BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF MECHANICAL ENGINEERING
INSTITUTE OF PHYSICAL ENGINEERING

FAKULTA STROJNÍHO INŽENÝRSTVÍ
ÚSTAV FYZIKÁLNÍHO INŽENÝRSTVÍ

USING LASER-INDUCED BREAKDOWN SPECTROSCOPY (LIBS) FOR MATERIAL ANALYSIS

VYUŽITÍ SPEKTROMETRIE LASEREM BUZENÉHO PLAZMATU (LIBS)
PRO MATERIÁLOVOU ANALÝZU

DISSERTATION THESIS
DISERTAČNÍ PRÁCE

AUTHOR: Ing. PAVEL POŘÍZKA
AUTOR PRÁCE

SUPERVISOR: prof. Ing. JOZEF KAISER, Ph.D.
VEDOUCÍ PRÁCE

BRNO 2014

ABSTRACT

This doctoral thesis is focused on further development of the Laser-Induced Breakdown Spectroscopy (LIBS) device for *in-situ* and in real-time classification and quantification of samples. The major part of work, namely the whole experimental part for this thesis, was conducted at the Federal Institute for Material Research and Testing (BAM) in Berlin, Germany where a simple LIBS system was constructed. In parallel to experimental work, the literature was surveyed with the aim to give a thorough view on the usage of chemometrics in the LIBS community. The application of chemometric algorithms on LIBS data is generally recommended when more complex data sets are obtained.

The research was primarily aimed on the LIBS capability of quantitative analysis and classifying the igneous rocks. Variety of samples was measured employing a simple LIBS system. The sample set was compiled from certified reference materials as well as from real samples collected directly at copper mines in Iran. The samples from Iran were classified *in-situ* by an experienced geologist and the copper content was estimated at the University of Clausthal, Germany. Even though the certified reference materials were analysed, the resulting calibration curve was highly nonlinear. For each individual rock type the relevant part of the calibration curve was observed under different trend. This separation of the calibration curve was assigned to the so-called matrix effect, which strongly affects the LIBS measurement. In other words, when different matrices with complex composition are analysed at once, the quantitative analysis employing the univariate calibration curves may not be reliable. Moreover, the normalization of such calibration curves using the intensity of selected matrix element lines did not let to a significant improvement in their linearity. It is generally not possible to pick up one line, which could perform the linearization independently on the complex data matrices. Chemometric algorithms, such as principal component regression (PCR) and partial least squares regression (PLSR), were used for multivariate calibration. PCR and PLSR may compensate for the matrix effect only to a certain extent. Furthermore, samples were successfully classified based on their spectral fingerprint (*i.e.* composition of matrix elements) employing principal component analysis (PCA) and Kohonen's selfs-organizing maps.

On the basis of theory and results, a solution for the reliable classification and quantification of unknown samples is proposed. The whole study should contribute to the processing of the analytical data measured by the *in-situ* stand-off LIBS device which is currently being constructed at Brno University of Technology in Brno, Czech Republic. However, LIBS can fulfil its potential as the versatile and irreplaceable technique for *in-situ* classification and quantitative analysis only when utilized with chemometric algorithms and data libraries. For those purposes, a fragment of the data library has already been established and tested for the application of LIBS to the mining industry.

ABSTRAKT

Tato doktorská práce je zaměřena na vývoj algoritmu ke zpracování dat naměřených zařízením pro spektrometrii laserem indukovaného plazmatu (angl. LIBS). Zařízení LIBS s tímto algoritmem by mělo být následně schopno provést třídění vzorků a kvantitativní analýzu analytu *in-situ* a v reálném čase. Celá experimentální část této práce byla provedena ve Spolkovém institutu pro materiálový výzku a testování (něm. BAM) v Berlíně, SRN, kde byl sestaven elementární LIBS systém. Souběžně s experimentílní prací byl vytvořen přehled literárních zdrojů s cílem podat ucelený pohled na problematiku chemometrických metod používaných k analýze LIBS měření. Použití chemometrických metod pro analýzu dat získaných pomocí LIBS měření je obecně doporučováno především tehdy, jsou-li analyzovány vzorky s komplexní matricí.

