectrum of reference 123 is shown in Fig. 5. It can be seen that, in contrast to the C calibration, most of the selected spectral variables are in the region of intense emission lines. The same dependence is observed for the other chemical elements being calibrated, except for C, for which the scmwiPLS method, like the two previously applied methods, selects spectral variables outside the region of observed intense emission lines.

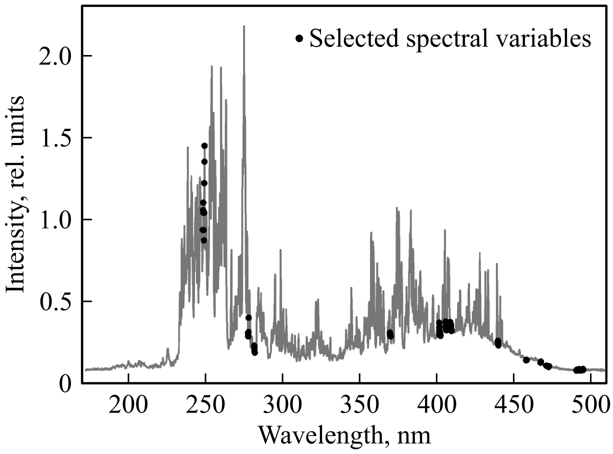

Figure 5. Emission spectrum of reference 123, which highlights the spectral variables used to calibrate the Mn concentration using the scmwiPLS method.

Table 3. Characteristics of multivariate scmwiPLS models for calibration of the concentration of six chemical elements in low-alloy steels by emission spectra

Element	Number of selected spectral variables in scmwiPLS	$RMSEC$, %	$RPDC$	$RMSEP$, %	$RPDP$
C	208	0.03	10.7	0.004	23.4
Mn	95	0.15	4.4	0.04	5.2
Si	208	0.11	2.5	0.003	20.7
Cr	153	0.13	2.9	0.04	3.1
Ni	240	0.10	2.5	0.01	4.8
Cu	325	0.05	3.9	0.04	1.4

Mn concentration calibration by scmwiPLS method is characterized by the following quality indicators: $RMSEC = 0.15\%$, $RMSEP = 0.04\%$, $RPDC = 4.4$ and $RPDP = 5.2$. Thus, the developed multivariate model is quantitative for both samplings.

Table 3 shows the characteristics of the scmwiPLS models for all six calibrated chemical elements.

Conclusion

It can be concluded that the use of methods for selecting spectral variables has improved the quality of multivariate models for calibrating the concentrations of the main technological impurities and alloying additives in low-alloy steels according to the data obtained using LIBS. Only for Cu in the concentration range 0.06–0.26%, the calibration created using the partial least squares method with the spectral variables selection by searching combination moving windows with a width that exceeds the number of latent variables per unity is qualitative. Similar calibration models for C in the concentration range from 0.13 to 0.43%, Mn (0.47–1.15%), Si (0.15–0.33%), Cr (0.09–0.43%) and Ni (0.05–0.25%) are quantitative.

