Spectrochimica Acta Part B 148 (2018) 65–82

Contents lists available at ScienceDirect

Spectrochimica Acta Part B

journal homepage: www.elsevier.com/locate/sab

Review article

On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis, a review

Pavel Pořízka[a,b,c,*], Jakub Klus[a,b], Erik Képeš[a], David Prochazka[a,b], David W. Hahn[c], Jozef Kaiser[a,b]

[a] Central European Institute of Technology, Brno University of Technology, Purkyňova 656/123, CZ-61200 Brno, Czech Republic
[b] AtomTrace a.s., Kolejní 9, CZ-61200 Brno, Czech Republic
[c] Department of Mechanical and Aerospace Engineering, University of Florida, 939 Sweetwater Drive, FL-32611 Gainesville, FL, USA

ABSTRACT

An implementation of a fast, robust, and effective algorithm is inevitable in modern multivariate data analysis (MVDA). The principal component analysis (PCA) algorithm is becoming popular not only in the spectroscopic community because it complies with the qualities mentioned above. PCA is, therefore, often used for the processing of detected multivariate signal (characteristic spectra). Over the past decade, PCA has been adopted by the Laser-Induced Breakdown Spectroscopy (LIBS) community and the number of scientific articles referring to PCA steadily increases. The interest in PCA is not caused only by the basic need to obtain a fast data visualization on a lower dimensional scale and to inspect the most prominent variables. Most recently, PCA has also been applied to yield unconventional data analyses, *i.e.* processing of large scale LIBS maps. However, a rapid development of LIBS-related instrumentation and applications has led to some non-uniform methodologies in the implementation and utilization of MVDA, including PCA. Thus, in this work, we critically assess and elaborate on the approaches to utilize PCA in LIBS data processing. The aim of this article is also to derive some implications and to suggest advice in data preprocessing, visualization, dimensionality reduction, model building, classification, quantification and non-conventional multivariate mapping. This review reflects also other MVDA algorithms than PCA and consequently, presented conclusions and recommendations can be generalized.

1. Introduction

Sample characterization using Laser-Induced Breakdown Spectroscopy (LIBS) technique has been dynamically advancing in recent years. The parameters of conventionally utilized analytical instrumentation (lasers, spectrometers, and detectors) are being constantly improved. Moreover, the complicated or basic lab-built systems have been transformed to the sophisticated and commercially available systems, which enable an effortless and fast spectroscopic analysis. Contemporary state-of-the-art LIBS systems are capable of a high-end performance analysis (repetition rate, resolution, sensitivity). The high-end performance of LIBS is in certain cases superior to the performance of its analytical counterparts or reference techniques, such as Laser-Ablation Inductively Coupled Plasma (LA-ICP) based techniques, X-ray Fluorescence (XRF), *etc.*

LIBS is a well-established technique in many different applications, such as biology [1–4], geology [5], and industry [6]. The reason is the simplicity and robustness of the LIBS instrumentation together with its capability of a fast-throughput multielemental analysis. Its potential has been repeatedly demonstrated by its high-end lab-based [7], *in-situ* and stand-off [8,9], and even extraterrestrial [10,11] utilization.

LIBS is one of the atomic emission spectroscopic techniques [6,12,13] based on the laser ablation sampling. Thorough articles were published with the aim to review the basic theory of the Laser-Induced Plasma (LIP) formation [14–16] and LIBS in general [17–20].

The introduction covering the basic theory about LIBS technique was brief because this review article targets namely the aspects of data processing. The reader should follow referenced books and review articles for more detailed background of LIBS theory prior to any further data processing through MVDA algorithms. As it was emphasized by Hahn and Omenetto [17]: "*advanced chemometric algorithms must be used with knowledge of what emission features (e.g. atomic or molecular emission peaks) are providing the associated discrimination.*"

A typical LIBS system is able to provide a high number of measurements (given by its repetition rate) when each measurement is described by a high number of variables (especially in the case of

* Corresponding author at: Central European Institute of Technology, Brno University of Technology, Purkyňova 656/123, CZ-61200 Brno, Czech Republic.
E-mail address: pavel.porizka@ceitec.vutbr.cz (P. Pořízka).

https://doi.org/10.1016/j.sab.2018.05.030
Received 16 March 2018; Received in revised form 23 May 2018; Accepted 31 May 2018
Available online 03 June 2018
0584-8547/

echelle spectrometers). Note that the high repetition rate systems are mentioned for their leading edge in the LIBS applications, however, obtaining large number of measurements is not strictly related to LIBS systems with high repetition rate. The collected LIP spectrum is rich in information and represents the sample from which it originated, *i.e.* the chemical/spectral fingerprint of the sample [21,22]. The processing of large scale datasets is a demanding task which can be accomplished by using the so-called multivariate data analysis (MVDA; often related to as chemometric, exploratory data analysis or pattern recognition). It is noteworthy that unique LIP spectra are strongly affected by the matrix effect [19] which requires special attention when it comes to the conventional univariate calibration and quantitative analysis. On the contrary, the relation to the sample matrix enables a classification of samples according to their spectral fingerprints using simple MVDA algorithms. When processing large datasets, there are two more requirements to be met, namely, to process the data in the least possible time and in the most efficient manner. Efficiency can be measured by the conservation of variance during the dimensionality reduction, the sensitivity to outliers and the specificity to discriminate between individual matrices of analytes.

MVDA algorithms are massively spread throughout the LIBS community and are used in a number of applications. It may be stated that the future of the LIBS data analysis lies in the implementation of MVDA algorithms. The use of multivariate algorithms for processing of spectroscopic data has already been well documented [23–26]. Moreover, several review articles [5,17,27,28] dealt solely with the multivariate processing of LIBS data. A full chapter in the LIBS book by Cremers and Radziemski [12] was also dedicated to this topic. Based on the literature survey, the most popular MVDA algorithm in the LIBS community is the Principal Component Analysis (PCA). This simple linear algorithm provides powerful means of data visualization and pattern recognition on a lower-dimensional scale.

Based on our thorough literature research, the methodological approaches in the processing of LIBS data through MVDA algorithms significantly differ. This is given *i)* by the needs of a particular application, *ii)* by the uniqueness of the data acquisition and data size, *iii)* by the data topology, *iv)* by the variety of MVDA algorithms and also *v)* by the internal methodology of each research group. Consequently, there is not a unified approach and it might not exist in the future. Moreover, a wide range of MVDA algorithms together with the available software for the processing of data creates an option to perform a reachable and easy-to-use analysis. This might lead to the misguided implementation of these algorithms and software, *i.e.* when their use leads to aesthetic improvement of low-quality data (high fluctuation, low sensitivity, *etc.*) [29]. Nevertheless, it has to be stressed that a stable and optimized analytical system providing a reproducible high-performance analysis (high-quality data) should be the cornerstone of any experimental work. The same is valid for the understanding of the theory of *i)* LIBS (*e.g.* laser-ablation and plasma dynamics and its properties) and *ii)* MVDA algorithms and their considerate and judicious implementation in the data analysis process [17].

In this work we bring a summary of the most common approaches in the implementation of PCA in LIBS data analysis for: low-dimensional visualization, clustering, outliers filtering, variable selection, quantification, classification, and non-conventional multivariate mapping. Additionally, general suggestions for the data preprocessing and the model building, as well as a comparison with the performance of other MVDA counterparts, are given.

2. Multivariate data analysis

2.1. Data preprocessing

Prior to an implementation of any MVDA algorithm such as PCA and its variations, it is strongly advised to preprocess the obtained data [28,30]. Detected multivariate signal in its raw state is burdened with unwanted background signal, fluctuation in the experimental parameters, *etc.* It has to be kept in mind that the data structure is changing during the data handling. This leads to consecutive changes in the performance of MVDA algorithm applied to the final data [31]. In general, there is a whole list of data preprocessing algorithms that should be considered prior to any MVDA. The order of the implementation of individual preprocessing steps is not given, therefore we propose the following one based on our experience and data processing algorithms found throughout the literature research. For the sake of briefness, individual steps are reviewed only tangentially. References to relevant, more detailed literature are given.

First of all, we advise to collect a high number of representative spectra per each sample to get robust statistics. Obtained data (the sample from the parent distribution) are organized into the data matrix $\boldsymbol{X}$ with n rows (n is the total number of spectra from all investigated samples) and p columns (variables related to wavelengths or even processed intensities of spectral lines). Harrington [32] also suggests checking the structure of the matrix and its rank.

Background subtraction mitigates the contribution from the background noise and any non-gated background continuum radiation. Van Veen and de Loos-Vollebregt [33] reviewed and compared several algorithms for the background correction of optical emission signal in ICP applications. Smoothing the spectrum is a possible way of correction [34]. Both preprocessing steps truncate certain amount of information that could be sacrificed for the sake of improved performance [35]. Most often, time-resolved detectors are used in order to avoid unwanted background from the continuum emission. Contrary to that, Myakalwar et al. [36] proved that it is possible to classify samples based on non-gated spectra detected using low-cost Czerny-Turner spectrometers and time-integrated detectors.

Row-wise spectra normalization, *i.e.* internal standardization [37–39] might reduce the influence of any potential fluctuations and dependence on the sample matrix. Most common ways of internal standardization (to matrix line intensity, to total spectral intensity, *etc.*) are described and compared in the aforementioned publications [40]. However, Colao et al. [41], with reference to Sirven et al. [42], suggested avoiding the utilization of matrix element for internal standardization, especially in cases when the content of the matrix element significantly changes throughout the sample set. Sirven et al. [43] proposed to use the background intensity as an alternative to the conventionally used standardization procedures. This approach was also applied in other works [41,44]. To mention another possibility for signal normalization, Death et al. [45] normalized spectra to regions of minimal variance. By contrast, they used the total spectral intensity in their further work [46].

The row-wise standardization is often followed by the column-wise mean centering and scaling. Column-wise mean centering is required to fulfill the underlying mathematical assumptions used to derive certain MVDA algorithms. In that case, mean value of each column (variable) is subtracted and thus moving the center of mass of data points (so-called scores) towards the beginning of a coordinate system. In the column-wise scaling, each variable is divided by its standard deviation and thus provides a sort of standardization of variables, *i.e.* the significance of variables is then unified. The process of mean centering and scaling is also referred to as standard normal variate (SNV). The whole process assumes that data follow the Gaussian distribution. However, this is not the case of LIBS data where data are also found under extreme value distribution [47,48]. Doucet et al. [49] compared impact of mean-centering and scaling on the performance of the principal component regression (PCR) and the partial least squares regression (PLSR) quantification. Porizka et al. [39] investigated the impact of column-wise mean centering and scaling in the case of steel, Al alloys and sedimentary ores data. They concluded that column-wise scaling should be avoided in the case of raw, unprocessed spectra that include background noise.

Detected spectra contain raw information described as the intensity

versus wavelength. In the most ideal case, the individual spectral lines should be assigned to elements and their relative intensities should be quantitatively processed prior to any exploratory data analysis. That leads to the significantly lower dimension of the variables based on the sample composition (*i.e.* content of individual analytes). The intensity estimation of the spectral line may be done in several ways [6,12,13]. Deriving the intensity of the analyte is more convenient from the perspective of further processing of the data matrix **X** by MVDA algorithms, *i.e.* the lower the rank of **X**, the faster the computation response. This approach is not so straightforward due to the necessary calibration of the system to individual analytes and related sample matrices. To make the preprocessing simple and convenient, the spectra are mostly kept in their raw state when introduced to MVDA, as it is evident in many publications, see Section 3.3.

Spectral lines are usually manually assigned by comparing the wavelengths to spectral libraries [50,51]. Those libraries contain a high number of spectral lines and the decision making is mostly up to the researcher who has an incomplete list of parameters (ionization, energy levels, Einstein coefficients or degeneracy levels). Apart from that, libraries specially dedicated to the LIBS purposes with limited selected spectral lines also exist [52,53]. The automatic approach in line assignment was also demonstrated by Amato et al. [54].

To deliver a complete set of steps in the data processing, we introduce also a brief discussion on the subsetting of the data set and an estimation of figure of merits. This is, however, not directly related to the implementation of PCA itself, but leads to PCA-based algorithms used in the classification (Soft Independent Modelling of Class Analogies - SIMCA) and quantification (Principal Component Regression - PCR). In the process of quantitative analysis and classification, the model is built using a training subset of the original data. Then the robustness of this model is cross-validated by a testing subset of data, usually complementary to the training subset. The dissection of the original data set into the subsets demands special attention. On one hand, the data set should be split into the set for modelling and for testing as suggested by Gottfried in the chapter dedicated to chemometrics [12], *i.e.* the “split-sample design” [32]. On the other hand, El Haddad [28] recommends to use another independent subset of original data - a validation subset - for the validation of the model in order to avoid its under- or over-fitting. Naturally, a model performs better when similar data are in both the training and the test sets. Nevertheless, Anderson et al. [55] avoided the over-optimistic assessment of results by assigning the similar spectra only to either the training or the test set. They tested five different ways to select the training set. The testing part should be followed by an estimation of figures of merit.

Porizka et al. [56] repeated the random selection of the model and test data subsets and showed slight differences in resulting figures of merit. This study supported the urgent need for an elaborate selection of data in order to cover the natural fluctuation within the original dataset. The figures of merit serve mainly as a metric for comparison of LIBS systems and approaches suggested by individual research groups. Individual figures of merit are summarized elsewhere [17,28]. The original body of work by Voigtman [57] should be studied prior to the estimation of LODs.

2.2. Principal component analysis

If we consider investigating *p*-dimensional space by the 2D projections of pairs of variables then we will have to depict $\frac{1}{2}p(p-1)$ scatter plots. However, that is a tedious and misleading process. As a response to this, PCA was introduced by Pearson in 1901 [58] with the aim to find lines and planes that could fit the set of points in the *p*-dimensional space. Later on, in 1933, Hotelling [59] derived the algebraic form of PCA similar to the one of Factor Analysis (FA). In both cases, PCA can be generalized to an algorithm as it is known today, *i.e.* visualizing complex multivariate data on a low-dimensional scale. Moreover, the suggested dimensionality reduction retains the most of the variance carried by the original data. From that essence, PCA is used as an effective tool for the exploratory data analysis and pattern recognition. PCA provides a dimensionality reduction of the original data set by the generalization of the original variance. This is done by a transformation of the original high-dimensional space (wavelength as variables) into a smaller set of independent variables, *i.e.* principal components (PCs). The data projection by PCA concerns only the variances between the objects and their structure and it is unsupervised, *i.e.* there is no direct relation to, for instance, a sample class or the content of investigated analyte. From the geometric point of view [60], there is a line (the first principal component) going through the center of gravity of *n* points in the *p*-dimensional space and carrying the information of the highest variance within the respective data set.

The PCA algorithm is based on the covariance or correlation matrix, depending on the utilized algorithm, and therefore the data should satisfy certain assumptions [61]. Two main assumptions are *i)* **linearity**: all variables should be linearly correlated and *ii)* **normal distribution**: each variable should follow the normal distribution. However, both main assumption are not met in the LIBS experiment. The linearity of spectral response (intensity of analytical line *versus* analyte content) breaks at certain point. This issue is, for instance, well described in the review by Hahn and Omenetto [17]. The normality of data distribution was disproved by Klus et al. [48]. However, this fact is not considered throughout the LIBS literature. Thus, we will proceed with our review regardless of validity of mentioned assumptions.

PCA essentially provides three outputs: *i)* variance, *ii)* loadings and *iii)* scores. It is possible to create a PCA model of the original data matrix in several ways: the power method, Singular Value Decomposition (SVD), Nonlinear Iterative Partial Least Squares (NIPALS), *etc*. To overcome the problems of non-linear data, the kernel [62] or robust [63,64] PCA algorithms may be used, showing improved figures of merit in comparison with typical PCA. Other algorithms, for instance the Discriminant Function Analysis (DFA) [65–67], Independent Component Analysis (ICA) [68] and Self-Organizing Maps (or Kohonen Maps) [69] present an alternative to PCA and might be applied for the projection and visualization on a low-dimensional scale. Topological Data Analysis (TDA) is also promising in the data analysis, yet, to the best of our knowledge, still not applied for the visualization of LIBS data.

2.2.1. Variance and selecting the number of PCs

The data points are distributed in the *k*-dimensional (*k* being the number of constructed PCs) space and the data set may be described by a parameter called variance [70]. Considering the sample of points from a given population, the variance is given as a sum of standard deviations from the sample mean divided by the total number of objects minus one. Thus, as it is summarized by Kendal [60], the sum of eigenvalues may be estimated as the sum of squares of the distances of all the objects (data points, scores) from their mean. This sum may be assigned *S*, then the *i*th PC carries a portion of λ_i/S of the original data set's total variation.

PCs are arranged in descending order with respect to the variation that they carry. Therefore, only a first few of them are capable to satisfactorily describe the original data set. However, the judicious estimation of the optimal number of PCs depends on many factors, such as the structure and topology of the original data, the application in which they are going to be used and mainly the experience and expertise of the researcher. For that reason, the records of the number of PCs in published articles differ. Most often the optimal number is found to be in units of PCs but also extreme cases in tens of PCs exist, which is rather doubtful. In general, finding the optimal number of PCs leads to the tradeoff between the robustness and over-training of the model.