Vývoj algoritmu byl zaměřen na kvantitativní analýzu a třídění vyvřelých hornin na základě měření pomocí LIBS aparatury. Sada vzorků naměřených použitím metody LIBS sestávala z certifikovaných referenčních materiálů a vzorků hornin shromážděných přímo na nalezištích mědi v Íránu. Vzorky z Íránu byly následně na místě roztříděny zkušeným geologem a množství mědi v daných vzorcích bylo změřeno na Univerzitě v Clausthalu, SRN. Výsledné kalibrační křivky byly silně nelineární, přestože byly sestaveny i z měření referenčních vzorků. Kalibrační křivku bylo možné rozložit na několik dílčích tak, že závislost intenzity měděné čáry na množství mědi se nacházela v jiném trendu pro jednotlivé druhy hornin. Rozdělení kalibrační křivky je zpravidla přisuzováno tzv. matričnímu jevu, který silně ovlivňuje měření metodou LIBS. Jinými slovy, pokud určujeme množství analytu ve vzorcích s různou matricí, je výsledná kalibrační křivka sestavená pouze z jedné proměnné (intenzity zvolené spektrální čáry analytu) nepřesná. Navíc, normalizace takto vytvořených kalibračních křivek k intenzitě spektrální čáry matrčního prvku nevedla k výraznému zlepšení linearity. Je obecně nemožné vybrat spektrální čáru jednoho matričního prvku pokud jsou analyzovány prvky s komplexním složením matric. Chemometrické metody, jmenovitě regrese hlavních komponent (angl. PCR) a regrese metodou nejmenších čtverců (angl. PLSR), byly použity v multivariační kvantitatvní analýze, tj. za použití více proměnných/spektrálních čar analytu a matričních prvků. Je potřeba brát v potaz, že PCR a PLSR mohou vyvážit matriční jev pouze do určité míry. Dále byly vzorky úspěšně roztříděny pomocí analýzy hlavních komponent (angl. PCA) a Kohonenových map na základě složení matričních prvků (v anglické literatuře se objevuje termín ‚spectral fingerprint‘)

Na základě teorie a experimentálních měření byl navržen algoritmus pro spolehlivé třídění a kvantifikaci neznámých vzorků. Tato studie by měla přispět ke zpracování dat naměřených *in-situ* přístrojem pro dálkovou LIBS analýzu. Tento přístroj je v současnosti vyvíjen v Brně na Vysokém učení technickém. Toto zařízení bude nenahraditelné při kvantifikaci a klasifikaci vzorků pouze tehdy, pokud bude použito zároveň s chemometrickými metodami a knihovnami dat. Pro tyto účely byla již naměřena a testována část knihoven dat v zaměření na aplikaci metody LIBS do těžebního průmyslu.

ABSTRAKT

Diese Doktorarbeit beschreibt Entwicklungen in der Laser-Induced Breakdown Spectroscopy (LIBS) für *in-situ* und real-time Klassifizierungen und Quantifizierungen von Gesteinsproben. Der gesamte experimentelle Teil dieser Arbeit ist an einem selbst gebauten LIBS System an der Bundesanstalt für Materialforschung und –prüfung (BAM) in Berlin, Bundesrepublik Deutschland, entstanden. Parallel zu den Experimenten ist eine Literaturrecherche gemacht worden, um einen Überblick und eine Bewertung der chemometrischen Analyseverfahren in der LIBS-Community zu erhalten. Generell werden chemometrische Algorithmen für komplexe und schwierig-auszuwertende Datensätze verwendet, daher ist deren Anwendung ratsam und ein Grundbaustein für die folgenden Arbeiten.

Mein Forschungsthema befasst sich mit dem Potential der LIBS zur quantitativen Analyse und Klassifizierung von vulkanischen Gesteinsproben; hierzu wird eine Probenvielzahl mit Hilfe eines einfachen LIBS Aufbaus untersucht. Es sind zertifizierte Referenzmaterialien als auch Proben aus Kupferminen im Iran, welche quasi *in-situ* von einem erfahrenen Geologen klassifiziert sowie der Kupferanteil an der Universität Clausthal, Bundesrepublik Deutschland, bestimmt worden sind. Obwohl zertifizierte Referenzmaterialien analysiert worden sind, kann keine einheitlich-lineare Kalibriergerade erhalten werden. Für jeden Gesteinstypen ist daher eine eigene Kalibriergerade aufgenommen worden. Abweichungen, welche zu den Problemen bei der Erstellung einer Kalibriergeraden führen, liegt der so genannte Matrix-Effekt zugrunde. Dieser beeinflusst die erhaltenen Daten so stark, dass keine univariate Kalibriergerade erstellt werden kann, da diese keine verlässlichen Ergebnisse liefern würde. Eine Normalisierung der Kalibriergeraden bezogen auf bestimmte Matrix Signale führt ebenfalls zu keiner Verbesserung der Linearität. Somit ist es generell nicht möglich nur ein Signal zu wählen, welches unabhängig von Matrixeinflüssen eine lineare Kalibriergerade erzeugt. Chemometrische Algorithmen, wie die principal component regression (PCR) und die partial least squares regression (PLSR), werden für die insofern notwendige multivariate Kalibrierung genutzt. PCR und PLSR können den Matrixeffekt jedoch nur bis zu einem gewissen Ausmaß kompensieren. Um bessere Ergebnisse zu erhalten werden Proben deshalb basierend auf charakteristischen Spektralbereichen mit Hilfe der Hauptkomponentenanalyse (engl. PCA) und Kohonenkarte klassifiziert.

