

Selection of Spectral Variables in Multivariate Calibration of C, Mn, Si, Cr, Ni, and Cu Concentrations in Low-Alloy Steels by Laser-Induced Breakdown Spectroscopy

M. V. Belkov^a, D. A. Borisevich^a, K. Yu. Katsalap^a, and M. A. Khodasevich^{a, *}

^a Institute of Physics, National Academy of Sciences of Belarus, Minsk, 220072 Belarus

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

Received July 5, 2022; revised July 5, 2022; accepted July 20, 2022

Abstract—Multivariate models for calibration of C, Mn, Si, Cr, Ni, and Cu concentrations from low-resolution emission spectra (190–440 nm, resolution 0.4 nm, spectral step 0.1 nm) by the least-squares method were developed in sets containing from 31 to 39 reference samples of low-alloy steels. Three methods of selection of spectral variables are considered, namely, method of ranking spectral variables by their correlation coefficient with the sought parameter, a successive projection algorithm, and an original modification of the method of searching combination moving window. The partial least-squares model with selection of spectral variables by searching combination moving window for C is quantitative (root-mean-square deviation 0.004%, residual deviation in the test sample 23.4 in the concentration range from 0.13 to 0.43%). The concentration calibrations are also quantitative for 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%). The calibration for Cu in the concentration range of 0.06–0.26% is qualitative (0.04% and 1.4).

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

DOI: 10.1134/S0030400X22100010

INTRODUCTION

In recent years, the use of laser-induced breakdown spectroscopy (LIBS) made it possible to considerably improve the characteristics of fixed and portable instruments for quantitative and qualitative analysis of samples [1–3]. LIBS is used for express analysis of the composition of various substances with minimal sample preparation or even without it, which is a considerable advantage over standard chemical methods. For example, LIBS was used in [1] for quantitative measurements of the chemical composition of welding joints of stainless steels with temporal and spatial resolution directly during welding. A possible solution of the problem of disclosure of various types of powder milk adulteration by using LIBS and machine learning methods is considered in [2]. These adulterations may cause serious indigestion of people. Work [3] is devoted to the use of LIBS for searching the methods of decreasing the errors in carbon calibration due to trace amounts of contaminants on the surfaces of low-alloy steels.

The steels and iron-based alloys are widely used almost in all human activity areas and hold a special place among the objects studied by LIBS. The presence of technological impurities, as well as doping

with chromium, manganese, and other chemical elements, determines their physical, chemical, and technological properties. The quantitative analysis methods for determining the concentrations of dopants and technological impurities are important for classification or sorting of steels. Usually, this analysis is performed using mass-spectroscopy [4] and optical emission spectroscopy using spark discharge [5], inductively coupled plasma [6], and glow-discharge [7]. The LIBS advantages are the possibility of rapid multivariate analysis in open air and relatively inexpensive instruments with accuracy sufficient for qualitative analysis. As the main disadvantage of LIBS, we should note the insufficient accuracy of quantitative measurements [8]. Nevertheless, numerous works are devoted not only to qualitative but also to quantitative LIBS applications (see, for example, [9–15]). Due to different experimental conditions (wavelength, pulse duration, laser beam energy and focusing, spectrometer's spectral range and resolution, delay time and interval of spectral measurements, the number of preliminary and measuring laser pulses, the number of accumulations, blowing of the object with gas), which considerably affect the measured spectra, LIBS is considered as a semiquantitative method [16]. Construction of

uni- or multivariate quantitative models with different preprocessing of spectra also leads to results with considerably different accuracies.

The spectrometers used for LIBS in compact portable and mobile systems usually have a low resolution. Therefore, due to a strong overlap of emission wings, the classical univariate approach to construction of the calibration dependence by the intensity of an isolated analytical line is hardly suitable for such spectrometers. In this case, wide use is made of multivariate calibration models [1, 17] covering the entire recorded spectrum. According to [18], calibration is a process used to create a model that relates two types of measured data. In the present work, we calibrate the concentration of main elements in low-alloy steels, i.e., create a mathematical model that relates the sought concentrations of chemical elements in a set of known reference samples of low-alloy steels to the spectral data obtained by LIBS.