Funding

The study was partially supported by the State Program for Scientific Research of the Republic of Belarus „Photonics and Electronics for Innovations“ under the state assignment 1.5.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] L. Quackatz, A. Griesche, T. Kannengiesser. Forces in Mechanics, **6**, 100063 (2022). DOI: 10.1016/j.finmec.2021.100063
[2] W. Huang, L. Guo, W. Kou, D. Zhang, Z. Hu, F. Chen, Y. Chu, W. Cheng. Microchemical Journal, **176**, 107190 (2022). DOI: 10.1016/j.microc.2022.107190
[3] M. Cui, H. Guo, Y. Chi, L. Tan, C. Yao, D. Zhang, Y. Deguchi. Spectrochimica Acta Part B: Atomic Spectroscopy, **191**, 106398 (2022). DOI: 10.1016/j.sab.2022.106398
[4] Y. Wei, R.S. Varanasi, T. Schwarz, L. Gomell, H. Zhao, D.J. Larson, B. Sun, G. Liu, H. Chen, D. Raabe, B. Gault. Patterns, **2** (2), 1 (2021). DOI: 10.1016/j.patter.2020.100192
[5] S. Grünberger, S. Eschlböck-Fuchs, J. Hofstadler, A. Pissenberger, H. Duchaczek, S. Trautner, J.D. Pedarnig. Spectrochimica Acta Part B: Atomic Spectroscopy, **169**, 1 (2020). DOI: 10.1016/j.sab.2020.105884
[6] M.W. Vaughan, P. Samimi, S.L. Gibbons, R.A. Abrahams, R.C. Harris, R.E. Barber, I. Karaman. Scripta Materialia, **184**, 63 (2020). DOI: 10.1016/j.scriptamat.2020.03.011
[7] Ch.J. Rao, S. Ningshen, J. Philip. Spectrochimica Acta Part B: Atomic Spectroscopy, **172**, 1 (2020). DOI: 10.1016/j.sab.2020.105973
[8] D. Syvilay, J. Guezenoc, B. Bousquet. Spectrochimica Acta Part B: Atomic Spectroscopy, **161**, 1 (2019). DOI: 10.1016/j.sab.2019.105696
[9] H. Kim, S.-H. Na, S.-H. Han, S. Jung, Y. Lee. Optics and Laser Technology, **112**, 117 (2019). DOI: 10.1016/j.optlastec.2018.11.002
[10] M. Cui, Y. Deguchi, Ch. Yao, Zh. Wang, S. Tanaka, D. Zhang. Spectrochimica Acta Part B: Atomic Spectroscopy, **167**, 1 (2020). DOI: 10.1016/j.sab.2020.105839
[11] W. Lee, J. Wu, Y. Lee, J. Sneddon. Applied Spectroscopy Reviews, **39** (1), 27 (2004). DOI: 10.1081/ASR-120028868
[12] N. Reinhard. Laser-induced breakdown spectroscopy: fundamentals and applications. (Springer, Berlin, Heidelberg, 2012).
[13] M. Markiewicz-Keszycka, X. Cama, M.P. Casado, Y. Dixit. Trends in Food Science & Technology, **65**, 80 (2017). DOI: 10.1016/j.tifs.2017.05.005
[14] N. Reinhard, B. Holger, B. Adriane, M. Kraushaar, I. Monch, P. Laszlo, S. Volker. Spectrochimica Acta Part B: Atomic Spectroscopy, **56** (6), 637 (2001). DOI: 10.1016/s0584-8547(01)00214-2

[15] J.L. Gottfried, F.C. De Lucia Jr.. Laser-Induced Breakdown Spectroscopy: Capabilities and Applications. Report ARL-TR-5238. (Aberdeen Proving Ground, MD, USA: Army Research Laboratory, 2010). DOI: 10.21236/ada528756

[16] V. Motto-Ros, D. Syvilay, L. Bassel, E. Negre, F. Trichard, F. Pelascini, J. El Haddad, A. Harhira, S. Moncayo, J. Picard, D. Devismes, B. Bousquet. Spectrochimica Acta Part B: Atomic Spectroscopy, **140**, 54 (2018). DOI: 10.1016/j.sab.2017.12.004

[17] Z. Wang, M. Sher Afgan, W. Gu, Y. Song, Y. Wang, Z. Hou, W. Song, Z. Li. TrAC Trends in Analytical Chemistry, **143**, 116385 (2021). DOI: 10.1016/j.trac.2021.116385

[18] American Society for Testing and Materials standard ASTM E 1655-05 „Standard Practices for Infrared Multivariate Quantitative Analysis“ (National standard of the Republic of Kazakhstan ACTM E 1655-2011 „Standard Practices for Infrared Multivariate Quantitative Analysis“)

[19] M.V. Belkov, D.A. Borisevich, K.Yu. Katsalap, M.A. Khodasevich. ZhPS, **88** 737 (2021).(in Russian).

[20] R.W. Kennard, L.A. Stone. Technometrics, **11**, 137 (1969). DOI: 10.1080/00401706.1969.10490666

[21] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin. Analytica Chimica Acta, **667**, 14 (2010). DOI: 10.1016/j.aca.2010.03.048

[22] S.F.C. Soares, A.A. Gomes, M.C.U. Araujo, A.R.G. Filho, R.K.H. Galvão. Trends in Analytical Chemistry, **42**, 84 (2013). DOI: 10.1016/j.trac.2012.09.006.

[23] M.A. Khodasevich, V.A. Aseev. Optika i spektroskopiya, 124 (5), 713 (in Russian). (2018). DOI: 10.21883/OS.2018.05.45958.22-18

[24] Y.P. Du, Y.Z. Liang, J.H. Jiang, R.J. Berry, Y. Ozaki. Analytica Chimica Acta, **501**, 183 (2004). DOI: 10.1016/j.aca.2003.09.041

[25] P. Geladi, B. Kowalski. Analytica Chimica Acta, **186**, 1 (1986). DOI: 10.1016/0003-2670(86)80028-9

[26] Y. Li, C.M. Altaner. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, **213**, 111 (2019). DOI: 10.1016/j.saa.2019.01.060

[27] R. Zornoza, C. Guerrero, J. Mataix-Solera, K.M. Scow, V. Arcenegui, J. Mataix-Beneyto. Soil Biology & Biochemistry, **40**, 1923 (2008). DOI: 10.1016/j.soilbio.2008.04.003