Ausgehend von Theorie und experimentell erhaltenen Ergebnissen ist eine Herangehensweise zur verlässlichen Klassifizierung und Quantifizierung unbekannter Gesteinsproben entwickelt worden. Die gesamte Arbeit soll zur Handhabung von den in *in-situ* stand-off LIBS Experimenten erhaltenen Daten beitragen, welche derzeit an der Technischen Universität Brünn, Tschechische Republik, durchgeführt werden. Jedoch kann LIBS nur den Anforderungen einer robusten, unersetzlichen Technik für *in-situ* Klassifizierungen und quantitativen Analyse entsprechen, wenn eine anschließende Aus- und Bewertung mittels chemometrischer Analyse und dem Abgleich mit Datenbanken durchgeführt wird. Zu diesem Zweck ist bereits eine LIBS-Spektren Datenbank im Rahmen meiner Forschung zur Anwendungen in der Bergbauindustrie eingeführt worden.

KEY WORDS

Laser-induced breakdown spectroscopy, LIBS, plasma, *in-situ* analysis, multivariate data analysis, data libraries, chemometrics, principal component analysis, PCA, partial least squares regression, PLSR, artificial neural networks, ANN.

KLÍČOVÁ SLOVA

Spektrometrie laserem buzeného plazmatu, LIBS, plazma, *in-situ* analýza, multivariační analýza, knihovny dat, chemometrie, analýza hlavních komponent, PCA, regrese metodou nejmenších čtverců, PLSR, umělé neuronové sítě, ANN

SCHLÜSSELWÖRTER

Laserinduzierte Plasmaspektroskopie, LIBS, Plasma, *in-situ*-Analyse, Multivariate Datenanalyse, Datenbibliothek, Chemometrie, Hauptkomponentenanalyse, PCA, partial least squares regression, PLSR, Künstliches neuronales Netz, ANN.

POŘÍZKA, P. *Využití spektrometrie laserem buzeného plazmatu (LIBS) pro materiálovou analýzu.* Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2014. 105 s. Vedoucí disertační práce prof. Ing. Jozef Kaiser, Ph.D..

Notification

I hereby declare that I wrote this doctoral thesis on my own with a supervision of prof. Ing. Jozef Kaiser, Ph.D., Dr. Jens Riedel and Dr. Igor B. Gornushkin. All the literature used is listed in references.

Pavel Pořízka

Acknowledgement

I am grateful to prof. Ing. Jozef Kaiser, Ph.D., Dr. Jens Riedel and Dr. Igor B. Gornushkin for providing me with useful pieces of advice and giving me the inspiration during their supervising work.

I would also like to thank the members of the LIBS research and development team at Brno University of Technology as well as the members of the group 1.0 at BAM, Federal Institute for Material Research and Testing, Berlin, Germany.

CONTENTS

INTRODUCTION

> *"The success of LIBS as a field-portable point detector of chemical and biological warfare agents requires the development and optimization of statistical methods for rapidly analysing complex spectra obtained in the field,"*
> Munson et al. from the U.S. Army Research Laboratory, USA [1].

Recently, the needs for direct *in-situ* and in real-time elemental analysis have been rapidly increasing in every field of interest (*e.g.* mining, archaeology, industry, biology, etc.). Laser-Induced Breakdown Spectroscopy (LIBS; in the literature the term Laser-Induced Plasma Spectroscopy (LIPS) can be also found) can meet the challenge as an analytical technique for fast, real-time, *in-situ*, and preparation-less measurement. LIBS devices may be capable to stand the harsh environments and difficult to reach sites due to their robust and relatively simple instrumentation [2, 3, 4]. LIBS has already found its place among the other well established techniques for elemental analysis (namely Laser Ablation Inductively Coupled Plasma (LA-ICP) coupled with Mass-Spectrometry (MS) or Optical Emission Spectrometry (OES) techniques).