EXPERIMENTAL

Previously, we solved the problem of calibration over the entire range of measured low-resolution emission spectra (190–440 nm, resolution 0.4 nm, spectral step 0.1 nm). The experimental setup and the measurement conditions are given in [19]. We studied 44 reference samples of low-alloy steels UG0d–UG7d, UG9d (Russia) and 51/1–58/1, 72–76, 101–103, 110–125 (IMZ, Poland); from this set, we used for calibration from 31 to 39 samples with different (non-coinciding) concentrations of C (in the range below 0.8%), Mn (2.0%), Si (1.2%), Cr (1.0%), Ni (0.8%), and Cu (0.5%).

METHODS AND RESULTS

The training and test sets approximately identical in the number of reference samples were formed according to the conventionally used Kennard–Stone algorithm [20], i.e., the first sample selected into the training set has a concentration closest to the studied range center and the concentration of each subsequent sample should be most distant from the already selected. In our case, this algorithm allows obtaining of more stable models in comparison with the uniform or random distribution of the training sample due to narrowing of the estimation intervals of concentrations of chemical elements in the training set. After normalization of the spectra to the intensity at the characteristic iron emission wavelength of 252.0609 nm, calibration models were developed using the least-squares method with the root-mean-square error of prediction (RMSEP) with respect to the corresponding reference values in the test sample being 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 increase the calibration accuracy, we partially took into account the requirements of [18] when form-

ing the training and test sets. Since the requirement for the minimum number of samples in the training (24) and test (20) sets were in total not satisfied, about 60% of samples form the training set and remaining 40% form the test set. This proportion is determined by the requirement of [18] that the number of samples for training multivariate models should be identical to the number of used latent variables multiplied by a factor of six. For the test sample, this factor is four. This problem was solved at the first stage of the work. Then, we applied to the full-range multivariate model three methods of selection of spectral variables, namely, ranking of spectral variables (RSV) [21] according to their correlation with the sought parameter, successive projections algorithm (SPA) [22], and an original modification [23] of searching combination moving window interval partial least squares (scmwPLS) [24]. Let us characterize each of these methods of selection of spectral variables and discuss the results obtained.

In the RSV method, all spectral variables are ranked in the order of decreasing coefficient of correlation with the calibrated concentration. Then, one variable with the minimal correlation is removed at each step and multivariate modeling is performed by the least-squares method [25] with determination of the optimal number of latent variables. The spectral variables are selected by the minimal root-mean-square deviation of the estimated concentration of the sought element in the test set samples from the reference values. If the selection of variables is restricted by the correlation coefficient value, this RSV method modification is called the significance multivariate correlation (SMC) method [26]. This method is characterized by introduction of arbitrariness of the researcher in the choice of this restriction, which is absent in the RSV method.

The characteristics of the full-range partial least squares (PLS) and PLS + RSV models are compared in Table 1.

One can see that the change in the proportion of the numbers of samples in the training and test sets from 1:1 to 3:2 leads to slight changes in the RMSEP only in the case of full-range calibration of Si and Ni concentrations (by 0.02% and 0.01%, respectively). This confirms the stability of multivariate PLS models to changes in the sizes of sets. Comparison of the RMSEP of the PLS and PLS + RSV models shows that the use of the correlation method of selection of spectral variables is insufficiently effective in the cases considered. The quality of the calibration models for Cr and Cu did not change because the number of variables selected from 3630 spectral lines was 3629 and 3625, respectively. The calibration quality for the other four elements improved, but insignificantly. Let us illustrate the obtained results of multivariate calibration of the C concentration by the PLS + RSV method. Figure 1 presents the dependence of RMSEP on the number of spectral variables eliminated from