There are several ways where to cut-off the non-informative PCs [71]. The first one is the *variance explained criteria* [72] which gives a threshold to cumulative variance. Most often, the threshold is set to 90% or 95%. This threshold is set by the operator and can be varied at

his will. The *Cattell's scree plot* [73] depicts the explained variance *versus* the PCs. The plot is in descending manner and the cut-off is advised to be done in the elbow of the diagram, *i.e.* in the place where the steep part of the diagram levels-off and flattens. The *Kaiser's criterion* [74] dictates to cut-off PCs with the eigenvalue lower than 1. This is enabled by the fact that the average of all eigenvalues is equal to unity.

It is important that the loadings of retained PCs are investigated after the cut-off in order to check the information that they carry. It is reasonable to expect that the loadings of higher orders' PCs will describe only noise. Defernez and Kemsley [75] stress a potential misuse of MVDA algorithms by the overfitting of the model. In other words, the higher the number of components the better the MVDA analysis. But from a certain number of components the model gets over-trained and fails in predictions of unseen data. It has higher, yet illusive performance, giving the results that may be deceptive.

2.2.2. Loadings

The relation between the newly constructed PCs and the original data are represented by loadings values. Geometrically, the loadings describe a projection of the original space in the direction of the highest variation. Thus, the composition of each PC loading describes the most significant lines contributing to each latent variable. This information is fundamentally important in the line selection and related dimensionality reduction, see Section 3.3.

The loadings matrix **P** has (m, p) dimensions. Loadings are usually plotted against the wavelengths. But in some articles [44,76,77] loadings were presented in the form of cross-plots, showing the relation of each wavelength variable in respective PCs.

2.2.3. Scores

When a PCA model of the original data is constructed, a certain number of PCs is selected giving a k-dimensional space. The scores are coordinates of the original objects in this newly constructed low-dimensional space. In contrast to the original space, which is organized in arbitrary order (determined by the experiment), PCs are organized according to the structure of the dataset (described by the variance). Thus, plotting of the most significant PCs quickly shows the structure of the dataset. The information from scores can be used in many applications as described in following sections.

The pattern recognition approach is based on the fact that the spectra which are mutually similar group together and form a joint cluster. Of course, it is supposed that, prior to the multivariate visualization, the spectra of each sample are highly correlated, having similar spectral fingerprint, and thus should share the same cluster. The size of the cluster is related to the fluctuation in the measurement (analytical system instability, local sample inhomogeneity, *etc.*). It may be stated that the higher the fluctuation, the broader the distribution of points in the cluster. Yet still, this projection is unsupervised and any conclusion drawn from this subspace (such as clustering or classification) reflects the real structure of the data, regardless of their origin, *i.e.* relation to individual sample.

3. PCA in LIBS

Advances in instrumentation development enable measurements with higher repetition rates, broader spectral ranges and better resolutions. Nowadays, an analysis results in datasets with thousands of variables [78] and millions of spectra [7]. Thus, the state-of-the-art LIBS system routinely provides big datasets (high number of spectra and variables) and so it is crucial to manage an effective and fast-response data processing. The MVDA algorithms must be applied into the analytical data correctly and with information-sensitivity.

When it comes to the exploratory analysis of spectroscopic data, PCA is the most often used algorithm. The first goal is to find similarities and patterns in the data that can lead to more complex conclusions about the sample set under investigation, *i.e.* clustering. However, PCA is an unsupervised learning algorithm and thus the projection created by PCA has to be carefully evaluated. Moreover, PCA deals with the overall variation in the data and may be overloaded with the fluctuation or outliers present in the data set. Consequently, PCA might not be able to accurately cluster individual samples [27].

PCA and its alterations (PCR and SIMCA) have already been successfully implemented to the analysis of LIBS data in many applications. De Lucia and Gottfried [27] brought a brief review on the utilization of chemometrics in a LIBS data analysis of hazardous and geological materials, especially focused on the endeavors of the US Army Research Laboratory. Harmon et al. [5] reviewed a combination of LIBS and MVDA algorithms in geology. Some more general reviews on processing of LIBS data through MVDA algorithms [17,28] were also published.

3.1. Data visualization

Structure, topology and quality of the data should be of interest prior to any further analysis. Therefore, a simple step of data visualization in a lower dimensional space is of great advantage. Most often, PCA is used solely for the visualization purposes and understanding of the fundamental relationship within the data, prior to the classification and quantification by other algorithms. However, De Lucia et al. [79] concluded a necessary condition for obtaining a meaningful clustering of LIBS data *via* PCA. This is that PCA is convenient when the intraclass variability is lower than the interclass variability, *i.e.* the discrimination gets worse when the intraclass fluctuation increases. Thus, other MVDA algorithms different from the PCA were also used for the low-dimensional visualization. Sirven et al. [80] showed that data in the PC space, *i.e.* their distances to the center of their respective cluster, follow a normal distribution. This is, however, in contrast to what was published in aforementioned articles [47,81] assuming the data to follow rather Extreme Value Distribution (EVD) over the normal distribution. Lazic et al. [82] suggested improving the linearity of data prior the creation of multivariate models.

The clustering of data points in the space given by selected original variables or PCs reflects the differences in unique and characteristic spectral fingerprints of individual sample matrices. In certain studies, the loading values were depicted directly in the PC space to show the significant variables in each PC [83,84]. It is obvious that a clustering of samples with distinct differences in their composition of matrix elements is quite effortless. Goode et al. [85] used PCA and the hierarchical cluster analysis (HCA) to show the possibility of clustering of LIBS data from various alloys and steel samples. Erdem et al. [86] used PCA to study archaeological pottery from eastern Turkey, suggesting some correlations between the composition of pottery and local sources. Gregoire et al. [87] used PCA to visualize polymer samples prior their classification by PLS. Vitkova et al. [88] explored clustering of data from various solid samples with PCA and then applied LDA and SVM for classification.

When LIBS is used in geological applications, it results in a large amount of data for the consequent MVDA analysis. Gottfried et al. [83] studied the topology of the minerals spectra in the PC space prior their classification using PLS-DA. Their spectra showed significant differences in detected spectral lines and relative intensities. A similar phenomenon was shown by Porizka et al. [89]. They presented a study describing, among other things, the differences in the visualization of minerals by PC space based on the broadband echelle spectra (200–1000 nm) and narrowband Czerny-Turner spectra (60 nm centered around 305 nm and 405 nm), see Fig. 1. The distinct separation and compactness of data clusters was retained. But the eigensystem, *i.e.* structure, of the data matrices changed accordingly to the composition of variables in the selected wavelength ranges. This is natural for data obtained by using different LIBS systems and/or under varying analytical settings.

Clustering and classification of material based on already established data libraries would be of great interest to many research groups.

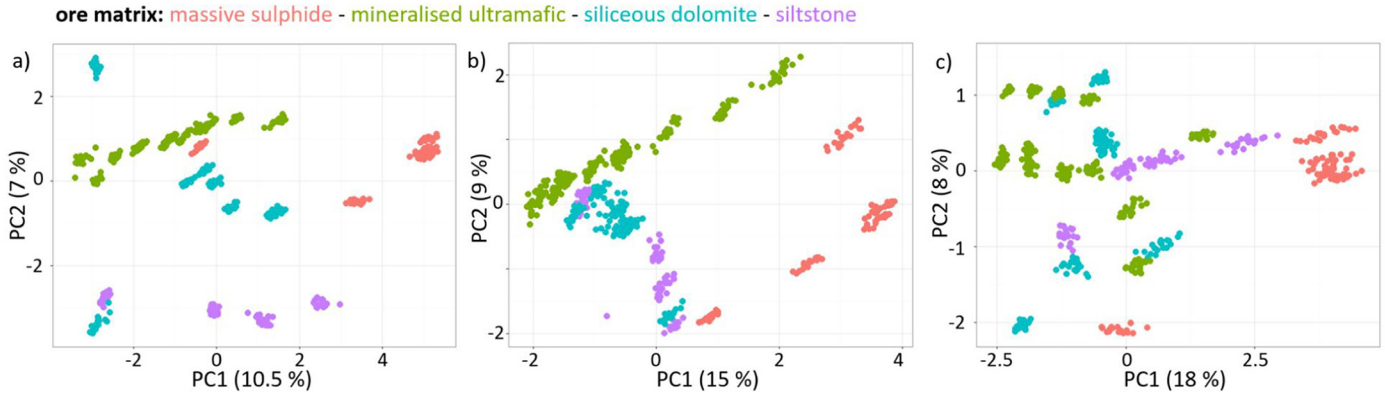

Fig. 1. This figure depicts a difference in PCA projection based on data from a) broad-range echelle spectra (200–1000 nm), and short-band (60 nm window) Czerny-Turner spectra at b) 305 nm central wavelength and c) 405 nm central wavelength. Data represent 28 sedimentary rocks in four different matrices. Obtained from [89] with permission provided by Elsevier and Copyright Clearance Center, license number: 4347640902794. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Lanza et al. [90] provided a discrimination of carbonate minerals in the PCA space; data were measured using the MSL Curiosity. Moreover, such analysis naturally leads to a provenance study. McManus et al. [91] used PCA for a provenance study of gemstones from 3 US states. The major concern of their provenance study was a diverse database covering varieties of geological materials including complex metadata (including original GPS locations of collected samples). That was, however, a very bold idea that would be hardly met in reality. Alvey et al. [92] executed an extensive provenance study of 157 garnet minerals from 92 world-wide locations and arranged them in a reduced PC space.

Studying organic material and polymers is also of considerable interest, however, the discrimination of related data depends only on the ratios and relative intensities of major elements (C, H, N, and O) and molecular bands (CN and C_2). This aspect makes the classification of such samples very challenging, moreover, some more sophisticated and non-linear algorithms are needed. Despite that, De Lucia and Gottfried [93] showed a possibility to discriminate spectra of explosives samples in PC space (3 PCs accounting for 92% of variance). Labbe et al. [94] presented a PCA of biological samples where the first two principal components stood for 99% of variance; the first PC represented 98% and the second only 1% of total variance. The influence of the substrate on the bacterial spectra was also studied [84,95]. It was shown that PCA was capable to discriminate bacteria of different strains, even when the samples were analyzed on different surfaces. During the sampling of thin films (such as bacteria strains spread on the substrates) a certain amount of the substrate itself was also ablated and its characteristic radiation contributed to the detected spectra. Therefore, it was necessary to filter this contribution out because it had no particular analytical meaning and could be confusing for the consecutive MVDA.

Another interesting conclusion concerns the analysis of biological samples and polymers by LIBS in general. Spectra of those samples reflect compositions that form complex molecules namely in C, H, N and O in the samples. Thus, the techniques providing molecular information (such as Raman spectroscopy) are more suitable for this kind of analysis. Despite that, LIBS proved to be a powerful alternative. Landström et al. [96] applied PCA directly on echellograms (CCD images of plasma characteristic radiation spectrally resolved *via* an echelle spectrometer) to discriminate biological and chemical warfare agents.

Tandem systems (such as the complementary LIBS and Raman spectroscopy techniques) are often used to gain more information (elemental and molecular composition) from the sample [97]. This, in turn, improves the MVDA algorithms' performance. Hoehse et al. [98] used PCA to visualize inks and pigments measured by LIBS and Raman spectroscopy, and a similar analysis was done by Prochazka et al. [99] on bacteria samples. In both cases, the LIBS and Raman spectra were simply merged together. Naturally, the combination of LIBS and Raman spectra improved the visualization and classification of samples.

It is also worth mentioning that PCA was used for a visualization of biological warfare agents based on the typical 1D spectra (intensity *vs.* wavelength) and 2D echellograms (images downloaded directly from the CCD detector on an echelle spectrometer) [100]. Porizka et al. [78] further developed the concept of echellograms proposed by Larsson et al. [100]. They proved the possibility to use directly the truncated information obtained from a CCD detector and to accelerate the data processing. In this kind of analysis it is not necessary to slow down the computing process by downloading and transforming the echellograms to spectra. Transforming a large amount of data is the bottleneck of the analysis when it comes to a potential utilization of echelle spectrometers in the high-repetition rate experiments. In fact, only a limited number of pixels from the whole megapixel image is necessary. Such an approach may lead to a significant enhancement of the repetition rate with an intensified detection of broadband spectra (using echelle spectrometers and sCMOS detectors).

Lasue et al. [101] compared PCA, ICA and Sammon's map (SM) for a visualization of ChemCam data. Of course, non-linear SM proved its supremacy over the other algorithms but this supremacy appeared only because of inconsiderably selected figures of merit. They used the stress function by applying the distances between points of the same cluster (the stress value of 26% for PCA, 17% for ICA and 4% for SM). Nevertheless, this indicator (stress value) is unfair because the SM model of the original data space is provided by finding the global minimum of the stress function. Moreover, no optimization of the number of PCs or ICs is done and the methods are compared strictly in the 2D space. Harmon et al. [102] used PLS-DA latent variable to study the clustering distribution of minerals spectra. Klus et al. [103] used SOMs in a similar analysis to show clustering of spectra of similar minerals in the grid of a self-organized map.

3.2. Multivariate elemental mapping

Mapping of elements and their distribution on the sample surface by using LIBS is a well-known method [1]. In the elemental mapping, a location on the sample surface, *i.e.* on the map, is represented by a relative intensity of selected spectral line of the element or even by the content of this element. But, most recently, the application of PCA has transformed the data analysis from a basic visualization in a reduced dimensional space to more sophisticated approaches. When the scores of selected PCs are depicted in the map (hereinafter referred to as multivariate map) it can represent a distribution of elements' combinations or even of individual matrices. This approach gives an

additional insight into a sample composition and a relation of individual phases (depending on the sample and the structure of original data). The meaning of PCs is then described as the original data by respective loadings and their spectral structure. The surface mapping of heterogeneous samples surfaces is in the spotlight of geological analysis, while it enables an elaborate understanding of individual phases and their distribution on the sample surface. This is due to the convenience in sample form (hard solid matrix) and its fast, straightforward preparation (cutting and polishing). Nevertheless, this approach in multivariate mapping of geological-sample surfaces may spread also to other applications.

Amador-Hernandez et al. [104] used LIBS to analyze printed electrodes in 3D. The mapping of the surface was provided in 25 × 13 spots and 10 consecutive layers. Intensities of Ag, Al, Au, C, Pd, Pt, and Ti spectral lines were estimated and proceeded to a PCA for further dimensionality reduction of data from layers 1, 3, 6 and 10. Cluster analysis (k-means clustering) was then applied on selected PCs, separately on data from each individual layer. From each layer, up to 3 PCs covering joint variance from 88.6% to 94.1% were selected. Data points were then assigned to a defined number of classes and depicted in the map showing the distribution of materials rather than elements in the investigated layers. In their following study [105], authors further worked with the same data set showing only PC scores biplots with distinct clusters.

In the work by Klus et al. [81], the distribution of uranium content in a sandstone-hosted uranium ore was investigated using maps of uranium line distribution (U II 409.01 nm), variation in the intensity of background (find out more about this in [106]), and values of the first principal component. All the maps showed a high degree of mutual numeric correlation. Thus it was found out that the first principal component (with 82% of variance) is namely representative of the uranium content variation. Another step in the MVDA analysis was done in order to reduce the original data set for PCA modelling and in turn to reduce demands on the computing power and computing time. The original data set was truncated to ~11% (randomly selected spectra) which resulted in creation of a model. The PC scores of the rest of the data set were then predicted based on the PCA model. A predicted map which was consequently created correlated well with other maps. This approach turned out to be a promising way of data set truncation with preserving the information carried by the original data.

Romppanen et al. [107] employed the PCA algorithm (SVD in fact) in their study of the rare earth elements distribution on the surface of yttrium-bearing ore minerals. They depicted maps of first 4 PCs (covering together 99.5% of variance). Individual PCs were related *via* their loadings to four sets of elements (Al and Ca; Ca, Y, Fe, and Al; Si, Fe, and Y; Fe and Y respectively to each PC). In the next step, data points were plotted in the PC space given by 3 components (PC 2–4) showing the clusters of individual matrices. It was assumed that spectra would cluster in this 3D space based on their similarities. Data points were then colored by their position in this artificial directional cube and replotted in the map. Finally, the individual clusters were assigned to minerals present in the sample. Despite a rather demanding data processing, the obtained results represented well individual minerals with a focus on yttrium distribution.

Moncayo et al. [108] performed a study similar to those already mentioned. They tested the performance of PCA in an analysis of megapixel elemental maps - a data set compiled of more than 1 million LIP spectra, see Fig. 2. The sample was highly heterogeneous and contained matrices of turquoise, pyrite and silica-bearing mineral. It was once again shown that the stand-alone PCA is capable to dissect the multidimensional signal and to assign various minerals (matrices) to individual PCs. The contribution of Si and Ti-bearing minerals was found. Those minerals were present in low amounts and distributed on the sample surface. They were present in high order PCs.