LIBS is becoming a popular method for elemental analysis due to its advantages (simplicity, cost, size, robustness, fast and non-demanding measurements), however, despite that LIBS still has some disadvantages (detection limits, repeatability) compared to other techniques (namely LA-ICP techniques). In spite of these negatives, LIBS might compete and overcome the other techniques in near future due to the improvements in instrumentation, application of data mining algorithms, and understanding of plasma evolution and its diagnostics. The usage of LIBS in various fields is summarized in many review articles [5, 6, 7, 8, 9]. In 2004, LIBS was considered to be the future super star among other spectroscopic techniques [10].

The biggest achievement in the field of LIBS is a state-of-the-art LIBS device developed by NASA, the so-called ChemCam LIBS device, which is attached to the Curiosity rover, the Mars Science Laboratory. ChemCam LIBS device is remotely controlled from the Earth and provides the analysis of the planet Mars surface, resp. Gale crater. Mars rover is a unique device and its construction was very challenging. The approaches used to build the device and mainly to create the data libraries could be an inspiration how to overcome the problems arising in this kind of research and development [11].

LIBS can probe samples in any state of matter (solid, liquid, and gaseous) and obtained spectra give complex information about the elemental composition of the sample under study, the so-called chemical fingerprint. The sampling rate is limited only by the frequency of the laser or the readout of the detector. This can lead to large data sets that may contain so-called latent variables, *i.e.* hidden relationship between samples. A deeper insight into the laser-induced breakdown, plasma formation, matrix effect, relevant analysis of the spectrum, etc. is given in section 1. However, bulky data sets are difficult to work with. When standard univariate approaches, in which the intensity is simply assumed to scale linearly with the concentration of the element of interest within the solid sample material, are utilized then the data may give misleading results. Therefore, more advanced algorithms ought to be used. There is a need of algorithms and data mining processes to handle large amount of data due to the ability of LIBS of fast and preparation-less measurements. Chemometric algorithms are standardly used in the LIBS community for data set analysis and processing. Chemometrics can be seen as useful and dilative tool for LIBS analysis that have already proved their usefulness in many various applications of LIBS, see section 2.4. Individual samples can be classified into groups, clustered together or be discriminated based on the LIBS spectra. Chemometric approaches are used for the discrimination of unknown samples in recently the most enormous successful application of LIBS device, the Mars Rover [11].

The team of the Laboratory of Laser Spectroscopy at Brno University of Technology is currently developing a stand-off LIBS device, figure 1. The typical LIBS spectrum and ablation crater are shown in figure 2. Creation of the necessary libraries for the remote LIBS device and the testing of the chemometric algorithms for sample classification are discussed further in this thesis. The data libraries are valuable and indispensable for the proposed analysis. The creation of the libraries is essential due to their uniqueness and their non-existence on the market.

Figure 1, schematic diagram of the stand-off LIBS device.
This picture was created by Ing. Jan Novotný, Ph.D and Ing. Michal Brada, both from Brno University of Technology.

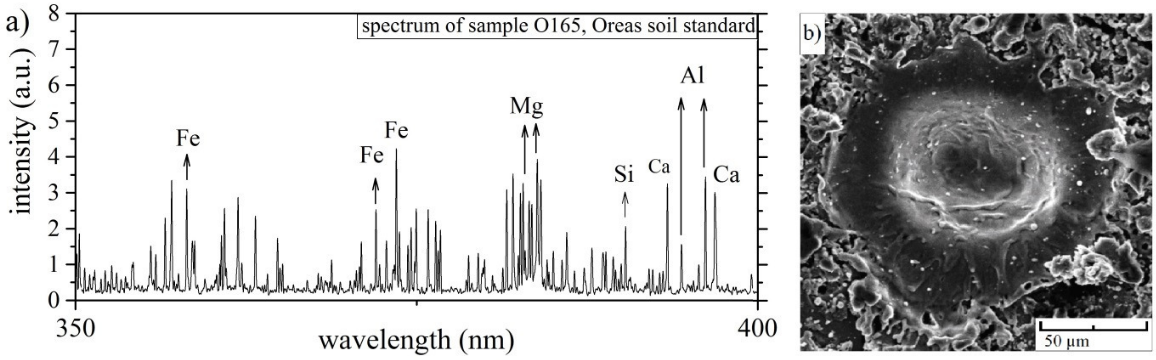


Figure 2, a) typical LIBS spectrum of Oreas soil standard, O165, b) typical crater created during the ablation of material.