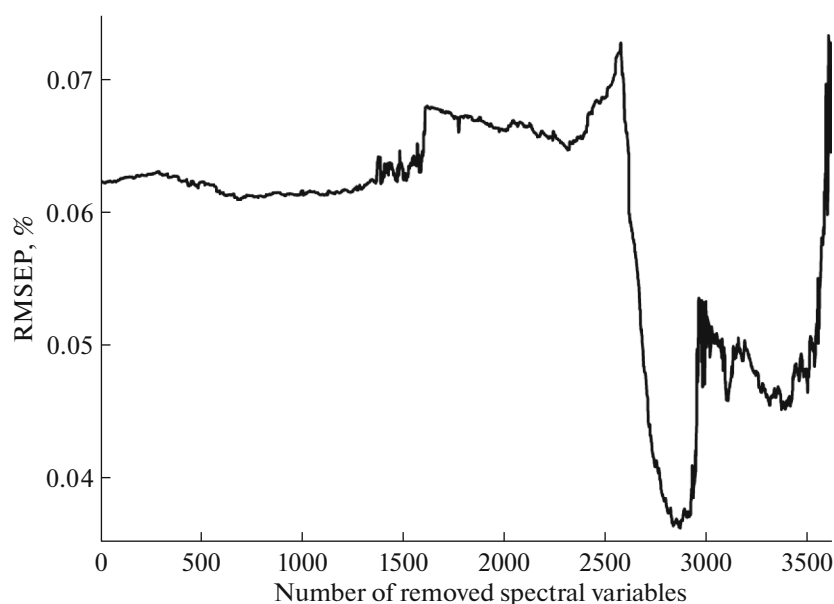
Table 1. Characteristics of multivariate PLS and PLS + RSV models for calibration of concentrations of six chemical elements in low-alloy steels by emission spectra

Element	Number of samples			Number of latent variables in the full-range PLS	RMSE in test in full-range PLS, %	Number of selected spectral variables for the PLS + RSV method	Number of spectral variables for the PLS + RSV method	RMSE in test (%) for the PLS + RSV method
	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

the model, which are ranked according to their correlation coefficient with the C concentration in the samples. The minimal RMSEP is achieved when the model includes 757 spectral variables, which allows one to obtain the calibration dependence shown in Fig. 2 with the use of four latent variables in the PLS method. The RMSEP is 0.04%, and the residual ratio of performance to deviation (RPD) in the test set (ratio of the root-mean square deviation of a parameter in the set to the root-mean-square deviation of the prediction from reference value) is 2.7. It is the best model among the models constructed for six considered elements by the PLS + RSV method, but it is only semiquantitative ($2.5 < \text{RPD} < 3$) [27]. Attention is drawn to the disposition of selected spectral variables

shown in Fig. 3 for the emission spectrum of reference sample 123. One can see that only one spectral variable (252.06 nm) lies in the region of intense emission lines, which may provide useful information for calibration. The other selected spectral variables lie at the edges of the measured spectra, which, as expected, provide little information.

The second method we used for selection of spectral variables for multivariate calibration is the successive projection algorithm (SPA) [22]. At the first SPA stage, for each of available 3630 spectral variables, an ordered sequence of all the other variables is formed. In these sequences, the second variable will be chosen according to the maximum projection to the subspace orthogonal to that of the first variable. This procedure

**Fig. 1.** Root-mean-square deviation of the estimated C concentration from reference values in the test set of samples of low-alloy steels as a function of the number of spectral variables ranked according to the correlation coefficient and removed from the partial least squares model.