The presented PCA approach, or the mapping using MVDA algorithms in general, may be further extended by using other linear (ICA, TDA) and even non-linear (SOM) MVDA algorithms. Special attention will be given to SOM because they provide a powerful alternative to PCA [69]. Klus et al. used SOM for multivariate mapping of U-Zr-Ti-Nb in sandstones-hosted uranium ores [103]. The total number of 22,500 spectra was modeled, based on their similarities, into 900 neurons. Neural responses were then correlated with individual spectra and depicted in the original map. That showed the isolation of individual matrices (Si-, Ti- and U-based minerals) and also the affiliation of different elements (U, Zr, and Nb). A similar work was done by Pagnotta et al. [109] in the study of ancient mortars. LIBS maps of selected elements were processed through the SOM algorithm to achieve a segmentation of different materials present on the sample surface.

Another non-conventional utilization of PCA was presented by Lanza et al. [110]. They used a combination of depth-profiling LIBS data with PCA for the study of weathered surfaces of basaltic rocks. The PC scores showed a clear dependence on the depth-profile along the first PC. The investigation of PC loadings suggested that the amount of Ca, Na and Mg was increasing when going through the weathered layer and *vice-versa* the content of Mn and Ba was decreasing. Such an application proves the versatility and flexibility of the data analysis done by using PCA.

The additional value of multivariate mapping implemented to the LIBS applications is the fast response and robust processing of MVDA algorithm. Multivariate mapping may be a new, progressive way in the exploratory data analysis, enabling to process large data sets without prior knowledge. Inferred information is clear and can be directly used for deciphering of the contribution of different matrices and their mutual relations.

3.3. Dimensionality reduction

A typical size of obtained data matrix ($n \times p$, i.e. number of objects *vs.* number of variables) varies from experiment to experiment. First of all, the total spectra set is given by the number of samples and the number of detected spectra per sample and are usually collected *ad hoc* for feasibility studies. Total number of spectra per sample set is rather low, in many cases not even reaching 100 samples and 1000 spectra, as it is demonstrated by data collected from reviewed articles and depicted in the Fig. 3. The reason for that is that the experiment is usually done using low repetition rate LIBS systems and obtaining large data sets is thus tedious. From a statistical point of view, large data sets are more convenient. With higher repetition rates the number of objects (obtained spectra) can grow over thousands. In the case of aforementioned multivariate elemental mapping, Section 3.2, the number of objects was in the range of 10^5 to 10^6; giving millions of pixels in the elemental map.

The total number of variables depends on the spectrometer used during the experiment. Going to extremes, the number of variables can reach over 40,000 with echelle spectrometers. Overall, the improvement in the instrumentation results in higher number of objects and variables. This is a shift in the paradigm and a new challenge for data processing.

In the most extreme case, the size of the data matrix obtained during a LIBS experiment can be in the range of 10^6 objects × 10^4 variables. Processing such data matrix is then demanding on computing power and improvement in this aspect is beneficiary. Therefore, in following chapters, we are going to dissect individual approaches in data processing with special attention to a dimensionality reduction in objects and variables.

3.3.1. Dimensionality reduction in variables

The literature research does not reveal any decisive viewpoint which would indicate how the number of variables should be reduced, or if the reduction should be done at all. Keeping the whole spectral range without any selection of spectral lines and processing of peak intensities can lead to increased computation demands. It also

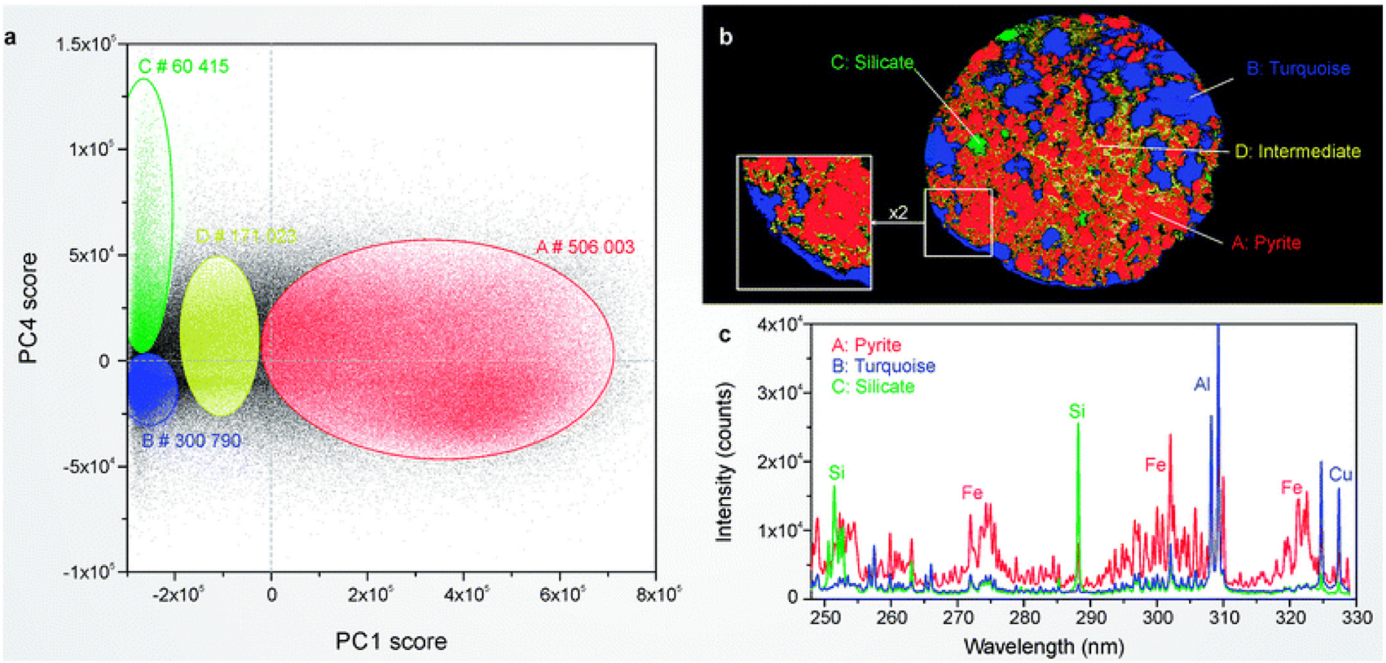

Fig. 2. This figure depicts the processing of a megapixel elemental map. Data were collected throughout the analysis of a heterogeneous turquoise mineral. a) Scatter plot of PC4 *vs.* PC1 with highlighted clusters of data points; clusters represent a similarity in spectral fingerprint of the highlighted spectra. b) Clustered data are further depicted in the pseudo-colored map to respective location demonstrating the distribution of individual matrices within the heterogeneous sample. c) Spectral fingerprints of selected matrices: pyrite, turquoise and silicate. Obtained from [108] with permission provided by Royal Society of Chemistry and Copyright Clearance Center, license number: 4347680572130. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

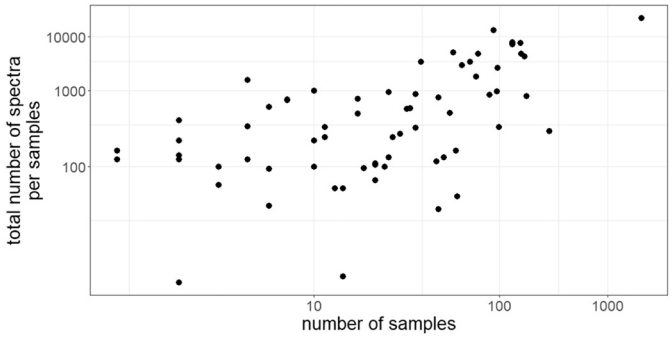

Fig. 3. This figure depicts the total number of spectra over total number of samples collected from 70 articles reviewed in this paper.

introduces a lot of noise and some correlated, redundant variables into the model. It is indisputable that the background information may be beneficial in certain extreme cases. Klus et al. [81] proved the importance of background signal in the multivariate mapping of uranium. The detection of uranium is challenging due to the high density of lines that exceed the resolution of a typical spectrometer. Thus, its content in the sample may be correlated with systematic variations of background signal.

The utilization of echelle spectrometer gives the possibility to detect abundant broad-range spectra. Such analysis can be provided also using broad-range Czerny-Turner spectrometers, however with significantly lower resolution compared to their echelle counterparts. In general, such spectrometers are beneficially utilized in the LIBS experiments and provide extensive data sets for MVDA. On the contrary, Colao et al. [41] suggested selecting the most significant lines and a related proper wavelength range when using a short-band Czerny-Turner spectrometer. A similar study was introduced by Porizka et al. [89]. It was shown that the utilization of short-band Czerny-Turner spectrometer instead of broad-range echelle spectrometer leads to the same classification power when the spectral window is judiciously selected. Moreover, limits of detection of selected analytes improved in this experiment. Sirven et al. [111] used PCA to discriminate between three types of soil prior further classification and a quantitative analysis by ANN. They showed that a short spectral window, 10 nm, is sufficient for the discrimination of data in the PC space. However, shorter band widths led to a low performance.

Another possible way is to filter out the background noise by thresholding. The level of threshold may be set by the 3σ limit of detection. This approach is not universal and does not consider low intensities of the trace and minor elemental lines. Moreover, detected intensity below 3σ level is not accepted by the spectroscopic community as an analytical signal and should not be used in further data processing. Thus, there is no reason to associate this under limit intensity as a contribution from any particular element. Sahoo et al. [112] set manually the threshold to filter the noise from the spectra. Kong et al. [113] compared the classification performance after dimensionality reduction by a selection of spectral lines and a reduction into PCs. A similar comparison was provided by Lee et al. in the case of classification of salts using LIBS [114] and tandem LIBS and LA-ICP-MS techniques [115].

Especially in the case of broad-range echelle spectra, the number of the informative lines might be overloaded by the number of variables representing only the background noise. In the worst scenario, such data could lead to a confusion of the learning method, such as PCA, and consequently to its poor performance and to biased results [116]. Bousquet et al. [76] applied PCA on the full spectra, regardless of any information carried by the background and considering good SBR and SNR. This fact may be acceptable assuming that the systematic change in strong spectral lines intensities will simply overshadow the contribution from randomly fluctuating background noise. On the contrary, in their former study [44], they recommended to select only a small part of the spectra, *i.e.* relevant information from matrix element lines (68 lines in total; 4 for Al, 10 for Ca, 31 for Fe, 3 for K, 4 for Mg, 2 for

Mn, 11 for Ti, and 3 for Si). Darby Dyar et al. [117] showed that PLSR was unsuccessful in the quantification of sulfur while using the whole spectral range. This was then confirmed by the investigation of PCA loadings where the contribution of sulfur was overshadowed by other more significant lines. However, from the perspective of this review, this could be avoided or mitigated by the selection of spectral lines and their scaling prior to MVDA. The scaling could then lead to the enhancement of the sulfur significance in the eigensystem of the data matrix.

A rigorous wavelength selection and manual elemental line assignment is time consuming and tedious. Automatic algorithms designed especially for LIBS applications are still rather lab-made and commercially not available. Processed line intensities combined with their ratios can be used but this process demands prior knowledge of the sample composition. However, the qualitative analysis concerns the "usual suspects" and an experienced spectroscopist can predict the composition in major elements and identify the most common spectral lines typically detected in LIP spectra. Sirven et al. [80] also suggested selecting only spectral lines of major elements because they are supposed to dominate the MVDA performance. The suggestion was then not to consider the spectral lines from minor and trace elements. On the contrary, in their other work [42], they recommended to include spectral lines of trace elements in the data set. Moreover, they compared the classification power of the system based on spectra with the artificially lowered resolving power (from the original 10,000 to final 2800). After a variable down selection, the physical interpretation of the loadings was much easier. They reduced the number of variables to 41 lines of 10 elements based on the PC loadings in consecutive steps of a global PCA approach. Then the number of variables was limited even to 10 lines (one line per each element) with a high classification accuracy proving that a low number of variables is not limiting when they are judiciously selected.

Gottfried et al. [83] used PCA for a significant feature identification of mineral spectra, naturally lines of Ca, Si, Na, and K together with Sr, Li, and Mg were identified. Lewis et al. [84] cross-plotted loadings of PCs to see the impact of individual variables on the variance describing the data set. Then, variables with a low impact were discarded from further computation. The authors showed that the contribution of Ca, Mg and trace elements had a significant impact on the analysis of bacteria. Pereira et al. [118] used PCA loadings to select major constituents, macro- and micro-nutrients responsible for the discrimination of spectra from LIBS analysis of plants, citrus leaves.

Munson et al. [119] investigated various approaches in dimensionality reduction in order to improve the classification capability between bacterial and chemical warfare agents. They used a) six lines (P, C, H, and O) and combination of their ratios and b) simple masks omitting the whole spectral range except the regions around selected lines. The approach using various ratios of line intensities brings a way to artificially increase the number of information-carrying variables. This may in turn improve the performance of MVDA algorithms. Moreover, the selection and processing of spectral lines enables to scale the variables and thus to unify their impact on the performance of MVDA. Anderson et al. [55] used five different approaches to improve the quantitative performance of PLSR. This was achieved *via* pre-selection of training and test subsets using PCA prior to k-means and hierarchical clustering, and SIMCA. They have simulated the Mars conditions (ChemCam system, pressure, distance) for over than 220 ore samples. They used the full spectral ranges despite the statement which claims the improved performance after variable selection. Anabitarte et al. [120] used kernel PCA to reduce dimensions prior to SVM in order to detect Al impurities in steel coatings. A relatively high number of 100 PCs was used in this study, then this number was further reduced to 8, 5, 3 and 1 PC showing that at least 3 PCs were necessary to achieve accurate results.

Selecting the right number of PCs is also of great interest in this case. This has to be done in order to avoid any overtraining and biased results. Typically, the high number PCs cover mostly only noise and, therefore, are usually omitted from computation. Marcos-Martinez et al. [121] stated that PCA for dimensionality reduction actually led to a decrease of the ANN's analytical performance. However, this is in contrast with aforementioned papers and should be accepted with care.

Variable Importance in Projection (VIP) presents a powerful way to identify the most prominent lines. The VIP is connected to factors provided by PLS models of the original data [122]. De Lucia et al. [123] studied the influence of the number of variables *versus* the whole spectra approach in the classification process of explosives spectra through PLS-DA. The set of variables was chosen by VIP having a score higher than 1.0. Naturally the contribution of C, H, N, O and molecules of C_2 and CN were the most significant in the analysis of explosives. After this selection, the number of selected variables was extended by their ratios, sums and combinations. It is noteworthy that this process is not fully correct from the mathematical point of view and should be approached cautiously. Regardless of our comment, this process was done twice, in the first case they got 40 variables and in the second case 177 variables. In the latter case, the ratios and combinations of variables followed stoichiometry and relative amounts of C, H, N, and O in the molecular bonds of analyzed samples. It was shown that the latter model had performed better than the first one, which proves the necessity to incorporate a higher degree of variability. The whole spectra model gave the best results, however, its performance was doubted by the authors due to the fact that the VIP procedure indicated a possible classification by factor not related to the contribution of C, H, N and O. Fink et al. [124] used a genetic algorithm to select variables prior to PCR and PLSR of analytes (such as Sb, Sn, and Ti) in plastic samples. They selected from 9 to 87 variables or the whole spectral ranges (up to 200 nm) for the PCR and PLSR models. The performance of the models based on the whole spectral ranges was significantly inferior. This once again proved the necessity to preprocess the data appropriately prior to any MVDA.

Corsi et al. [125] introduced an interesting way of the dimensionality reduction, similarly it was suggested in Section 3.1. The content of 11 elements within the archaeological samples (copper-based artefacts) was estimated by the Calibration Free (*CF*) LIBS [126]. The reduced data set (13 objects described by 11 variables) was then introduced into the PCA (66.3% of joint variance in the first two PCs). This is in contrast to typical analysis of raw LIP spectra.

PCA provides the dimensionality reduction of the original data set to a smaller set of independent variables but preserving the most of the variance. This essential feature of PCA may then be used primarily for data processing while the lower number of variables can fully represent the original data set. It is necessary to overcome the curse of dimensionality and select a right number of latent variables [127]. Then the selected number of PCs can be further used for classification and quantification purposes. A review of articles in which the authors implemented the reduced PC space for an exploratory data analysis and consecutive unsupervised classification is given in Section 4.

3.3.2. Dimensionality reduction in objects

The distinct separation of individual data points in PCA space suggests that samples can be classified based on characteristic LIBS spectra. However, PCA focuses on variances among data. Thus, any significantly different data overload the discrimination power of PCA and some smaller differences among the rest of the data are mitigated.