1. LASER-INDUCED BREAKDOWN SPECTROSCOPY

> *"There wouldn´t be any LIBS device on Mars Rover developed by NASA*
> *if LIBS was not a very robust and feasible technique,"*
> Richard E. Russo[1], EMSLIBS 2013 conference, Bari, Italy

The LIBS measurement is based on the laser-induced plasma (LIP) formation and (in most cases) time-resolved detection of its radiation. Despite the relative simplicity of the LIBS system the phenomenon of plasma formation is very complex. In following chapters, a basic LIBS instrument is described as well as the necessary theoretical background on laser-induced plasma. More detailed information on the fundamentals of LIP properties and consecutive spectral analysis (LIBS in general) can be found in the books [2, 3, 4] or in the recent comprehensive review articles [12, 13, 14].

1.1. HISTORY OF LIBS

The fifty years long history of LIBS is essentially connected with the invention and development of lasers [15]. The milestones in the LIBS research are briefly reviewed in this chapter.

In 1962, two studies were published about the work with non-Q-switched lasers. Dacey [16] informed about the formation of non-vigorous luminous plasmas on the iron and carbon targets. Brech and Cross [17] vaporized metallic and non-metallic materials using a ruby laser, then the vapours were analysed with an electrical spark. This study is considered to be the beginning of the LIBS itself [10]; though the laser-induced plasma is not produced directly. In 1963, the Q-switched ruby laser was introduced for the matter vaporization, though the plasma was still formed employing an electrical spark. Figures of merit of the system were evaluated on geological samples [18] and biological tissues [19]. Interest in the LIBS started to grow when the lasers with improved Q-switches were brought to the market [20]. The plasma was then induced directly by the laser beam without the assistant electrode excitation. In the meantime, Runge *et al.* [21] published the first calibration plots, in the scientific community of LIBS, for Cr and Ni in the steel standards. Debras-Guédon and Liodec [22, 23] made series of measurements with optimized LIBS setup. In their study, 25 elements and CN and AlO molecular bands were observed, however, CN bands were not present in all plasmas but of carbonate samples. That is the first note on the detection of molecular bands employing LIBS. Authors as well suggested the formation of CN in ambient atmosphere only, where carbon atoms in the plasma plume interacted with the nitrogen coming from the ambient air. Moreover, in this extended study, Debras-Guédon and Liodec could not detect several elements due to the lower sensitivity in the UV region and as well due to the intense continuum emission.

---

[1] Richard E. Russo is one of the leading scientists at Lawrence Berkeley National Laboratory, USA. He is a long-term expert in the field of Laser-Induced Breakdown Spectroscopy with more than 200 scientific publications on his account.

More pioneering works were done on the plasma diagnostics [24], laser-matter interaction [25] and temporally gated detection [26]. Nevertheless, the first wave of extent interest in the LIBS technique faded in the late 1960s away despite the first commercial LIBS systems by Carl Zeiss and LMA [10, 27]. In 1981, Loree and Radziemski [28] came up with the acronym LIBS, which refers to the breakdown of ambient atmosphere during the formation of a laser-induced plasma. In the mid of 1980s, the invention of intensified charge-coupled device (ICCD) renewed the interest in LIBS. These detectors enabled a temporal detection of the plasma radiation. Moreover, the evasion of the continuum emission of the LIP leads to the improvement in the signal-to-noise ratio and subsequently to the improvement of the limits of detection (LOD).

LIBS is becoming popular technique among other atomic emission spectroscopy (AES) techniques [10]. LIBS excels due to its simple, versatile and relatively low-cost instrumentation on one hand and analytical flexibility on the other. Nevertheless, the intense study of LIBS took place practically only in the last two decades, when the more advanced instruments (*e.g.* more compact and robust lasers, spectrometers with improved spectral resolution and more sensitive detectors) were introduced. There is continuous increase in the number of articles published in the field of LIBS in the last 20 years, figure 3, [4, 29]. Since 1990s, the application of LIBS scattered among various applications (*e.g.* on-line analysis, heterogeneity and quality control in steel industry, mineralogy, mining, space objects investigation, archaeology, cultural heritage, medical applications, biological samples, environmental analysis, bioremediation).