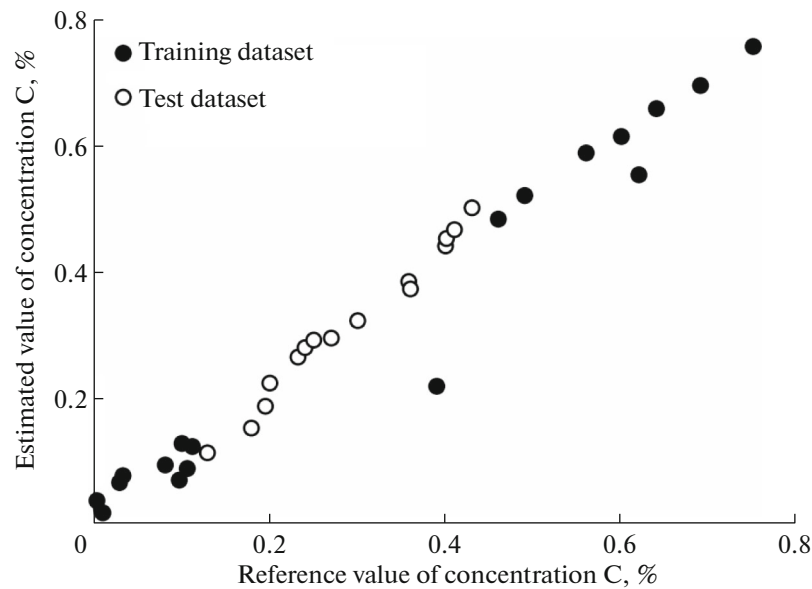


Fig. 2. Relation between the estimated C concentrations and reference values in the case of the multivariate model constructed by the partial least squares method including 757 spectral variables with the best correlation with the calibrated parameter.

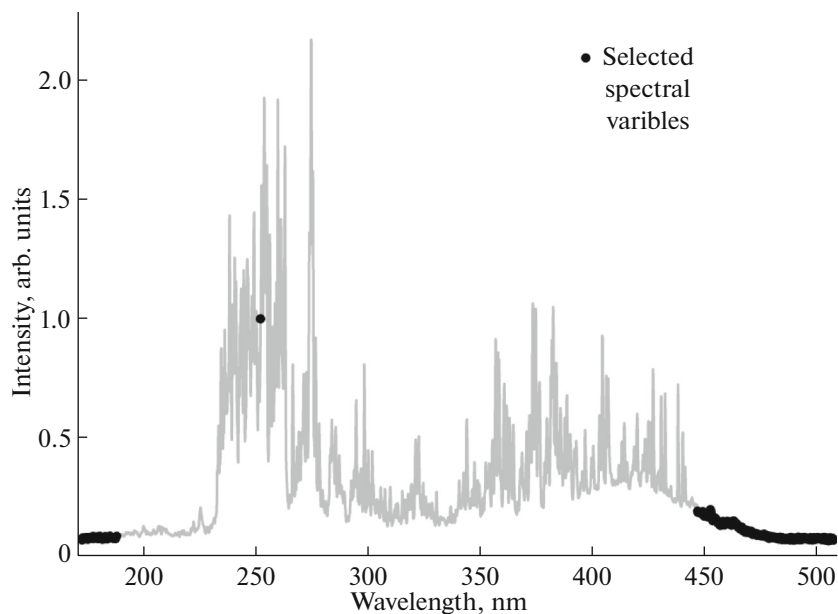


Fig. 3. Emission spectrum of reference sample 123 with indicated spectral variables used for calibration of carbon concentration by the PLS + RSV method.

is continued until all the measured spectral variables are included into the sequence. At the second SPA stage, based on the increasing sets of elements in each sequence, PLS models are constructed with selection of the optimal number of latent variables. In our case of restriction of the number of latent variables by ten, the number of these models is $3630^2 \times 10 \approx 1.3 \times 10^8$. At the third stage, by the minimum RMSEP, one determines the best multivariate model and, corre-

spondingly, the sought set of spectral variables providing the minimum calibration error. Table 2 present the examples of the best multivariate models for the PLS + SPA method.

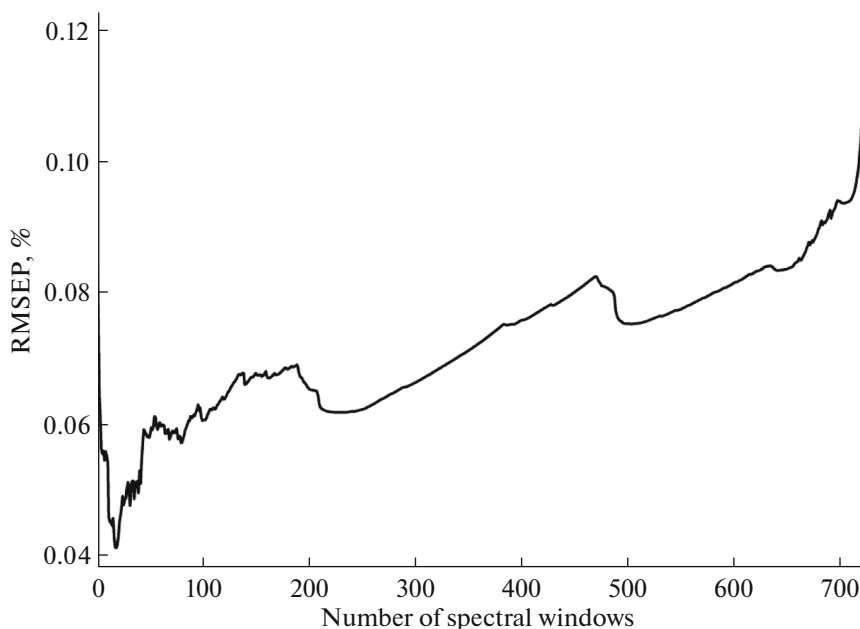
Comparison of Tables 1 and 2 shows that the quality of the PLS + SPA calibration models is better than that of PLS + RSV for Mn, Si, Cr, and Cu, but worse for C. For Ni, the RMSEP is the same in both cases.