Multari et al. [128] suggested an approach to overcome this problem. Any well-discriminated and distinctly outlying cluster of data (spectra of a sample) is removed from the data set. Consequently, PCA is applied once again on the truncated data set when the differences between the formerly overlapped data points in the PC space are now revealed. This step-by-step approach may be regarded as a sort of dimensionality reduction in the sense of number of objects. This in turn reduces the variation in the data and changes significantly the data structure and topology. A US patent was issued [129] claiming

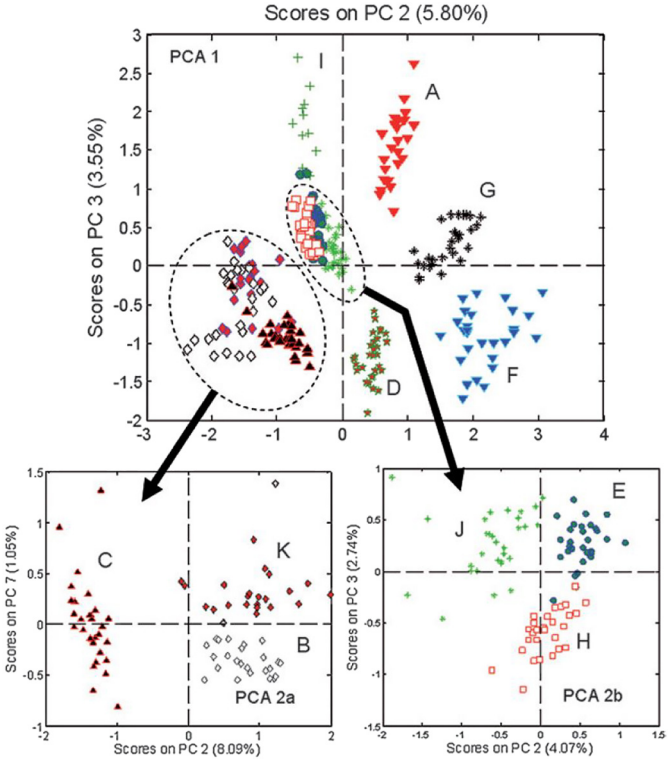

Fig. 4. This figure shows the capability to provide a provenance study. The distribution of yellow cake spectra in PC space suggests a possibility to classify samples according to their location of origin. Moreover, PCA was calculated three times; first for the complete data set (top figure) and then separately for both highlighted regions (bottom figures). This approach demonstrates the increased performance of PCA when the number of objects is reduced. Thus, the PCA algorithm is not overloaded with variance from a high number of sample matrices. Obtained from [42] with permission provided by Royal Society of Chemistry and Copyright Clearance Center, license number: 4347711143996. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

aforementioned data processing steps. Before this patent, Sirven et al. [42] suggested a so-called global PCA that works in step-by-step mode for estimation of geographical origin of uranium ores, see Fig. 4. This algorithm was used for a mitigation of the number of data and a consequent implementation of MVDA to reduced space; this was done similarly to the approach patented by Multari. This successive approach was then adopted by Colao et al. [41] for the identification of historical building materials.

Anderson et al. [55] studied a possibility to reduce the number of samples in order to reduce the computation burden. Only samples lying furthest from a cluster center were selected and used for a furthest-neighbor clustering tree. Other samples were discarded from modelling. Thus, this approach retains the biggest amount of variation in the data and at the same time avoids redundancy. The authors also reduced the original data set by PCA prior to the construction of a dendrogram. Despite their efforts, the most reliable results were obtained when they used the full training sets, covering quite a large amount of variation and diversity. Fink et al. [124] claimed to use PCA to select the samples with the highest statistical leverage for the calibration purposes. However, this process was not further commented.

3.3.3. Outliers filtering

It is crucial to preprocess the data prior to any further implementation of MVDA algorithms, Section 3.1. Outliers filtering should be also considered; especially when the internal standardization is not sufficiently effective in mitigating the fluctuations of analytical signal. In that respect PCA is a powerful candidate because its essence is being sensitive to high order of variation, which might lead to detection of outliers.

In general, it could be assumed that spectra detected from a shot-to-shot analysis of the homogeneous sample should possess a high degree of mutual correlation. Each spectrum is described by a unique set of values, *i.e.* descriptors, (correlation coefficient, Euclidean distance to the center of PC space, total energy of the spectrum, intensity of matrix lines, *etc.*). Selected descriptors represent analyzed spectra and relate them to the rest of the data set. The descriptors also enable a comparison of analyzed spectra with the rest of the data set. Those values are then compared and thresholded (either visually or numerically) which leads to a truncated data set with more similar spectra. This high mutual correlation may be demonstrated by a compact cluster of points in the PC space. The threshold level should be set carefully. Harder filtering should be avoided because it significantly influences the topology of the data, especially in the case of modelling. Fluctuation is natural to certain extent and describes the response of analytical system [32]. Therefore, only the most outlying points should be filtered.

Signal fluctuation is of great concern in LIBS and is a result of many independent parameters influencing the laser-matter interaction, which is usually concluded as the matrix-effect. The source of fluctuation is laser temporal instability, local heterogeneities on the sample surface, *etc.* The process of material ablation and consecutive plasma formation is therefore not stable from shot-to-shot perspective. The properties of laser-induced plasma (described by its temperature and electron density) are significantly affected together with emitted analytical signal. Those aspects increase the standard deviation (variance) of data points in the *p*-dimensional space and decrease the performance of the MVDA technique, which is then reflected in the resulting figure of merits.

Outliers filtering is a possible preprocessing step mitigating the analytical signal fluctuation. But, as it is typical in LIBS applications, the experimental parameters and protocols in data processing are optimized *ad hoc* and various approaches and algorithms are used. This leads also to non-unified ways of outliers filtering such as linear correlation [130], dendrogram [112], Student's *t*-test after the computation of the Mahalanobis distances of individual spectra in the original data space [131], total energy of each spectrum [132], or intensity of matrix lines [133,134]. PCA for outliers filtering was introduced into the processing of LIBS data by Sirven et al. [80]. They used basic visual thresholding and omitted up to 30% of spectra prior to a rocks classification in a preflight ChemCam testing. In their further work [42] they judged outliers based on the abnormal value of their residual fit. Myakalwar et al. [36] used a dendrogram after the PCA dimensionality reduction in order to detect outliers in the analysis of pharmaceutics. Porizka et al. [77] improved the quantitative analysis of Cu in igneous rocks by filtering based on the distribution of data points in the PCA space. Prochazka et al. [99] applied PCA on the set of measurements of the same sample. In the space given by PC1 and PC2, the most outlying points (those with the biggest distance from the weighted center) were omitted and PCA was recalculated. This procedure was repeated until the selected number of the most similar measurements remained.

Porizka et al. [56] compared impact of three different approaches for outliers filtering (PCA, linear correlation, and total energy) on classification based on the selected figures of merit (overall accuracy, sensitivity, and specificity). Any of those approaches leads to an improvement in the classification accuracy. The total energy proved its supremacy over PCA and the linear correlation. It is, however, necessary to mention that the case study involved only steel samples. It was also found that each approach led to filtering of different data points (spectra), see Fig. 5. That, in turn, affected the topology of filtered data and also the differences in resulting figures of merit. Therefore, it should be advised to compare different approaches when samples of other matrices (other than steel) are subjected to LIBS.

It needs to be stressed that outliers filtering should be applied only to the classification or quantification (calibration and regression) of

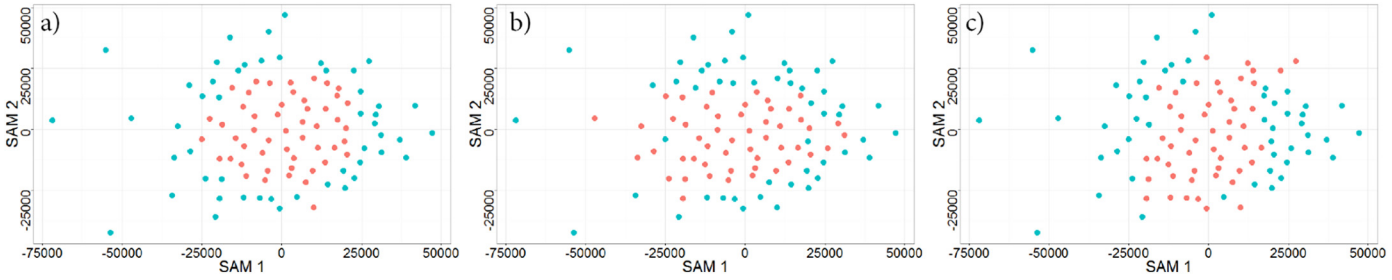

Fig. 5. This figure depicts a distribution of data points filtered using a) PCA, b) linear correlation, and c) total intensity approach. Depicted in the first two components of Sammon's map, data left for further data processing are marked in salmon-pink and outliers are marked in cyan. The Figure was obtained from [56] with permission from Elsevier and Copyright Clearance Center, license number: 4347640902794. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

homogeneous sets of samples. In the case of multivariate elemental mapping, it is naturally counterproductive to omit outliers from a data set which describes heterogeneous samples. In that particular case, outliers are mostly the subject of the analysis, for instance the detection of gold in ores [135].

3.4. Regression

A problem of modelling a continuous variable is called a regression analysis. The extensive theory and examples of univariate calibration may be found in books dedicated to LIBS [13]. The Principal Component Regression (PCR) is an adaptation of the PCA for the purposes of a multivariate regression [30].

In regression analysis, or calibration, two variables are correlated *via* an equation describing this relationship. The most basic concept deals with the relation of the elemental line's intensity and the content of the analyte. Having an option to incorporate the information from more variables at once opens new possibilities in the regression analysis, *i.e.* the multivariate regression. This suggests that more than one variable contribute to the response related to the content of the analyte. In the regression, the data matrix **X** is supplied with a vector **c** describing the known concentration of an analyte in each object. Note that this vector may be related to any parameter characterizing each sample and in turn may affect the quality and structure of LIP and related emission spectra (such as sample hardness, roughness, *etc.*).

Bousquet et al. [44] concluded that the univariate calibration of Cr in soil samples does not provide good results due to a vast variation of matrices. Instead of a single univariate calibration, several calibration curves have to be constructed. For the classification, they suggest to use PCA directly and to separate samples according to their location in the PC space. Choi et al. [136] analyzed 21 samples with distinctly different matrix and used PCA for visualization prior to the univariate calibration. Of course, a significant impact of the matrix effect was evident in the PC space and also in the structure of univariate calibration that was split according to each individual sample matrix. Thus, for the regression analysis, the samples should be of the same matrix in order to provide the best possible results. Porizka et al. [77] used PCA to filter the outliers and then to classify samples according to their matrix. The univariate calibration of Cu in soil samples was then assessed giving lower bias. In fact, the matrix effect was not mitigated but completely avoided using this kind of data processing. They also compared PCR and PLSR to the univariate calibration after classification proving that a multivariate data analysis was not a necessary step in data analysis. Regardless of this finding, it is generally accepted and proved in the LIBS literature that multivariate algorithms may compensate the matrix effect to certain extent and may influence the LIBS analysis itself [43,137].

Wisbrun et al. [138] brought a pioneering study on calibration of heavy metals (Cd, Cr, Cu, Ni, Pb, and Zn) in soils, sand, and sewage sludge using PCR. They selected and weighed 50 spectral lines of analytes prior to a PCR computation; only two PCs were used. They also estimated limits of detection of tens of ppm. Death et al. [45] performed a quantitative analysis of six elements (Al, Fe, K, Mn, P, and Si) in iron ores by applying PCR based on a maximum number of ten PCs forming a model of short band spectral region (centered at 250, 400 or 750 nm). This analysis showed results in good agreement with referenced values. An example of PCR used for the modelling and prediction in the quantification process of Si is shown in Fig. 6. In their further work, Death et al. [46] used PCR for quantification of elements in three sets of samples from Australia and West Africa showing good calibration results. Tripathi et al. [139] used PCR to predict the content of plutonium oxide surrogates (CeO_2, Fe_2O_3, Cr_2O_3, MoO_3, NiO).

Most often, the results of PCR are compared with those of PLSR, the latter is generally considered to have better performance. Both algorithms are collated across various applications of LIBS. Fink et al. [124] analyzed a set of polymers. They also showed the improvement of one order of magnitude in the detection limits when comparing the performance of multivariate and univariate algorithms. Doucet et al. [140] compared PCR and PLSR to find out which one performs better in a quantitative analysis of elements (Cu, Fe, Mg, Mn, Si, and Ti) in 260 aluminum alloys. Data were obtained using a LIBS system with Paschen-Runge spectrometer focused on 11 elemental lines (no matrix line

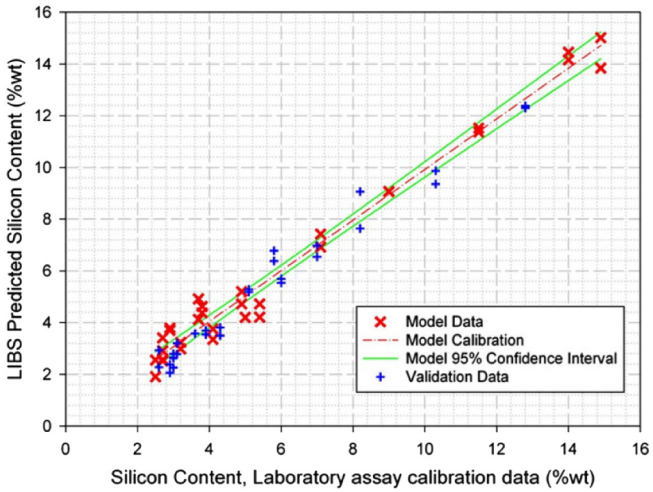

Fig. 6. This figure shows a PCR calibration model constructed from 16 samples and validated by 15 samples unknown to the model. The sample set was collected from iron ores and the quantification study was in this case focused on the Si content. This figure also demonstrates a good performance of PCR when the sample set is well chosen; the coefficient of determination was estimated to 0.99 and average relative errors were estimated to 2.5% for model and 5.6% for prediction. Obtained from [45] with permission provided by Elsevier and Copyright Clearance Center, license number: 4347691418259. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of aluminum). They stressed that it is necessary to linearize the data prior the implementation of MVDA algorithms. This data pretreatment resulted in a dramatic improvement of figures of merit. In their following study [49] they compared PCR and PLSR used for a quantification of pharmaceutical tablets and their composition through the emission of molecular bands (CN, CH, C_2). Yaroshchyk et al. [132] provided a calibration and quantification of Fe in ore pellets using PCR, PLSR and variations of PLSR (serial PLSR and multi-block PLSR). The data were obtained using a LIBS system with two spectrometers covering two separate spectral ranges in UV and VIS, resulting in two different datasets. In the case of PCR and PLSR, the considered data set was either from UV range, VIS range, or merged UV and VIS ranges. S-PLSR and MB-PLSR were able to work with both ranges (UV and VIS) at once without merging. Thus, S-PLSR and MB-PLSR present another way to process data from different spectrometers or even techniques, considering hyphenated systems (such as LIBS and Raman). Devandag et al. [141] compared the performance of univariate (6 various spectral lines) and multivariate (PCR and PLSR) calibration for the purpose of a quantitative analysis of Mn in glass matrix. One order of magnitude improvement in detection limits was obtained with the utilization of MVDA.

3.4.1. PCR counterparts

PCR is not so massively utilized due to its relatively low performance compared to other algorithms. Despite that we deliver articles where PCA was used as the first step of an analysis for visualization prior to any MVDA regression. Sirven et al. [43] investigated the influence of matrix effect on the performance in a quantitative analysis of Cr by PLSR and ANN. In another study [111] they used PCA prior to ANN for a quantification of Cr in soil. El Haddad et al. [137] visually inspected the ICP-OES data through PCA in order to check their clustering. As a result, no clusters in PC space were observed which directly implies that the chemical composition of samples obtained through the ICP-OES analysis does not significantly differ. It might be stated that on a certain level the samples collected from various places of the mine could be considered similar. Martin et al. [142] used PCA for discrimination of four wood species. PLSR was then used for a quantitative analysis of Cu and Zn. Martin et al. used PCA (scores and loadings) prior to a quantitative study using PLSR of the forensic samples [143] and nuclear materials (Ce, Cs, and Sr) [144]. Anderson et al. [145] used PCA for a visualization and loadings inspection prior the comparison of PLSR and ANN performance in quantitative analysis of soils for the ChemCam instrument. Bacterial contamination of food was assessed by a combination of PCA and PLSR [146]. Bilge et al. used PCA for a visualization of LIBS data from meat [147] and milk [148] samples prior to a quantitative analysis by PLSR. Moncayo et al. [149] visualized milk samples prior to their quantitative analysis done by ANN.

It is worth mentioning that PLSR was used in other applications, such as a quantification of components in geo-samples under Martian conditions [150], of micronutrients in pellets of plant material [151], of loss on ignition in iron ore [152] and of Cl in concrete [153].