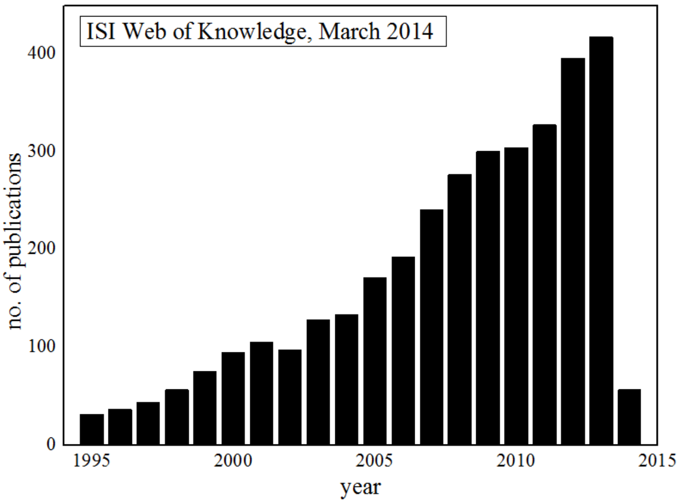


Figure 3, number of LIBS publication in the last 20 years **[30]**, ‘laser-induced breakdown spectroscopy’ used as a keyword.

Moreover, the development and optimization of LIBS systems for various applications and the theoretical understanding of the spatial and temporal evolution of a plasma plume are of great interest in many scientific groups. In past ten years, many review articles reflected the vast diversity of the LIBS applications. In two follow-up comprehensive studies, Hahn and Omenetto [12, 13] exhaustively reviewed fundamentals and diagnostics of LIP and applications in which the LIBS community scattered to. Other reviews covering the history of LIBS, plasma evolution, fundamental instrumentation and advanced applications can be found [5, 31, 32, 14, 33]. The review on modelling of the LIP was brought by Gornushkin and Panne [34]. In this review, calibration-free (CF) models were considered as a possible candidate for standardless spectroscopic analysis. It is noteworthy, CF-LIBS is based on the assumption that an optically thin LIP has to be in the state of local thermodynamic equilibrium (LTE) [2]. Then the spectral line of an analyte is parameterised in order to compensate the matrix effect. As a result, there is no need for the calibration of the LIBS system, *i.e.* constructing the calibrations curves for the element of interest from the measurement of reference samples. Another review on modelling of LIP was brought by Dijk *et al.* [35], 20 contributors were invited to give an in-depth discussion.

The quantitative analysis employing LIBS has been limited due to the inadequate level of the analytical figures of merit (accuracy, precision, detection limits) [2]. LIBS systems are vastly used in industry for their fast, non-demanding and robust analysis [4, 36, 37]. Tognoni *et al.* [38] focused their article on the LIBS for quantitative micro-analysis (to map the surface and check for the local inhomogeneity) in industry and discussed the instrumental part and the optimization in detail. A summary on the analysis of minerals, rocks, and soils is given in [7, 39, 40, 41], on explosives [42], space and cultural heritage [41]. LIBS proved to be competitive in challenging analysis of molecularly complex biomedical materials or clinical specimens [6] and biological samples in general [9, 43]. LIBS is capable of fast analysis resulting in bulky data sets. Each single sample is represented by a unique set of variables forming complex matrix. Univariate analysis (classification or quantification) may be therefore misleading and result in false positives, *i.e.* an unknown sample is wrongly classified or the content of the analyte in an unknown sample is incorrectly estimated. Data mining is provided utilizing advanced statistical algorithms of chemometrics [44]. The applications of LIBS using chemometric algorithms are detailed in [7, 39, 42].

Hahn and Omenetto [13] discussed possible ways how to improve the sensitivity and detection power (reciprocal to the limits of detection) of the LIBS technique. The task is to increase the signal-to-noise ratio which results in better detection limits. More laser pulses can be used to ablate the material and produce the plasma or to pump the energy into the existing plasma plume. Double (or multiple) pulse experiments lead to the increase of the LIBS signal. The investigation of double pulse LIBS is mostly focused on collinear (where both laser pulses are introduced on the sample surface from the same direction) or orthogonal (laser beams are parallel to each other) arrangements. Unfortunately, those arrangements bring more variables into the optimization of the whole system (interpulse delay, laser wavelengths, and irradiances of the laser pulses). Naturally, the properties of DP LIP are completely different to those of single pulse (SP) LIP. The selection of the beam path geometry is crucial as well. The orthogonal beam geometry can be beneficially used in mapping applications and the collinear geometry offers larger mass removal. Review articles describing above mentioned difficulties of DP LIBS are given in [45, 46].