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

Element	Number of selected spectral variables for the PLS + SPA method	Number of latent variables in the PLS + SPA model	RMSE in test (%) for the PLS + SPA method
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

The third method used by us for selecting spectral variables was the scmwPLS method. This method deals with spectral intervals (windows) of particular widths rather than with individual spectral variables. The original modification of the scmwPLS method [23] contains three stages. The first stage consists in the construction of a full-range multivariate PLS model and determination of the optimal number n of latent variables by the minimum RMSEP value. The width of the spectral windows, in which the number of spectral variables exceeds n by unity, is fixed at the second stage. This condition minimizes the window width and retains the possibility of selecting latent variables even in one window. Then, the first window is shifted by one spectral variable per step and a multivariate model, which is also characterized by the RMSEP value, is constructed at each step. As the window reaches the edge of the measured spectral range,

the optimal position of this window is determined by the minimum RMSEP and fixed. This procedure is repeated with the second spectral window. The modeling is performed over the spectral variables belonging to the first fixed and the second moving spectral windows. Each subsequent window adds to the model the $n + 1$ spectral variables to take into account all measured variables. For correct scmwPLS operation, it is necessary to preliminary reduce the total number of spectral variables to a value divisible by $n + 1$. The total number of multivariate scmwPLS models in the considered case for, e.g., four latent variables, is $3630^2/10 \approx 1.3 \times 10^6$, which is two orders of magnitude smaller than for PLS + SPA. The third stage consists in the selection of spectral variables corresponding to a combination of windows such that the scmwPLS model based on them has the minimal RMSEP.

**Fig. 4.** Dependence of the root-mean-square error of prediction (RMSEP) of the Mn concentration in the test set of low-alloy steels by emission spectra with the use of the partial least-squares model with selection of spectral variables by searching combination moving window on the number of spectral windows.

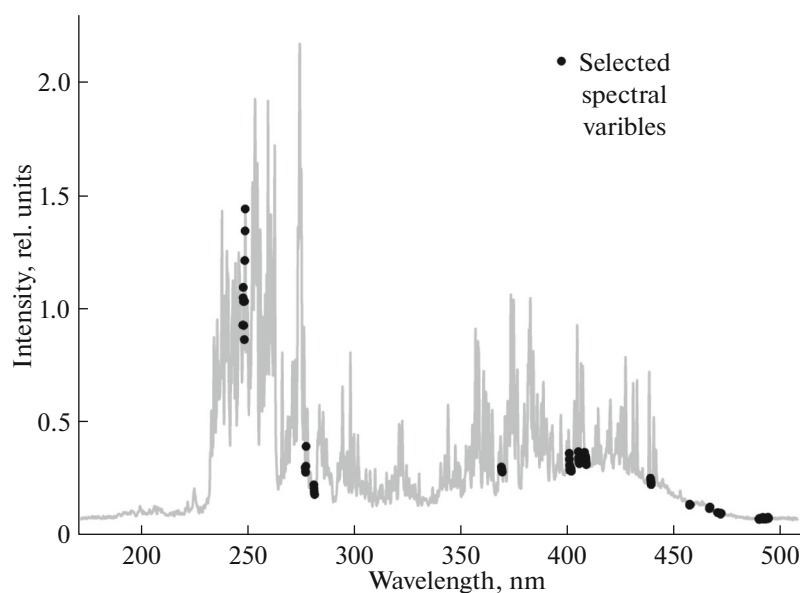


Fig. 5. Emission spectrum of reference sample 123 with indicated spectral variables used for calibration of the Mn concentration by the scmwiPLS method.