3.5. Clustering and classification

The PCA model transforms the original data space and provides a low-dimensional visualization of the data. This projection provides only the unsupervised pattern recognition. Despite that, general relationships between data may be observed. The most basic kind of analysis is the clustering of data points according to their mutual similarities. The level of similarity is determined using a metric set by the Pearson correlation coefficient or by the distance (Euclidean, Manhattan or Mahalanobis) between data points in the k-dimensional space (k is the number of PCs, $k < m$). Resulting values are often organized in a dendrogram (a result of hierarchical clustering) and utilized for the discrimination of objects based on their characteristic spectra.

In contrast to the unsupervised pattern recognition (clustering), there is the supervised approach of classification. In the case of the latter, data matrix **X** is extended with a vector describing the class membership of individual objects. Afterword, a model is constructed using a MVDA algorithm and then unknown data are classified. For those purposes, the PCA-based algorithm may be used - Soft Independent Modelling of Class Analogyies (SIMCA). The concept of *soft modelling* enables to classify one object to more groups at the same time or to leave the object unclassified. At this point we would like to emphasize that MVDA algorithms classify the obtained spectra, not the samples themselves. Only after a careful validation of obtained statistical results, a classification of the real samples might be considered. Therefore, it is important to include figures of merit describing the performance of the classification process.

This section overviews articles that deal with a classification of LIBS data through i) the implementation of classification algorithms after the dimensionality reduction by PCA or ii) the direct utilization of the SIMCA algorithm. In contrast to quantification, an application of LIBS technique benefits from the strong matrix-dependence when it comes to classification. Possession of an extensive data library, *i.e.* collection of unique spectra covering the diversity of materials, is absolutely crucial in order to provide a reliable classification. No data library is currently available commercially, and even if it was, the transfer between different LIBS systems is not possible yet. Therefore, each research group is building its own database. Bohling et al. [154] suggested an approach where it would be possible to further extend an already existing model with new measurements.

The most challenging-to-build database seems to be the collection of various minerals and ores. Though, the application of LIBS to geology is one of the most promising in terms of LIBS potential (*in-situ* analysis). For instance, it is expected that MVDA discrimination of samples can emulate the distribution of the igneous rock types in the so-called QAPF and TAS diagrams and therefore can provide a routine provenance study. The provenance study of minerals was claimed in the patent [155] utilizing the PCA and other MVDA algorithms. The feasibility of LIBS combined with MVDA for the classification of geological samples is also demonstrated by the Curiosity rover equipped with the ChemCam device [156].

PCA is usually utilized only for a brief inspection of the spectra and a visualization of hidden patterns within the data set. Therefore, the preliminary exploratory data analysis using PCA is followed by an application of linear and non-linear MVDA algorithms for classification and quantification purposes. Nevertheless, distinct clusters in the PCA scores plot suggest that it is possible to use simple algorithms and that the use of advanced non-linear methods is in certain cases (low number of samples) unnecessary. Thus, in several articles the classification of samples was demonstrated only by a visual inspection of the PCA scores plot [44,76,77,111]. Samuels et al. [116] presented a discrimination of biological samples, such as bacterial spores, molds, pollens and the protein ovalbumin. They proved that it is possible to separate biosamples by applying unsupervised multivariate method (PCA) on their typical LIP spectra (full echelle range). PCA can also suggest the class of the new object (spectrum, set of spectra representing a sample) in a visual way by projecting it into a model given by PCs subspace [80]. Yet they concluded that in this approach there were insufficiencies, such as lacking any supervised process and related figures of merit.

A rather pioneering work by Fink et al. [124] claimed a clustering of spectra from plastic samples in the PC space (3 PCs covering 96% of variance) and suggested their potential classification. Hybl et al. [157] presented a discrimination of biological aerosols from natural background by PCA. 12 samples in four classes were analyzed by a broad-

band LIBS system and 30 most prominent lines were selected. Data points in the PC space indicated possible inter-class but no intra-class discrimination, *i.e.* the system showed a low specificity. However, as the authors stated, this work did not push the limits of chemometrics due to the simplicity of the presented data, *i.e.* the number of sample classes was quite low. Baudelet et al. [158] demonstrated the possibility to discriminate between individual bacteria using their Ca, Na and K signals obtained from a fs ablation. De Lucia et al. [159] showed a classification between explosive and non-explosive organic materials in the PCA space. Only first PC (70%) was used and a threshold was applied in order to discriminate between samples. Xia and Bakker [160] compared the performance of PCA-Adaboost (a hybrid PCA designed for classification) and PLS-DA in a real-life application of LIBS providing a sorting of materials. Considering the significantly different matrices that are subject to classification, the estimated optimal number of 100 PCs for PCA-Adaboost and 80 latent variables for PLS-DA is rather disturbing - especially when a good separation of clusters was shown in the space of 3 PCs.

3.5.1. Classification after dimensionality reduction by PCA

PCA is first used to reduce the number of dimensions, which can get over 30,000 of variables (*i.e.* wavelengths in the detected spectrum) in the case of LIBS data. Then, only the most descriptive PCs are kept in the model. This step preserves only the most valuable information and mitigates the contribution of fluctuations and noise within the data. Thus, this preprocessing step increases the prediction power of the model on which other algorithms are applied to yield a sample classification. This approach was successfully used also in other techniques, such as Raman spectroscopy [161], Fourier Transform Infrared [162], and X-ray fluorescence [163]. Yet, it has to be stressed that this approach is unsupervised and the classification results may be biased.

Samek et al. [164] applied the Mahalanobis distance on the reduced PC space as a metric to classify teeth caries. In another work [165] they used the same approach to classify solid standards. Yueh et al. [166] used PCA (11 PCs) followed by HCA to classify murine soft tissue samples (brain, kidney, liver, lungs, muscle, and spleen). PCA was used to reduce the original multivariate data to a lower number of variables while preserving the latent variance at the same time. Obtained results were compared with the performance of the PLS-DA and ANN applied on the original dataset. Moderate results were found regardless of the utilized algorithm. A higher number of measurements encompassing the system fluctuation and improved outliers filtering was suggested.

Gottfried et al. [167] used PCA for a visualization of several hazardous materials (explosives and bacteria). Three PCs constituted 99.39% of total variance in the model based on 6 ratios of the most significant elemental lines (C, O, H and N). Then simple classification was done directly in the PC space by drawing ellipses and counting positive and false hits.

Unnikrishnan et al. [168] used PC scores for a clustering based on Mahalanobis distance and also used spectral residuals for the classification of plastics where the diagnostic threshold of classification was estimated using Youden's index plot and Receiver Operating Characteristics (ROC). Spectra were masked except for three 15 nm spectral regions including lines from C, H, and CN. Poor SNR and SBR which were due to the composition of the samples (C, H, N, and O) spoiled the classification accuracy. Pokrajac et al. [169] introduced PCA to LIBS spectra of four proteins. After the PCA dimensionality reduction they implemented various classifying algorithms (kNN, LDA, SVM, and ANN). They found the best classification performance for a relatively high number of PCs (from 12 to 41). That is rather extreme taking into consideration the total number of samples, spectra per sample, and spectral lines per spectrum. This fact might be caused by a low number of informative spectral lines with poor SBR and SNR despite the broad wavelength range (200–950 nm) of echelle spectrometer. On the contrary to this finding, they claimed that the optimal number was 6 PCs, as they estimated from the Cattell's scree plot.

Porizka et al. [77] used PCA for the dimensionality reduction and following classification in the space of first three PCs (covering up to 97.5% of total variance) applying the Gaussian clustering algorithm. The Gaussian distribution of points in each cluster was assumed but not statistically tested. After the clustering, they did a univariate quantitative analysis. Vitkova et al. [170] presented a classification of 29 brick samples using LDA after PCA reduction. They used the table-top and stand-off LIBS systems. In PCA scores plots, the firing temperature of bricks was evident. This leads to the matrix effect influencing the quality of LIP spectra. Merk et al. [171] tested the PCA and PLS-DA capability to provide a fast turnaround time in metal scrap sorting, and they reached a repetition rate 25 samples per second.

Amador-Hernandez et al. [104,105] used PCA prior to unsupervised k-means clustering for class assignment and consecutive multivariate mapping of printed electrodes. In other articles, PCA was used for a dimensionality reduction prior to hazardous material analysis through neural networks [154], classification of proteins by SVM [172], classification of geological samples by k-means clustering [55], classification of wood samples [173], classification of bricks utilizing stand-off system and LDA [88], classification of soils by PLS-DA [174], salts by PLS-DA [175], and classification of explosives by kNN [112]. Pontes et al. [176] used the successive projection algorithm (SPA), genetic algorithm (GA), and stepwise formulation (SF) for a dimensionality reduction prior to LDA classification. The findings were then compared to SIMCA.

3.5.2. Supervised classification using SIMCA

SIMCA is one of the simplest classification algorithms and its concept was developed in 1970s by Wold and Sjöström [177]. In the modelling step, each class of objects is individually modeled using PCA; each PCA model can be described by a different number of PCs. Then, unknown or test data (for cross-validation) are projected in the multidimensional spaces represented by individual PCA models and their distances to the center of gravity of individual models are calculated. The geometric distance can be also converted to statistical probabilities. Finally, membership of the unknown or test data is assigned to a class which has the smallest (or under threshold) distance to the model.

Sirven et al. [42] presented a classification and a provenance study of uranium ores using SIMCA. They did such a supervised classification twice, first time, they used 41 lines and second time only 10 lines (one spectral line per element). In both cases, good figures of merit were obtained. They also suggested that supervised classification algorithms could provide more accurate results than the unsupervised ones, such as hierarchical clustering, k-means clustering, *etc.* Clegg et al. [178] concluded that both PCA and SIMCA algorithms will be beneficial for the Mars rover and related exploration. SIMCA was employed with a satisfactory accuracy of 88.1%. Colao et al. [41] used SIMCA to classify 35 unknown samples based on 10 reference samples. In other works, they used SIMCA to classify citrus leaves [118] and pharmaceutical tablets [131].

SIMCA is popular for its simplicity but it is generally accepted that it has lower performance than other methods. Thus it is usual to accompany SIMCA with other methods and compare their performance. Sirven et al. [80] showed that PLS-DA (with a classification rate of 85.9%) was more sensitive than SIMCA (with a classification rate of 77.5%) in the classification of rocks. However, SIMCA proved to be more robust and effective in the classification of a few similar samples. Therefore, the authors suggested using SIMCA at the beginning of the

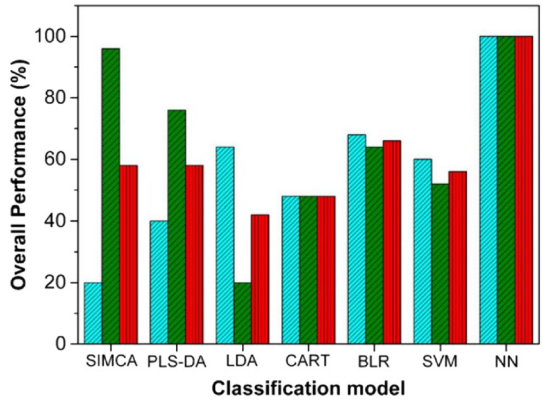

Fig. 7. This figure shows the overall performance of selected MVDA algorithms in classification of 25 bone samples; selected figures of merit are success rate (cyan), robustness (green) and accuracy (red). It is obvious that simple SIMCA is outperformed by other, sophisticated (even non-linear) algorithms. Abbreviations: PLS-DA – Partial Least Squares Discriminant Analysis, LDA – Linear Discriminant Analysis, CART – Classification and Regression Tree, BLR – Binary Logistic Regression, SVM – Support Vector Machines, NN – Neural Networks. Obtained from [183] with permission provided by Elsevier and Copyright Clearance Center, license number: 4347691418259. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ChemCam operation and the utilization of PLS-DA was planned for later stages when the library has grown. Moreover, they improved the classification rate to 100% when the SIMCA and PLS-DA algorithms were combined. Duchene et al. [179] showed similar results in the classification of archaeological material. Gottfried et al. [180] used PCA (99.27% of joint variance in 3 PCs) to visualize spectra of explosives based on 6 elemental line ratios. Then PCA of 29 variables (9 intensities of spectral lines and molecular bands and their ratios) was done. Classification of four samples was provided *via* SIMCA models using 5 to 8 PCs and *via* PLS-DA model using 17 latent variables. However, no figure of merits were provided which made the comparison of both MVDA algorithms difficult. Anderson et al. [55] compared SIMCA with other techniques in the classification of igneous rocks and showed SIMCA's lower performance.

3.5.3. SIMCA counterparts

Some more sophisticated and even non-linear algorithms were used. Godoi et al. [181] compared SIMCA, PLS-DA and KNN for a classification of toys, however their goal was to estimate the toxicity, *i.e.* the amount of Cd and Pb. Dingari et al. [182] compared the performance of SIMCA, PLS-DA and SVM, showing that a non-linear SVM algorithm provides higher classification performance than its linear counterparts. Myakalwar et al. [36] used SIMCA-derived models to provide an average correct classification rate of 88% of non-gated pharmaceutics spectra, whereas PLS-DA and ANN reached 100%. Moncayo et al. [183] compared 7 MVDA algorithms (including SIMCA and LDA done after PCA dimensionality reduction) for a classification of 55 bone samples of 5 individuals, see Fig. 7. They proved that ANN, due to its non-linear nature, provides the most sensitive and robust results.

PLS-DA is a very frequently used alternative to SIMCA for the classification of LIBS data [83,156]. Remus et al. [184] recommended using PLS-DA rather than PCA related algorithms because the latter were not capable to fully discriminate samples coming from different locations, *i.e.* to provide a provenance study. This obvious superiority of PLS-DA over PCA was summarized by Barker and Rayens [185].

Classification and feasibility of the provenance of conflict minerals were proved based on the conflict minerals' characteristic spectral fingerprints using PLS-DA [102]. Explosive residues were analyzed using LIBS as a thin films spread on some organic and inorganic substrates, consequently the collected spectra were classified by PLS-DA [186]. It is worth mentioning that other MVDA algorithms were also used for classification, such as linear and rank correlation [187], Linear Discriminant Analysis (LDA) [88], ICA [68], ANN [188–190], DFA [67,191,192], Graph Theory Method [193].

An interesting application of SIMCA arises due to its rather low performance (and in turn quite a high sensitivity to changes in the topology and structure of the data). Thus, SIMCA tends to be more biased than other MVDA algorithms (PLS-DA, SVM, SOM, ANN, *etc.*). Because of that, SIMCA can be utilized to study the effects of preprocessing (normalization, outliers filtering, *etc.*). It compares the changes in the data structure by resulting classification figures of merit [39,56].

For the sake of completeness, the combination of LIBS and Raman data aids to get a correct classification. Three data sets were created: i) LIBS, ii) Raman, iii) LIBS and Raman. Hoehse et al. [98] compared performance of SIMCA, PLS-DA and SVM in the classification inks and pigments. Prochazka et al. [99] used SOM to classify 6 bacterial strains.

4. Summary of publications (Table 1)

Table 1
This table summarizes overview of literature related to implementation of PCA in LIBS.

The way of PCA implementation	
Visualization	[4,43,44,78,79,83–88,90–96,99–101,110,111,116–118,131,136,137,141–43,145,147–149,157,168,170,173,183]
Classification	[36,39,41,42,55,76–78,80,89,98,112,118,119,124,125,131,136,154,160,164–166,171–174,176,178,179,180–183]
Regression (PCR)	[45,46,49,77,124,132,138,139–141]
Variable down-selection	[42,111,166]
Dimensionality reduction	[55,77,88,95,112,114,120,154,164–166,169,170,174,175,183]
Outliers filtering	[42,56,77,80]
Multivariate mapping	[81,104,105,107]
Application of LIBS	
Archaeology and forensics	[41,86,88,125,143,170,179]
Miscellaneous applications	[98,104,105,114,115,141,160,175]
Biology	[4,76,84,94,95,99,100,116,118,119,142,146,147,148,149,157,164,166,169,172,173]
Geology	[39,42–45,46,55,76–78,80,81,83,89,90–92,101,107,110,111,117,132,137–139,143,145,174,176,178]
Explosives and chemicals, nuclear materials	[79,93,96,112,119,144,154,159,167,180,186]
Pharmaceutics	[36,49,131,182]
Polymers	[87,124,168,181]
metals	[39,56,85,113,120,140,171]

5. Conclusion and future prospects

Based on the literature survey, LIBS combined with MVDA algorithms proved the capability to classify unknown samples and quantify analytes in many applications. However, the majority of reviewed articles represented only feasibility and preliminary studies. The impact of presented alterations in data pre-processing and MVDA algorithms on the resulting figures of merit was demonstrated on a limited number of samples, with a low number of spectra per sample, *etc.*

Generally, LIBS is on its rise and the hand-held as well as the table-top commercial systems attract more and more attention. The hand-held LIBS systems are used namely for the classification of steels and alloys, and for quantitative analyses. MVDA algorithms also penetrate this sector and will be decisive in terms of the performance and the resulting figures of merit.