It was already mentioned that LIBS, as a technique for direct elemental analysis, can be employed as well for the detection of molecular bands [13]. In later stages of a plasma evolution molecules are formed from the ablated atoms only, or by chemical reactions of the LIP with ambient air. The detection of CN, OH, C_2 can give the direct information on the chemical compounds in the analysed sample. Moreover, LIBS can be employed in connection with other techniques, such as Raman spectroscopy, to extend its analytical performance [13]. Raman spectroscopy may be utilized to detect molecular complexity of the sample under study. Both techniques can be coupled together also in the stand-off analysis of remote objects.

The potential of the LIBS system is also in the remote and *in-situ* detection of difficult-to-reach or hazardous objects. Fortes and Laserna [29] brought an extensive study on the field-portable LIBS instruments, distinguishing among portable, remote and stand-off detection. Sallé *et al.* [47] studied only the stand-off LIBS for open-path applications and explained the fundamentals of a stand-off system. Harmon *et al.* [40] investigated the performance of portable LIBS system in the analysis of minerals and geomaterials. Nevertheless, the need for more compact and reliable instruments resulted in a relatively late introduction of portable LIBS systems. The first portable devices were developed in the late 1990s by the groups of Cremers [48] and Winefordner [49]. In remote LIBS the laser pulse is led to the sample and focused on its surface by an optical fibre while the plasma radiation is collected by the same or another optical fibre. In 1995, remote LIBS system utilizing fibre optics was introduced for the analysis of ferrous materials [50] and soils [51] for the distances up to 100m. Stand-off LIBS, capable of an open path detection up to 2.4 m, was firstly demonstrated in 1987 by Cremers [52]. For more detailed information on the development of the field-portable LIBS please be referred to the review articles [29, 47]. To conclude, field-portable LIBS systems increase in general the flexibility and applicability of the LIBS in respect to the other AES techniques.

The prototype of Martian rover was designed and tested more than ten years prior the landing on Mars. The ChemCam instrument package, the LIBS system, on the Mars Science Laboratory rover, Curiosity, is one of the biggest achievements of the LIBS technique so far. The LIBS system is utilized to obtain the elemental composition of the surface of the planet Mars. Data library was established in advance to calibrate the system and to offer the possibility to classify unknown samples on-site based on their chemical fingerprint [11]. Moreover, LIBS is employed as well for the identification of organic materials via detection of CN bands and water via OH emission bands [53]. In European project, combined Raman/LIBS spectrometer will be used in the next Mars mission [54].

LIBS becomes one of the effective analytical techniques in the industry for real-time and *in-situ* analysis. Moreover, the detection limits and repeatability of the LIBS system is competitive to the other AES techniques [10]. The recent trend in the field of LIBS is to press down the costs of instrumentation and to make the system more compact and robust, figure 4.

Figure 4, advanced LIBS devices (from left to right): stand-off LIBS system during the measurement of the facade of the historical building in Malaga, Spain (the group of Javier Lasserna); handheld LIBS (Lasersec Systems) for direct analysis; LIBS device attached directly to the mining machine (the group of Reinhard Noll); ChamCam LIBS system attached to the Mars Rover, Curiosity.

1.2. SETUP FOR LIBS

Nowadays, LIBS is mainly studied on the academic ground. The stock with commercial LIBS instruments covers several industrial applications, however, their performance is limited [4]. Therefore, almost every LIBS device is custom built and optimized for case studies even though basic principles of instruments stay unchanged. The scheme of a basic LIBS setup with interaction (or vacuum) chamber is depicted on figure 5. For detailed information on the properties of individual instruments utilized in LIBS experiment refer to [2, 4].