Let us consider in detail the use of scmwiPLS for calibration of the Mn concentration. The full-range PLS demonstrates the minimum RMSEP = 0.12% for four latent variables. The corresponding error in training is RMSEC = 0.14%. For the training set, we have RPDC = 4.7, and, for the test set, RPDP = 1.8, which indicates degradation of this calibration quality parameter due to narrowing of the considered range of Mn concentration in the test set.

In the scmwiPLS method, the selection of spectral variables in the emission spectra for calibration over four latent variables is performed using windows with a width of five variables. The initial number of variables (3630) is divisible by five and should not be decreased. Figure 4 shows the dependence of RMSEP on the number of spectral windows taken into account in the PLS model. The minimum root-mean-square estimation error corresponds to 19 windows or 95 variables. The positions of the selected spectral variables

on the emission spectrum of reference sample 123 are shown in Fig. 5. One can see that, in contrast to the C calibration, most of the selected spectral variables lie in the region of intense emission lines. The same dependence was also observed for the other calibrated chemical elements except for C, for which the scmwiPLS method, as well as the two previously used methods, selects spectral variables outside the region of observed intense emission lines.

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

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

Table 3. Characteristics of the multivariate scmwiPLS models for calibration of concentrations 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

CONCLUSIONS

The use of the spectral variable selection methods makes it possible to improve the quality of multivariate models of calibration of concentrations of main technological impurities and dopants in low-alloy steels by the data of laser-induced breakdown spectroscopy. The calibration based on the partial least squares method with selection of spectral variables by searching combination moving window with a width exceeding the number of latent variables by unity is qualitative only for Cu in the concentration range of 0.06–0.26%. 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