Processing of LIBS data has undergone several dynamic developments in recent years. An increasing number of samples and detected representative spectra demands the implementation of MVDA algorithms in a more sophisticated way. The goal is to provide fast and robust data processing with desired selectivity and specificity, significantly reducing the dimensionality and information redundancy, yet retaining the most valuable information within the data matrix. Considering the trends and historical evolution, it seems that a progress in the data analysis will be continually made in the future. The implementation of MVDA algorithms will become a vital part of a routine LIBS analysis.

Historically, PCA and its variations were abundantly applied for multivariate analysis of LIBS data, such as visualization, clustering, classification, regression and recently also for multivariate surface mapping. It was also shown that both basic and advanced analyses of LIBS data can be provided using only PCA and its variations (SIMCA, PCR). Thus, when judicious experimental design and data preprocessing is provided, there is no need for switching to a more sophisticated MVDA algorithm (such as PLSR, PLS-DA, SVM, SOM, ANN, *etc.*). Therefore, it may be stated that PCA is essentially rooted in the LIBS community. Articles summarized in this review reflect the crucial issues that have to be considered when MVDA is of interest. Even though this review article was primarily aimed at PCA algorithm and its implementation for processing of LIBS data, the majority of presented suggestions, recommendations and conclusions can be considered in implementation of any other MVDA algorithm.

The reviewed articles represent mostly feasibility studies and preliminary results proving some capacities of LIBS technique. This was caused by the fact that there are no available data libraries, researches therefore established *ad hoc* libraries which were based most often only on limited number of collected samples. The analytical approach is different in every research group due to different applications, data structure, and experience of each group in relation to MVDA and the associated algorithm implementation. There are many contradictory suggestions and recommendations in the literature, which implies that there is no established receipt for a correct data analysis and it may not exist. Despite that, the data preprocessing and application of MVDA algorithms should follow several guidelines:

- **Theoretical background** - firm knowledge of the basic processes occurring in LIP and affecting the characteristic signal is the key. The MVDA algorithm should not be used as a “black box”.
- **Experimental design** - an optimized LIBS system and a coherent experimental design should deliver high-quality data, both repeatable and reproducible.
- **Robust dataset** - a statistically significant number of objects (*i.e.* spectra) describing the selected number of samples is important to cover a possible variation in the studied sample matrix and the fluctuation in the analysis itself.
- **Preprocessing** - well balanced data preprocessing should improve the data. At the same time, retaining the most of the variability within the dataset is desirable.
- **MVDA** – an optimization of the selected algorithm (avoiding overfitting) and an estimation of figures of merit should be done by using a model, validation and test subsets.
- **Detailed report** - the whole process (starting with samples, going through LIBS analysis and data collection, preprocessing and finally MVDA) should be described in detail. Figures of merit have to be estimated; this enables the means for a potential comparison with results of other research groups.

Looking at the issue from a future perspective, it will be necessary to deliver an optimized data processing algorithm. In the ideal case, this algorithm should enable to reduce the number of objects, which would decrease response time of the modelling and prediction step. All the same, it would be sensitive to the data and would conserve the information from matrix, minor, and trace elements. To summarize, the algorithms should be sensitive to outliers and should provide a high degree of specificity.

Acknowledgement

Authors affiliated with CEITEC would like to acknowledge the financial support obtained from the National Sustainability program - CEITEC NPU II (LQ1061) and supported also from the ERDFund-Project CEITEC Nano+ (CZ.02.1.01/0.0/0.0/16_013/0001728). PP is grateful to the Fulbright commission for supporting his research at the University of Florida (E0583833).

References

[1] J. Kaiser, K. Novotný, M.Z. Martin, A. Hrdlička, R. Malina, M. Hartl, R. Kizek, Trace elemental analysis by laser-induced breakdown spectroscopy—biological applications, Surf. Sci. Rep. 67 (11 − 12) (2012) 233–243.
[2] S.J. Rehse, H. Salimnia, A.W. Miziolek, Laser-induced breakdown spectroscopy (LIBS): an overview of recent progress and future potential for biomedical applications, J. Med. Eng. Technol. 36 (2) (2012) 77–89.
[3] D.J. Santos, L.C. Nunes, G.G.A. de Carvalho, M.d.S. Gomes, P.F. de Souza, F.d.O. Leme, F.J. Krug, Laser-induced breakdown spectroscopy for analysis of plant materials: a review, Spectrochim. Acta B At. Spectrosc. 71–72 (2012) 3–13.
[4] P. Pořízka, P. Prochazková, D. Prochazka, L. Sládková, J. Novotný, M. Petrilak, J. Kaiser, Algal biomass analysis by laser-based analytical techniques—a review, Sensors 14 (12) (2014) 17725–17752.
[5] R.S. Harmon, R.E. Russo, R.R. Hark, Applications of laser-induced breakdown spectroscopy for geochemical and environmental analysis: a comprehensive review, Spectrochim. Acta B At. Spectrosc. 87 (2013) 11–26.
[6] R. Noll, Laser-induced Breakdown Spectroscopy Fundamentals and Applications, Springer-Verlag Berlin Heidelberg, Heidelberg New York, 2012.
[7] J.O. Cáceres, F. Pelascini, V. Motto-Ros, S. Moncayo, F. Trichard, G. Panczer, J. Martín-Chivelet, Megapixel multi-elemental imaging by laser-induced breakdown spectroscopy, a technology with considerable potential for paleoclimate studies, Sci. Rep. 7 (1) (2017).
[8] F.J. Fortes, J.J. Laserna, The development of fieldable laser-induced breakdown spectrometer: no limits on the horizon, Spectrochim. Acta B At. Spectrosc. 65 (12) (2010) 975–990.
[9] J. Rakovský, P. Čermák, O. Musset, P. Veis, A review of the development of portable laser induced breakdown spectroscopy and its applications, Spectrochim. Acta B At. Spectrosc. 101 (2014) 269–287.
[10] S. Maurice, S.M. Clegg, R.C. Wiens, O. Gasnault, W. Rapin, O. Forni, A.R. Vasavada, ChemCam activities and discoveries during the nominal mission of the Mars Science Laboratory in Gale crater, Mars, J. Anal. At. Spectrom. 31 (4) (2016) 863–889.
[11] CHEMCAM Team, “ChemCam on Mars,” 2 February 2018. [Online], Available: http://www.msl-chemcam.com/index.php#.WnRdEahKswE.
[12] D.A. Cremers, L.J. Radziemski, Handbook of Laser-induced Breakdown Spectroscopy, 2nd Edition, John Wiley & Sons, Ltd, Chichester, West Sussex, 2013.
[13] A. Miziolek, V. Palleschi, I. Schechter, Laser Induced Breakdown Spectroscopy, Cambridge University Press, Cambridge, 2006.
[14] E. Tognoni, V. Palleschi, M. Corsi, G. Cristoforetti, Quantitative micro-analysis by laser-induced breakdown spectroscopy: a review of the experimental approaches, Spectrochim. Acta B At. Spectrosc. 57 (7) (2002) 1115–1130.
[15] C. Aragón, J.A. Aguilera, Characterization of laser induced plasmas by optical emission spectroscopy: a review of experiments and methods, Spectrochim. Acta B At. Spectrosc. 63 (9) (2008) 893–916.
[16] R.E. Russo, X. Mao, J.J. Gonzalez, V. Zorba, J. Yoo, Laser ablation in analytical chemistry, Anal. Chem. 85 (13) (2013) 6162–6177.
[17] D.W. Hahn, N. Omenetto, Laser-induced breakdown spectroscopy (LIBS), part II:

review of instrumental and methodological approaches to material analysis and applications to different fields, Appl. Spectrosc. 66 (4) (2012) 347–419.
[18] R. Gaudiuso, M. Dell'Aglio, O.D. Pascale, G.S. Senesi, A.D. Giacomo, Laser induced breakdown spectroscopy for elemental analysis in environmental, cultural heritage and space applications: a review of methods and results, Sensors 10 (12) (2010) 7434–7468.
[19] D.W. Hahn, N. Omenetto, Laser-induced breakdown spectroscopy (LIBS), part I: review of basic diagnostics and plasma—particle interactions: still-challenging issues within the analytical plasma community, Appl. Spectrosc. 64 (12) (2010) 335A–336A.
[20] A.P.M. Michel, Review: applications of single-shot laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 65 (3) (2010) 185–191.
[21] T. Kim, Z.G. Specht, P.S. Vary, C.T. Lin, Spectral fingerprints of bacterial strains by laser-induced breakdown spectroscopy, J. Phys. Chem. B 108 (17) (2004) 5477–5482.
[22] R.S. Harmon, J. Remus, N.J. McMillan, C. McManus, L. Collins, J.L.J. Gottfried, A.W. Miziolek, LIBS analysis of geomaterials: geochemical fingerprinting for the rapid analysis and discrimination of minerals, Appl. Geochem. 24 (6) (2009) 1125–1141.
[23] P. Geladi, Chemometrics in spectroscopy. Part 1. Classical chemometrics, Spectrochim. Acta B At. Spectrosc. 58 (5) (2003) 767–782.
[24] P. Geladi, B. Sethson, J. Nyström, T. Lillhonga, T. Lestander, J. Burger, Chemometrics in spectroscopy, Spectrochim. Acta B At. Spectrosc. 59 (9) (2004) 1347–1357.
[25] J.M. Andrade, M.J. Cal-Prieto, M.P. Gómez-Carracedo, A. Carlosena, D. Prada, A tutorial on multivariate calibration in atomic spectrometry techniques, J. Anal. At. Spectrom. 23 (1) (2008) 15–28.
[26] N. Kumar, A. Bansal, G.S. Sarma, R.K. Rawal, Chemometrics tools used in analytical chemistry: an overview, Talanta 123 (2014) 186–199.
[27] F.C.J. De Lucia, J.L. Gottfried, Rapid analysis of energetic and geo-materials using LIBS, Mater. Today 14 (6) (2011) 274–281.
[28] J. El Haddad, L. Canioni, B. Bousquet, Good practices in LIBS analysis: review and advices, Spectrochim. Acta B At. Spectrosc. 101 (2014) 171–182.
[29] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemom. 28 (4) (2014) 213–225.
[30] R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Antony Rowe Ltd, Chippenham, Wiltshire, 2003.
[31] R. Cangelosi, A. Goriely, Component retention in principal component analysis with application to cDNA microarray data, Biol. Direct 2 (1) (2007) 2.
[32] P. de Boves Harrington, Statistical validation of classification and calibration models using bootstrapped Latin partitions, Trends Anal. Chem. 25 (11) (2006) 1112–1124.
[33] E.H. van Veen, M.T.C. de Loos-Vollebregt, Application of mathematical procedures to background correction and multivariate analysis in inductively coupled plasma-optical emission spectrometry, Spectrochim. Acta B At. Spectrosc. 53 (5) (1998) 639–669.
[34] A.A. Urbas, S.J. Choquette, Automated spectral smoothing with spatially adaptive penalized least squares, Appl. Spectrosc. 65 (6) (2011) 665–677.
[35] P. Yaroshchyk, J.E. Eberhardt, Automatic correction of continuum background in laser-induced breakdown spectroscopy using a model-free algorithm, Spectrochim. Acta B 99 (2014) 138–149.
[36] A.K. Myakalwar, N.C. Dingari, R.R. Dasari, I. Barman, M.K. Gundawar, Non-gated laser induced breakdown spectroscopy provides a powerful segmentation tool on concomitant treatment of characteristic and continuum emission, PLoS One 9 (8) (2014) e103546.
[37] N.B. Zorov, A.A. Gorbatenko, T.A. Labutin, A.M. Popov, A review of normalization techniques in analytical atomic spectrometry with laser sampling: from single to multivariate correction, Spectrochim. Acta B At. Spectrosc. 65 (8) (2010) 642–657.
[38] J.P. Castro, E.R. Pereira-Filho, Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS), J. Anal. At. Spectrom. 31 (10) (2016) 2005–2014.
[39] P. Pořízka, J. Klus, A. Hrdlička, J. Vrábel, P. Škarková, D. Prochazka, J. Kaiser, Impact of laser-induced breakdown spectroscopy data normalization on multivariate classification accuracy, J. Anal. At. Spectrom. 32 (2) (2017) 277–288.
[40] T. Takahashi, B. Thornton, Quantitative methods for compensation of matrix effects and self-absorption in laser induced breakdown spectroscopy signals of solids, Spectrochim. Acta B 138 (2017) 31–42.
[41] F. Colao, R. Fantoni, P. Ortiz, M.A. Vazquez, J.M. Martin, R. Ortiz, N. Idris, Quarry identification of historical building materials by means of laser induced breakdown spectroscopy, X-ray fluorescence and chemometric analysis, Spectrochim. Acta B At. Spectrosc. 65 (8) (2010) 688–694.
[42] J.-B. Sirven, A. Pailloux, Y. M'Baye, N. Coulon, T. Alpettaz, S. Gossé, Towards the determination of the geographical origin of yellow cake samples by laser-induced breakdown spectroscopy and chemometrics, J. Anal. At. Spectrom. 24 (4) (2009) 451.
[43] J.-B. Sirven, B. Bousquet, L. Canioni, L. Sarger, Laser-induced breakdown spectroscopy of composite samples: comparison of advanced Chemometrics methods, Anal. Chem. 78 (5) (2006) 1462–1469.
[44] B. Bousquet, J.-B. Sirven, L. Canioni, Towards quantitative laser-induced breakdown spectroscopy analysis of soil samples, Spectrochim. Acta B At. Spectrosc. 62 (12) (2007) 1582–1589.
[45] D.L. Death, A.P. Cunningham, L.J. Pollard, Multi-element analysis of iron ore pellets by laser-induced breakdown spectroscopy and principal components regression, Spectrochim. Acta B At. Spectrosc. 63 (7) (2008) 763–769.
[46] D.L. Death, A.P. Cunningham, L.J. Pollard, Multi-element and mineralogical analysis of mineral ores using laser induced breakdown spectroscopy and chemometric analysis, Spectrochim. Acta B At. Spectrosc. 64 (10) (2009) 1048–1058.
[47] A.P.M. Michel, A.D. Chave, Analysis of laser-induced breakdown spectroscopy spectra: the case for extreme value statistics, Spectrochim. Acta B At. Spectrosc. 62 (12) (2007) 1370–1378.
[48] J. Klus, P. Pořízka, D. Prochazka, J. Novotný, K. Novotný, J. Kaiser, Effect of experimental parameters and resulting analytical signal statistics in laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 126 (2016) 6–10.
[49] F.R. Doucet, P.J. Faustino, M. Sabsabi, R.C. Lyon, Quantitative molecular analysis with molecular bands emission using laser-induced breakdown spectroscopy and chemometrics, J. Anal. At. Spectrom. 23 (5) (2008) 694.
[50] National Institute of Standards and Technology, "National Institute of Standards and Technology," 2 February 2018. [Online], Available: https://www.nist.gov/.
[51] C. Heise, "Atomic spectral line database," 2 February 2018. [Online], Available: https://www.cfa.harvard.edu/amp/ampdata/kurucz23/sekur.html.
[52] Applied Photonics, "Applied Photonics," 2 February 2018. [Online], Available: http://www.appliedphotonics.co.uk/libs/capabilities_libs.htm.
[53] AtomTrace®, "AtomTrace," 2 February 2018. [Online], Available: https://www.atomtrace.com/.
[54] G. Amato, G. Cristoforetti, S. Legnaioli, G. Lorenzetti, V. Palleschi, F. Sorrentino, E. Tognoni, Progress towards an unassisted element identification from laser induced breakdown spectra with automatic ranking techniques inspired by text retrieval, Spectrochim. Acta B At. Spectrosc. 65 (8) (2010) 664–670.
[55] R.B. Anderson, J.F. Bell, R.C. Wiens, R.V. Morris, S.M. Clegg, Clustering and training set selection methods for improving the accuracy of quantitative laser induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 70 (2012) 24–32.
[56] P. Pořízka, J. Klus, D. Prochazka, E. Képeš, A. Hrdlička, J. Novotný, J. Kaiser, Laser-induced breakdown spectroscopy coupled with chemometrics for the analysis of steel: the issue of spectral outliers filtering, Spectrochim. Acta B At. Spectrosc. 123 (2016) 114–120.
[57] J. Carlson, A. Wysoczanski, E. Voigtman, Limits of quantitation — yet another suggestion, Spectrochim. Acta B At. Spectrosc. 96 (2014) 69–73.
[58] K. Pearson, On lines and planes of closest fit to systems of points in space, Philos. Mag. 2 (11) (1901) 559–572.
[59] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (6) (1933) 417–441.
[60] M. Kendall, Multivariate Analysis, MacMillan, New York, 1980.
[61] L. Hatcher, N. Rourke, A Step-by-step Approach to Using SAS® for Factor Analysis and Structural Equation Modeling, SAS Institute, Cary, North Carolina, 2013.
[62] W. Wu, D.L. Massart, S. de Jong, Kernel-PCA algorithms for wide data part II: fast cross-validation and application in classification of NIR data, Chemom. Intell. Lab. Syst. 37 (2) (1997) 271–280.
[63] M. Hubert, P.J. Rousseeuw, S. Verboven, A fast method for robust principal components with applications to chemometrics, Chemom. Intell. Lab. Syst. 60 (1–2) (2002) 101–111.
[64] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: a new approach to robust principal component analysis, Technometrics 47 (1) (2005) 64–79.
[65] J. Diedrich, S.J. Rehse, S. Palchaudhuri, Escherichia coli identification and strain discrimination using nanosecond laser-induced breakdown spectroscopy, Appl. Phys. Lett. 90 (16) (2007) 163901.
[66] J. Diedrich, S.J. Rehse, S. Palchaudhuri, Pathogenic Escherichia colistrain discrimination using laser-induced breakdown spectroscopy, J. Appl. Phys. 102 (1) (2007) 14702.
[67] R.A. Putnam, Q.I. Mohaidat, A. Daabous, S.J. Rehse, A comparison of multivariate analysis techniques and variable selection strategies in a laser-induced breakdown spectroscopy bacterial classification, Spectrochim. Acta B At. Spectrosc. 87 (2013) 161–167.
[68] O. Forni, S. Maurice, O. Gasnault, R.C. Wiens, A. Cousin, S.M. Clegg, J. Lasue, Independent component analysis classification of laser induced breakdown spectroscopy spectra, Spectrochim. Acta B At. Spectrosc. 86 (2013) 31–41.
[69] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybern. 43 (1) (1982) 59–69.
[70] S. Kachigan, Multivariate Statistical Analysis: A Conceptual Introduction, New York: Radius Press, 1991.
[71] A.B. Costello, J.W. Osborne, Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis, Pract. Assess. Res. Eval. 10 (2005) 173–178.
[72] I. Jolliffe, Principal Component Analysis, Second Edition, Springer, New York, 2002.
[73] R.B. Cattell, The scree test for the number of factors, Multivar. Behav. Res. 1 (2) (1966) 245–276.
[74] H.F. Kaiser, The application of electronic computers to factor analysis, Educ. Psychol. Meas. 20 (1) (1960) 141–151.
[75] M. Defernez, E.K. Kemsley, The use and misuse of chemometrics for treating classification problems, Trends Anal. Chem. 16 (4) (1997) 216–221.
[76] B. Bousquet, G. Travaillé, A. Ismaël, L. Canioni, K. Michel-Le Pierrès, E. Brasseur, S. Belbèze, Development of a mobile system based on laser-induced breakdown spectroscopy and dedicated to in situ analysis of polluted soils, Spectrochim. Acta B At. Spectrosc. 63 (10) (2008) 1085–1090.
[77] P. Pořízka, A. Demidov, J. Kaiser, J. Keivanian, I. Gornushkin, U. Panne, J. Riedel, Laser-induced breakdown spectroscopy for in situ qualitative and quantitative analysis of mineral ores, Spectrochim. Acta B At. Spectrosc. 101 (2014) 155–163.
[78] P. Pořízka, J. Klus, J. Mašek, M. Rajnoha, D. Prochazka, P. Modlitbová, J. Kaiser, Multivariate classification of echellograms: a new perspective in laser-induced

breakdown spectroscopy analysis, Sci. Rep. 7 (1) (2017).
[79] J.F.C. De Lucia, J.L. Gottfried, C.A. Munson, A.W. Miziolek, Multivariate analysis of standoff laser-induced breakdown spectroscopy spectra for classification of explosive-containing residues, Appl. Opt. 47 (31) (2008) G112.
[80] J.-B. Sirven, B. Sallé, P. Mauchien, J.-L. Lacour, S. Maurice, G. Manhès, Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods, J. Anal. At. Spectrom. 22 (12) (2007) 1471.
[81] J. Klus, P. Mikysek, D. Prochazka, P. Pořízka, P. Prochazková, J. Novotný, J. Kaiser, Multivariate approach to the chemical mapping of uranium in sandstone-hosted uranium ores analyzed using double pulse laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 123 (2016) 143–149.
[82] V. Lazic, A. Palucci, S. Jovicevic, C. Poggi, E. Buono, Analysis of explosive and other organic residues by laser induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 64 (10) (2009) 1028–1039.
[83] J.L. Gottfried, R.S. Harmon, F.C.J. De Lucia, A.W. Miziolek, Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification, Spectrochim. Acta B At. Spectrosc. 64 (10) (2009) 1009–1019.
[84] D.E. Lewis, J. Martinez, C.A. Akpovo, L. Johnson, A. Chauhan, M.D. Edington, Discrimination of bacteria from Jamaican bauxite soils using laser-induced breakdown spectroscopy, Anal. Bioanal. Chem. 401 (7) (2011) 2225–2236.
[85] S.R. Goode, S.L. Morgan, R. Hoskins, A. Oxsher, Identifying alloys by laser-induced breakdown spectroscopy with a time-resolved high resolution echelle spectrometer, J. Anal. At. Spectrom. 15 (9) (2000) 1133–1138.
[86] A. Erdem, A. Çilingiroğlu, A. Giakoumaki, M. Castanys, E. Kartsonaki, C. Fotakis, D. Anglos, Characterization of Iron age pottery from eastern Turkey by laser- induced breakdown spectroscopy (LIBS), J. Archaeol. Sci. 35 (9) (2008) 2486–2494.
[87] S. Grégoire, M. Boudinet, F. Pelascini, F. Surma, V. Detalle, Y. Holl, Laser-induced breakdown spectroscopy for polymer identification, Anal. Bioanal. Chem. 400 (10) (2011) 3331–3340.
[88] G. Vítková, K. Novotný, L. Prokeš, A. Hrdlička, J. Kaiser, J. Novotný, D. Prochazka, Fast identification of biominerals by means of stand-off laser-induced breakdown spectroscopy using linear discriminant analysis and artificial neural networks, Spectrochim. Acta B At. Spectrosc. 73 (2012) 1–6.
[89] P. Pořízka, J. Klus, D. Prochazka, G. Vítková, M. Brada, J. Novotný, J. Kaiser, Assessment of the most effective part of echelle laser-induced plasma spectra for further classification using Czerny-Turner spectrometer, Spectrochim. Acta B At. Spectrosc. 124 (2016) 116–123.
[90] N.L. Lanza, R.C. Wiens, S.M. Clegg, A.M. Ollila, S.D. Humphries, H.E. Newsom, J.E. Barefield, Calibrating the ChemCam laser-induced breakdown spectroscopy instrument for carbonate minerals on Mars, Appl. Opt. 49 (13) (2010) C211.
[91] C.E. McManus, N.J. McMillan, R.S. Harmon, R.C. Whitmore, J.F.C. De Lucia, A.W. Miziolek, Use of laser induced breakdown spectroscopy in the determination of gem provenance: beryls, Appl. Opt. 47 (31) (2008) G72.
[92] D.C. Alvey, K. Morton, R.S. Harmon, J.L. Gottfried, J.J. Remus, L.M. Collins, M.A. Wise, Laser-induced breakdown spectroscopy-based geochemical fingerprinting for the rapid analysis and discrimination of minerals: the example of garnet, Appl. Opt. 49 (2010) C168–C180.
[93] F.C.J. De Lucia, J.L. Gottfried, Characterization of a series of nitrogen-rich molecules using laser induced breakdown spectroscopy, Propellants, Explos., Pyrotech. 35 (3) (2010) 268–277.
[94] N. Labbé, I.M. Swamidoss, N. André, M.Z. Martin, T.M. Young, T.G. Rials, Extraction of information from laser-induced breakdown spectroscopy spectral data by multivariate analysis, Appl. Opt. 47 (31) (2008) G158.
[95] D. Merdes, J. Suhan, J.M. Keay, D.M. Hadka, W.R. Bradley, The investigation of laser-induced breakdown spectroscopy for detection of biological contaminants on surfaces, Spectroscopy 22 (4) (2007) 28–38.
[96] L. Landström, A. Larsson, P.-Å. Gradmark, L. Örebrand, P.O. Andersson, P. Wästerby, T. Tjärnhage, Detection and Monitoring of CWA and BWA Using LIBS, Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XV, Baltimore, Maryland, United States, (2014).
[97] P. Pořízka, S. Kaski, A. Hrdlička, P. Modlitbová, L. Sládková, H. Häkkänen, J. Kaiser, Detection of fluorine using laser-induced breakdown spectroscopy and Raman spectroscopy, J. Anal. At. Spectrom. 32 (10) (2017) 1966–1974.
[98] M. Hoehse, A. Paul, I. Gornushkin, U. Panne, Multivariate classification of pigments and inks using combined Raman spectroscopy and LIBS, Anal. Bioanal. Chem. 402 (4) (2012) 1443–1450.
[99] D. Prochazka, M. Mazura, O. Samek, K. Rebrošová, P. Pořízka, J. Klus, J. Kaiser, Combination of laser-induced breakdown spectroscopy and Raman spectroscopy for multivariate classification of bacteria, Spectrochim. Acta B At. Spectrosc. 139 (2018) 6–12.
[100] A. Larsson, H. Andersson, L. Landström, Impact of data reduction on multivariate classification models built on spectral data from bio-samples, J. Anal. At. Spectrom. 30 (5) (2015) 1117–1127.
[101] J. Lasue, R.C. Wiens, T.F. Stepinski, O. Forni, S.M. Clegg, S. Maurice, Nonlinear mapping technique for data visualization and clustering assessment of LIBS data: application to ChemCam data, Anal. Bioanal. Chem. 400 (10) (2011) 3247–3260.
[102] R.S. Harmon, K.M. Shughrue, J.J. Remus, M.A. Wise, L.J. East, R.R. Hark, Can the provenance of the conflict minerals columbite and tantalite be ascertained by laser-induced breakdown spectroscopy? Anal. Bioanal. Chem. 400 (10) (2011) 3377–3382.
[103] J. Klus, P. Pořízka, D. Prochazka, P. Mikysek, J. Novotný, K. Novotný, J. Kaiser, Application of self-organizing maps to the study of U-Zr-Ti-Nb distribution in sandstone-hosted uranium ores, Spectrochim. Acta B At. Spectrosc. 131 (2017) 66–73.
[104] J. Amador-Hernández, J.M. Fernández-Romero, M.D. Luque de Castro, In-depth characterization of screen-printed electrodes by laser-induced breakdown spectrometry and pattern recognition, Surf. Interface Anal. 31 (4) (2001) 313–320.
[105] J. Amador-Hernández, J. Fernández-Romero, M. Luque de Castro, Three-dimensional analysis of screen-printed electrodes by laser induced breakdown spectrometry and pattern recognition, Anal. Chim. Acta 435 (2) (2001) 227–238.
[106] R.C. Chinni, D.A. Cremers, L.J. Radziemski, M. Bostian, C. Navarro-Northrup, Detection of uranium using laser-induced breakdown spectroscopy, Appl. Spectrosc. 63 (11) (2009) 1238–1250.
[107] S. Romppanen, H. Häkkänen, S. Kaski, Singular value decomposition approach to the yttrium occurrence in mineral maps of rare earth element ores using laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 134 (2017) 69–74.
[108] S. Moncayo, L. Duponchel, N. Mousavipak, G. Panczer, F. Trichard, B. Bousquet, V. Motto-Ros, Exploration of megapixel hyperspectral LIBS images using principal component analysis, J. Anal. At. Spectrom. 33 (2018) 210–220.
[109] S. Pagnotta, M. Lezzerini, L. Ripoll-Seguer, M. Hidalgo, E. Grifoni, S. Legnaioli, G. Lorenzetti, F. Poggialini, V. Palleschi, Micro-laser-induced breakdown spectroscopy (micro-LIBS) study on ancient roman mortars, Appl. Spectrosc. 71 (4) (2017) 721–727.
[110] N.L. Lanza, S.M. Clegg, R.C. Wiens, R.E. Mcinroy, H.E. Newsom, M.D. Deans, Examining natural rock varnish and weathering rinds with laser-induced breakdown spectroscopy for application to ChemCam on Mars, Appl. Opt. 51 (7) (2012) B74.
[111] J.-B. Sirven, B. Bousquet, L. Canioni, L. Sarger, S. Tellier, M. Potin-Gautier, I.L. Hecho, Qualitative and quantitative investigation of chromium-polluted soils by laser-induced breakdown spectroscopy combined with neural networks analysis, Anal. Bioanal. Chem. 385 (2) (2006) 256–262.
[112] T.K. Sahoo, A. Negi, M.K. Gundawar, Study of Preprocessing Sensitivity on Laser Induced Breakdown Spectroscopy (LIBS) Spectral Classification, International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India, (2015).
[113] H.Y. Kong, L.X. Sun, J.T. Hu, Y. Xin, Z.B. Cong, A comparative study of two data reduction methods for steel classification based on LIBS, Appl. Mech. Mater. 644–650 (2014) 4722–4725.
[114] Y. Lee, K.-S. Ham, S.-H. Han, J. Yoo, S. Jeong, Revealing discriminating power of the elements in edible sea salts: line-intensity correlation analysis from laser-induced plasma emission spectra, Spectrochim. Acta B At. Spectrosc. 101 (2014) 57–67.
[115] Y. Lee, S.-H. Nam, K.-S. Ham, J. Gonzalez, D. Oropeza, D.J. Quarles, R.E. Russo, Multivariate classification of edible salts: simultaneous laser-induced breakdown spectroscopy and laser-ablation inductively coupled plasma mass spectrometry analysis, Spectrochim. Acta B At. Spectrosc. 118 (2016) 102–111.
[116] A.C. Samuels, F.C. DeLucia, K.L. McNesby, A.W. Miziolek, Laser-induced breakdown spectroscopy of bacterial spores, molds, pollens, and protein: initial studies of discrimination potential, Appl. Opt. 42 (30) (2003) 6205.
[117] M.D. Dyar, J.M. Tucker, S. Humphries, S.M. Clegg, R.C. Wiens, M.D. Lane, Strategies for Mars remote laser-induced breakdown spectroscopy analysis of sulfur in geological samples, Spectrochim. Acta B At. Spectrosc. 66 (1) (2011) 39–56.
[118] F.M.V. Pereira, D.M.B.P. Milori, A.L. Venâncio, M.d.S.T. Russo, P.K. Martins, J. Freitas-Astúa, Evaluation of the effects of Candidatus Liberibacter asiaticus on inoculated citrus plants using laser-induced breakdown spectroscopy (LIBS) and chemometrics tools, Talanta 83 (2) (2010) 351–356.
[119] C.A. Munson, F.C.J. De Lucia, T. Piehler, K.L. McNesby, A.W. Miziolek, Investigation of statistics strategies for improving the discriminating power of laser-induced breakdown spectroscopy for chemical and biological warfare agent simulants, Spectrochim. Acta B At. Spectrosc. 60 (7–8) (2005) 1217–1224.
[120] F. Anabitarte, J. Mirapeix, O.M.C. Portilla, J.M. Lopez-Higuera, A. Cobo, Sensor for the detection of protective coating traces on boron steel with Aluminium–silicon covering by means of laser-induced breakdown spectroscopy and support vector machines, IEEE Sensors J. 12 (1) (2012) 64–70.
[121] D. Marcos-Martinez, J.A. Ayala, R.C. Izquierdo-Hornillos, F.J.M. de Villena, J.O. Caceres, Identification and discrimination of bacterial strains by laser induced breakdown spectroscopy and neural networks, Talanta 84 (3) (2011) 730–737.
[122] I.-G. Chong, C.-H. Jun, Performance of some variable selection methods when multicollinearity is present, Chemom. Intell. Lab. Syst. 78 (1–2) (2005) 103–112.
[123] F.C.J. De Lucia, J.L. Gottfried, Influence of variable selection on partial least squares discriminant analysis models for explosive residue classification, Spectrochim. Acta B At. Spectrosc. 66 (2) (2011) 122–128.
[124] H. Fink, U. Panne, R. Niessner, Process analysis of recycled thermoplasts from consumer electronics by laser-induced plasma spectroscopy, Anal. Chem. 74 (17) (2002) 4334–4342.
[125] M. Corsi, G. Cristoforetti, M. Giuffrida, M. Hidalgo, S. Legnaioli, L. Masotti, V. Palleschi, A. Salvetti, E. Tognoni, C. Vellabona, A. Zanini, Archaeometric analysis of ancient copper artefacts by laser-induced breakdown spectroscopy technique, Microchim. Acta 152 (1–2) (2005) 105–111.
[126] A. Ciucci, M. Corsi, V. Palleschi, S. Rastelli, A. Salvetti, E. Tognoni, New procedure for quantitative elemental analysis by laser-induced plasma spectroscopy, Appl. Spectrosc. 53 (8) (1999) 960–964.
[127] A.K. Myakalwar, N. Spegazzini, C. Zhang, S. Anubham, R.R. Dasari, I. Barman, M.K. Gundawar, Less is more: avoiding the LIBS dimensionality curse through judicious feature selection for explosive detection, Sci. Rep. 5 (1) (2015).
[128] R.A. Multari, D.A. Cremers, J.M. Dupre, J.E. Gustafson, The use of laser-induced breakdown spectroscopy for distinguishing between bacterial pathogen species and strains, Appl. Spectrosc. 64 (7) (2010) 750–759.
[129] R. Multari and D. Cremers, “Methods For Forming Recognition Algorithms For