This work was partially supported by the State Program of Scientific Research of the Republic of Belarus “Photonics and Electronics for Innovations” (order no. 1.5).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- L. Quackatz, A. Griesche, and T. Kannengiesser, *Forces Mech.* **6**, 100063 (2022).
<https://doi.org/10.1016/j.finmec.2021.100063>
- W. Huang, L. Guo, W. Kou, D. Zhang, Z. Hu, F. Chen, Y. Chu, and W. Cheng, *Microchem. J.* **176** (2022).
<https://doi.org/10.1016/j.microc.2022.107190>
- M. Cui, H. Guo, Y. Chi, L. Tan, C. Yao, D. Zhang, and Y. Deguchi, *Spectrochim. Acta, Part B* **191** (2022).
<https://doi.org/10.1016/j.sab.2022.106398>
- Y. Wei, R. S. Varanasi, T. Schwarz, L. Gomell, H. Zhao, D. J. Larson, B. Sun, G. Liu, H. Chen, D. Raabe, and B. Gault, *Patterns* **2** (2), 1 (2021).
<https://doi.org/10.1016/j.patter.2020.100192>
- S. Grünberger, S. Eschlböck-Fuchs, J. Hofstadler, A. Pissenberger, H. Duchaczek, S. Trautner, and J. D. Pedarnig, *Spectrochim. Acta, Part B* **169**, 1 (2020).
<https://doi.org/10.1016/j.sab.2020.105884>
- M. W. Vaughan, P. Samimi, S. L. Gibbons, R. A. Abrahams, R. C. Harris, R. E. Barber, and I. Karaman, *Scr. Mater.* **184**, 63 (2020).
<https://doi.org/10.1016/j.scriptamat.2020.03.011>
- Ch. J. Rao, S. Ningshen, and J. Philip, *Spectrochim. Acta, Part B* **172**, 1 (2020).
<https://doi.org/10.1016/j.sab.2020.105973>
- D. Syvilay, J. Guezenoc, and B. Bousquet, *Spectrochim. Acta, Part B* **161**, 1 (2019).
<https://doi.org/10.1016/j.sab.2019.105696>
- H. Kim, S.-H. Na, S.-H. Han, S. Jung, and Y. Lee, *Opt. Laser Technol.* **112**, 117 (2019).
<https://doi.org/10.1016/j.optlastec.2018.11.002>
- M. Cui, Y. Deguchi, Ch. Yao, Zh. Wang, S. Tanaka, and D. Zhang, *Spectrochim. Acta, Part B* **167**, 1 (2020).
<https://doi.org/10.1016/j.sab.2020.105839>
- W. Lee, J. Wu, Y. Lee, and J. Sneddon, *Appl. Spectrosc. Rev.* **39**, 27 (2004).
<https://doi.org/10.1081/ASR-120028868>
- N. Reinhard, *Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications* (Springer, Berlin, 2012).
- M. Markiewicz-Keszyccka, X. Cama, M. P. Casado, and Y. Dixit, *Trends Food Sci. Technol.* **65**, 80 (2017).
<https://doi.org/10.1016/j.tifs.2017.05.005>
- N. Reinhard, B. Holger, B. Adriane, M. Kraushaar, I. Monch, P. Laszlo, and S. Volker, *Spectrochim. Acta, Part B* **56**, 637 (2001).
[https://doi.org/10.1016/s0584-8547\(01\)00214-2](https://doi.org/10.1016/s0584-8547(01)00214-2)
- J. L. Gottfried and F. C. de Lucia, Jr., Final Report ADA528756 (2010).
<https://doi.org/10.21236/ada528756>
- V. Motto-Ros, D. Syvilay, L. Bassel, E. Negre, F. Trichard, F. Pelascini, J. el Haddad, A. Harhira, S. Moncayo, J. Picard, D. Devismes, and B. Bousquet, *Spectrochim. Acta, Part B* **140**, 54 (2018).
<https://doi.org/10.1016/j.sab.2017.12.004>
- Z. Wang, Sher M. Afgan, W. Gu, Y. Song, Y. Wang, Z. Hou, W. Song, and Z. Li, *Trends Anal. Chem.* **143** (2021).
<https://doi.org/10.1016/j.trac.2021.116385>
- ASTM E 1655-05: Standard Practices for Infrared Multivariate Quantitative Analysis.
- M. V. Bel'kov, D. A. Borisevich, K. Yu. Catsalap, and M. A. Khodasevich, *J. Appl. Spectrosc.* **88**, 970 (2021).
- R. W. Kennard and L. A. Stone, *Technometrics* **11**, 137 (1969).
<https://doi.org/10.1080/00401706.1969.10490666>
- Z. Xiaobo, Z. Jiewen, M. J. W. Povey, M. Holmes, and M. Hanpin, *Anal. Chim. Acta* **667**, 14 (2010).
<https://doi.org/10.1016/j.aca.2010.03.048>
- S. F. C. Soares, A. A. Gomes, M. C. U. Araujo, A. R. G. Filho, and R. K. H. Galvão, *Trends Anal. Chem.* **42**, 84 (2013).
<https://doi.org/10.1016/j.trac.2012.09.006>
- M. A. Khodasevich and V. A. Aseev, *Opt. Spectrosc.* **124**, 748 (2018).
<https://doi.org/10.1134/S0030400X18050089>
- Y. P. Du, Y. Z. Liang, J. H. Jiang, R. J. Berry, and Y. Ozaki, *Anal. Chim. Acta* **501**, 183 (2004).
<https://doi.org/10.1016/j.aca.2003.09.041>
- P. Geladi and B. Kowalski, *Anal. Chim. Acta* **186**, 1 (1986).
[https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
- Y. Li and C. M. Altaner, *Spectrochim. Acta, Part A* **213**, 111 (2019).
<https://doi.org/10.1016/j.saa.2019.01.060>
- R. Zornoza, C. Guerrero, J. Mataix-Solera, K. M. Scow, V. Arcenegui, and J. Mataix-Beneyto, *Soil Biol. Biochem.* **40**, 1923 (2008).
<https://doi.org/10.1016/j.soilbio.2008.04.003>

Translated by M. Basieva