Laser-Induced Breakdown Spectroscopy". Patent US20110246145 A1, 2012.
[130] I.B. Gornushkin, U. Panne, J.D. Winefordner, Linear correlation for identification of materials by laser induced breakdown spectroscopy: improvement via spectral filtering and masking, Spectrochim. Acta B At. Spectrosc. 64 (10) (2009) 1040–1047.
[131] A.K. Myakalwar, S. Sreedhar, I. Barman, N.C. Dingari, S. Venugopal Rao, P. Prem Kiran, G. Manoj Kumar, Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis, Talanta 87 (2011) 53–59.
[132] P. Yaroshchyk, D.L. Death, S.J. Spencer, Comparison of principal components regression, partial least squares regression, multi-block partial least squares regression, and serial partial least squares regression algorithms for the analysis of Fe in iron ore using LIBS, J. Anal. At. Spectrom. 27 (1) (2012) 92–98.
[133] J.E. Carranza, D.W. Hahn, Sampling statistics and considerations for single-shot analysis using laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 57 (4) (2002) 779–790.
[134] V. Lazic, F. Colao, R. Fantoni, V. Spizzicchino, Laser-induced breakdown spectroscopy in water: improvement of the detection threshold by signal processing, Spectrochim. Acta B At. Spectrosc. 60 (7–8) (2005) 1002–1013.
[135] D. Díaz, D.W. Hahn, A. Molina, Quantification of gold and silver in minerals by laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 136 (2017) 106–115.
[136] S.-J. Choi, K.-J. Lee, J.J. Yoh, Quantitative laser-induced breakdown spectroscopy of standard reference materials of various categories, Appl. Phys. B Lasers Opt. 113 (3) (2013) 379–388.
[137] J. El Haddad, M. Villot-Kadri, A. Ismaël, G. Gallou, K. Michel, D. Bruyère, B. Bousquet, Artificial neural network for on-site quantitative analysis of soils using laser induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 79–80 (2013) 51–57.
[138] R. Wisbrun, I. Schechter, R. Niessner, H. Schroeder, K.L. Kompa, Detector for trace elemental analysis of solid environmental samples by laser plasma spectroscopy, Anal. Chem. 66 (18) (1994) 2964–2975.
[139] M.M. Tripathi, K.E. Eseller, F.-Y. Yueh, J.P. Singh, Multivariate calibration of spectra obtained by laser induced breakdown spectroscopy of plutonium oxide surrogate residues, Spectrochim. Acta B At. Spectrosc. 64 (11–12) (2009) 1212–1218.
[140] F.R. Doucet, T.F. Belliveau, J.-L. Fortier, J. Hubert, Use of chemometrics and laser-induced breakdown spectroscopy for quantitative analysis of major and minor elements in aluminum alloys, Appl. Spectrosc. 61 (3) (2007) 327–332.
[141] P. Devangad, V.K. Unnikrishnan, M.M. Tamboli, K.M.M. Shameem, R. Nayak, K.S. Choudhari, C. Santhosh, Quantification of Mn in glass matrices using laser induced breakdown spectroscopy (LIBS) combined with chemometric approaches, Anal. Methods 8 (39) (2016) 7177–7184.
[142] M.Z. Martin, N. Labbé, T.G. Rials, S.D. Wullschleger, Analysis of preservative-treated wood by multivariate analysis of laser-induced breakdown spectroscopy spectra, Spectrochim. Acta B At. Spectrosc. 60 (7–8) (2005) 1179–1185.
[143] M.Z. Martin, N. Labbé, N. André, R. Harris, M. Ebinger, S.D. Wullschleger, A.A. Vass, High resolution applications of laser-induced breakdown spectroscopy for environmental and forensic applications, Spectrochim. Acta B At. Spectrosc. 62 (12) (2007) 1426–1432.
[144] M.Z. Martin, S. Allman, D.J. Brice, R.C. Martin, N.O. Andre, Exploring laser-induced breakdown spectroscopy for nuclear materials analysis and in-situ applications, Spectrochim. Acta B At. Spectrosc. 74–75 (2012) 177–183.
[145] R.B. Anderson, R.V. Morris, S.M. Clegg, J.F. Bell, R.C. Wiens, S.D. Humphries, R. Mcinroy, The influence of multivariate analysis methods and target grain size on the accuracy of remote quantitative chemical analysis of rocks using laser induced breakdown spectroscopy, Icarus 215 (2) (2011) 608–627.
[146] R.A. Multari, D.A. Cremers, J.A.M. Dupre, J.E. Gustafson, Detection of biological contaminants on foods and food surfaces using laser-induced breakdown spectroscopy (LIBS), J. Agric. Food Chem. 61 (36) (2013) 8687–8694.
[147] G. Bilge, H.M. Velioglu, B. Sezer, K.E. Eseller, I.H. Boyaci, Identification of meat species by using laser-induced breakdown spectroscopy, Meat Sci. 119 (2016) 118–122.
[148] G. Bilge, B. Sezer, K.E. Eseller, H. Berberoglu, A. Topcu, I.H. Boyaci, Determination of whey adulteration in milk powder by using laser induced breakdown spectroscopy, Food Chem. 212 (2016) 183–188.
[149] S. Moncayo, S. Manzoor, J.D. Rosales, J. Anzano, J.O. Caceres, Qualitative and quantitative analysis of milk for the detection of adulteration by laser induced breakdown spectroscopy (LIBS), Food Chem. 232 (2017) 322–328.
[150] B. Sallé, J.-L. Lacour, P. Mauchien, P. Fichet, S. Maurice, G. Manhès, Comparative study of different methodologies for quantitative rock analysis by laser-induced breakdown spectroscopy in a simulated Martian atmosphere, Spectrochim. Acta B At. Spectrosc. 61 (3) (2006) 301–313.
[151] J.W.B. Braga, L.C. Trevizan, L.C. Nunes, I.A. Rufini, D.J. Santos, F.J. Krug, Comparison of univariate and multivariate calibration for the determination of micronutrients in pellets of plant materials by laser induced breakdown spectrometry, Spectrochim. Acta B At. Spectrosc. 65 (1) (2010) 66–74.
[152] P. Yaroshchyk, D.L. Death, S.J. Spencer, Quantitative measurements of loss on ignition in Iron ore using laser-induced breakdown spectroscopy and partial least squares regression analysis, Appl. Spectrosc. 64 (12) (2010) 1335–1341.
[153] C. Gottlieb, S. Millar, T. Günther, G. Wilsch, Revealing hidden spectral information of chlorine and sulfur in data of a mobile laser-induced breakdown spectroscopy system using chemometrics, Spectrochim. Acta B At. Spectrosc. 132 (2017) 43–49.
[154] C. Bohling, K. Hohmann, D. Scheel, C. Bauer, W. Schippers, J. Burgmeier, W. Schade, All-fiber-coupled laser-induced breakdown spectroscopy sensor for hazardous materials analysis, Spectrochim. Acta B At. Spectrosc. 62 (12) (2007) 1519–1527.
[155] C. Mcmanus, J. Dowe, T. Likes, J. Dowe, P. Mcmanus, N. Winkless, A. Shelton, K. Baughn, R. Baughn and R. Orr, "Methods and systems for analyzing samples". Patent WO2014150696 A1, 2014.
[156] A.M. Ollila, J. Lasue, H.E. Newsom, R.A. Multari, R.C. Wiens, S.M. Clegg, Comparison of two partial least squares-discriminant analysis algorithms for identifying geological samples with the ChemCam laser-induced breakdown spectroscopy instrument, Appl. Opt. 51 (7) (2012) B130.
[157] J.D. Hybl, G.A. Lithgow, S.G. Buckley, Laser-induced breakdown spectroscopy detection and classification of biological aerosols, Appl. Spectrosc. 57 (10) (2003) 1207–1215.
[158] M. Baudelet, J. Yu, M. Bossu, J. Jovelet, J.-P. Wolf, T. Amodeo, P. Laloi, Discrimination of microbiological samples using femtosecond laser-induced breakdown spectroscopy, Appl. Phys. Lett. 89 (16) (2006) 163903.
[159] F.C.J. De Lucia, J.L. Gottfried, C.A. Munson, A.W. Miziolek, Double pulse laser-induced breakdown spectroscopy of explosives: initial study towards improved discrimination, Spectrochim. Acta B At. Spectrosc. 62 (12) (2007) 1399–1404.
[160] H. Xia, M.C.M. Bakker, Reliable classification of moving waste materials with LIBS in concrete recycling, Talanta 120 (2014) 239–247.
[161] B.K. Lavine, C.E. Davidson, A.J. Moores, P.R. Griffiths, Raman spectroscopy and genetic algorithms for the classification of wood types, Appl. Spectrosc. 55 (8) (2001) 960–966.
[162] F.L. Martin, M.J. German, E. Wit, T. Fearn, N. Ragavan, H.M. Pollock, Identifying variables responsible for clustering in discriminant analysis of data from infrared microspectroscopy of a biological sample, J. Comput. Biol. 14 (9) (2007) 1176–1184.
[163] B. Vekemans, L. Vincze, F.E. Brenker, F. Adams, Processing of three-dimensional microscopic X-ray fluorescence data, J. Anal. At. Spectrom. 19 (10) (2004) 1302.
[164] O. Samek, H.H. Telle, D.C. Beddows, Laser-induced breakdown spectroscopy: a tool for real-time, in vitro and in vivo identification of carious teeth, BMC Oral Health 1 (1) (2001).
[165] O. Samek, V. Krzyžánek, D. Beddows, H. Telle, J. Kaiser, M. Liška, Material Identification Using Laser Spectroscopy, Computer Analysis of Images and Patterns: 9th International Conference, LNCS 2124, Warsaw, Poland, (2001).
[166] F.-Y. Yueh, H. Zheng, J.P. Singh, S. Burgess, Preliminary evaluation of laser-induced breakdown spectroscopy for tissue classification, Spectrochim. Acta B At. Spectrosc. 64 (10) (2009) 1059–1067.
[167] J.L. Gottfried, F.C.J. De Lucia, C.A. Munson, A.W. Miziolek, Double-pulse standoff laser-induced breakdown spectroscopy for versatile hazardous materials detection, Spectrochim. Acta B At. Spectrosc. 62 (12) (2007) 1405–1411.
[168] V.K. Unnikrishnan, K.S. Choudhari, S.D. Kulkarni, R. Nayak, V.B. Kartha, C. Santhosh, Analytical predictive capabilities of laser induced breakdown spectroscopy (LIBS) with principal component analysis (PCA) for plastic classification, RSC Adv. 3 (48) (2013) 25872.
[169] D. Pokrajac, A. Lazarevic, V. Kecman, A. Marcano, Y. Markushin, T. Vance, N. Melikechi, Automatic classification of laser-induced breakdown spectroscopy (LIBS) data of protein biomarker solutions, Appl. Spectrosc. 68 (9) (2014) 1067–1075.
[170] G. Vítková, L. Prokeš, K. Novotný, P. Pořízka, J. Novotný, D. Všianský, J. Kaiser, Comparative study on fast classification of brick samples by combination of principal component analysis and linear discriminant analysis using stand-off and table-top laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 101 (2014) 191–199.
[171] S. Merk, C. Scholz, S. Florek, D. Mory, Increased identification rate of scrap metal using laser induced breakdown spectroscopy echelle spectra, Spectrochim. Acta B At. Spectrosc. 112 (2015) 10–15.
[172] T. Vance, N. Reljin, A. Lazarevic, D. Pokrajac, V. Kecman, N. Melikechi, S. McDaniel, Classification of LIBS Protein Spectra Using Support Vector Machines and Adaptive Local Hyperplanes, The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, (2010).
[173] J.R. Cordeiro, M.I.V. Martinez, R.W.C. Li, A.P. Cardoso, L.C. Nunes, F.J. Krug, J. Gruber, Identification of four wood species by an electronic nose and by LIBS, Int. J. Electrochem. (2012) 1–5.
[174] G. Kim, J. Kwak, K.-R. Kim, H. Lee, K.-W. Kim, H. Yang, K. Park, Rapid detection of soils contaminated with heavy metals and oils by laser induced breakdown spectroscopy (LIBS), J. Hazard. Mater. 263 (2013) 754–760.
[175] G. Park, H. Yoo, Y. Gong, S. Cui, S.-H. Nam, K.-S. Ham, Y. Lee, Feasibility of rapid classification of edible salts by a compact low-cost laser-induced breakdown spectroscopy device, Bull. Kor. Chem. Soc. 36 (1) (2015) 189–197.
[176] M.J.C. Pontes, J. Cortez, R.K.H. Galvão, C. Pasquini, M.C.U. Araújo, R.M. Coelho, B.E. Madari, Classification of Brazilian soils by using LIBS and variable selection in the wavelet domain, Anal. Chim. Acta 642 (1–2) (2009) 12–18.
[177] S. Wold, M. Sjöström, SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy, ACS Symposium Series, (1977), pp. 243–282.
[178] S.M. Clegg, E. Sklute, M.D. Dyar, J.E. Barefield, R.C. Wiens, Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques, Spectrochim. Acta B At. Spectrosc. 64 (1) (2009) 79–88.
[179] S. Duchene, V. Detalle, R. Bruder, J. Sirven, Chemometrics and laser induced breakdown spectroscopy (LIBS) analyses for identification of wall paintings pigments, Curr. Anal. Chem. 6 (1) (2010) 60–65.
[180] J.L. Gottfried, J.F.C. De Lucia, C.A. Munson, A.W. Miziolek, Strategies for residue explosives detection using laser-induced breakdown spectroscopy, J. Anal. At. Spectrom. 23 (2) (2008) 205–216.
[181] Q. Godoi, F.O. Leme, L.C. Trevizan, E.R. Pereira Filho, I.A. Rufini, D.J. Santos, F.J. Krug, Laser-induced breakdown spectroscopy and chemometrics for

classification of toys relying on toxic elements, Spectrochim. Acta B At. Spectrosc. 66 (2) (2011) 138–143.
[182] N.C. Dingari, I. Barman, A.K. Myakalwar, S.P. Tewari, M. Kumar Gundawar, Incorporation of support vector machines in the LIBS toolbox for sensitive and robust classification amidst unexpected sample and system variability, Anal. Chem. 84 (6) (2012) 2686–2694.
[183] S. Moncayo, S. Manzoor, F. Navarro-Villoslada, J.O. Caceres, Evaluation of supervised chemometric methods for sample classification by laser induced breakdown spectroscopy, Chemom. Intell. Lab. Syst. 146 (2015) 354–364.
[184] J.J. Remus, J.L. Gottfried, R.S. Harmon, A. Draucker, D. Baron, R. Yohe, Archaeological applications of laser-induced breakdown spectroscopy: an example from the Coso Volcanic Field, California, using advanced statistical signal processing analysis, Appl. Opt. 49 (13) (2010) C120.
[185] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemom. 17 (3) (2003) 166–173.
[186] J.L. Gottfried, F.C. De Lucia Jr., A.W. Miziolek, Discrimination of explosive residues on organic and inorganic substrates using laser-induced breakdown spectroscopy, J. Anal. At. Spectrom. 24 (3) (2009) 288.
[187] I.B. Gornushkin, B.W. Smith, H. Nasajpour, J.D. Winefordner, Identification of solid materials by correlation analysis using a microscopic laser-induced plasma spectrometer, Anal. Chem. 71 (22) (1999) 5157–5164.
[188] R. Sattmann, I. Mönch, H. Krause, R. Noll, S. Couris, A. Hatziapostolou, R. Miguel, Laser-induced breakdown spectroscopy for polymer identification, Appl. Spectrosc. 52 (3) (1998) 456–461.
[189] V. Motto-Ros, A.S. Koujelev, G.R. Osinski, A.E. Dudelzak, Quantitative multi-elemental laser-induced breakdown spectroscopy using artificial neural networks, J. Eur. Opt. Soc. 3 (2008).
[190] M. Boueri, V. Motto-Ros, W. Lei, L. Qain-LiMa, Z.H. Zheng, J. Yu, Identification of polymer materials using laser-induced breakdown spectroscopy combined with artificial neural networks, Appl. Spectrosc. 65 (2011) 307–314.
[191] S.J. Rehse, J. Diedrich, S. Palchaudhuri, Identification and discrimination of Pseudomonas aeruginosa bacteria grown in blood and bile by laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 62 (10) (2007) 1169–1176.
[192] S.J. Rehse, Q.I. Mohaidat, S. Palchaudhuri, Towards the clinical application of laser-induced breakdown spectroscopy for rapid pathogen diagnosis: the effect of mixed cultures and sample dilution on bacterial identification, Appl. Opt. 49 (13) (2010) C27.
[193] E. Grifoni, S. Legnaioli, G. Lorenzetti, S. Pagnotta, V. Palleschi, Application of graph theory to unsupervised classification of materials by laser-induced breakdown spectroscopy, Spectrochim. Acta B At. Spectrosc. 118 (1) (2016) 40–44.