Usually, the laser pulse impacts the sample surface perpendicularly from above. The most common LIBS device employs a Q-switched high energy solid-state Nd:YAG laser, lasing at its fundamental wavelength (1064 nm) or equipped with non-linear crystals to produce other harmonics (532, 366 and 266 nm). The lasers are usually operated at the frequency of 10 or 20 Hz with the pulse width of units of ns (typically 6 – 10 ns). Laser pulses of energies of hundreds mJ are delivered with Gaussian or flat-top beam profile. Lasers (continuously pumped solid state lasers with acousto-optical Q-switching and diode-pumped solid state lasers with electro-optic Q-switching) with higher frequency rates (up to tens of kHz) can be used depending on the application. However, the latter lasers produce pulses of lower energies (typically below 1 mJ/pulse). In general, any change in the repetition of the laser changes the parameters of produced laser beam (pulse energy, pulse width, and beam profile). In general, the irradiance decreases with the increasing repetition rate.

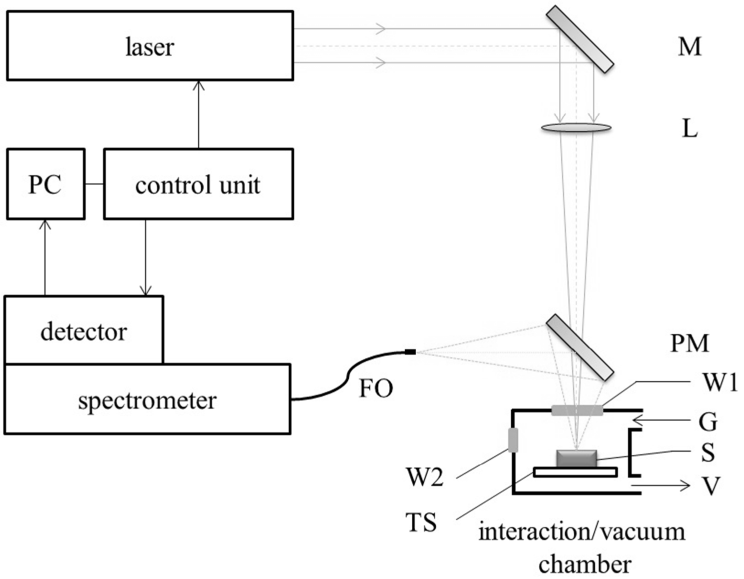


Figure 5, the scheme of a basic LIBS setup; M – mirror, L – focusing lens, PM – pierced mirror for the collection of plasma radiation from above, W1 – transparent window, W2 – transparent windows for side view/collection, G – gas inlet, S – sample, V – vacuum pump, TS – translation stage, FO – fibre optics

The pulse width is crucial parameter in the LIBS experiment [4]. For nanosecond laser pulses, a LIP is generated and spatially evolves while the laser radiation still persists and consequently, a LIP is affected by the impacting laser pulse. For femtosecond pulses, laser beam transmits the energy to the material before the ablation process starts. However, lasers generating pulses of shorter widths (ps or even fs) are rarely used in the LIBS experiment. Furthermore, the utilization of fs-lasers leads to the reduction of the matrix effect [55] which is balanced by higher price of the instrument. Moreover, a fast transition of the energy results in shorter persistence of a LIP with weaker emission because particles forming the LIP plume are not heated up by the laser pulse.

Laser pulse is focused into a tight spot (typically several hundreds of μm) in order to obtain an increase in the irradiance on this interaction spot. Laser beam can be considered as a parabolic approximation of a spherical wave [55] because its properties change very slowly along the path of the beam. In the case of Gaussian beam profile, where the transverse energy is distributed according to the Gaussian function, the smallest spot is called a waist w_0. The size of the waist can be varied (focused or diverged) with a lens as follows:

$$w_0 = \frac{2 f_L \lambda}{\pi D_L}, \tag{1.1}$$

where λ is the wavelength of the laser radiation, f_L is the focal length of the lens, and D_L is the diameter of the laser spot on the lens. It is obvious from this equation that focused spot is linearly dependent on the wavelength of the laser beam. Furthermore, the absorption of a laser pulse depends on the laser wavelength as well as on the properties of the surface (roughness) in the region of interaction [4]. Smaller craters can be therefore produced utilizing lower pulse energies, sub-nanosecond laser pulses, the use of optimized focusing optics, lower laser wavelengths or any combination of aforementioned. The size of the crater is important in the LIBS analysis, *e.g.* in the mapping of the sample surface where the lateral resolution increases with a lower laser spot. However, the reduction of the ablation spot increases the probability of the prespark occurrence. The irradiance of the laser pulse focused into a tight spot may reach the threshold of the ambient gas. Therefore the size of the laser spot as well as the parameters of the incident laser should be optimized. Furthermore, during the extensive LIBS measurement particles are released into the ambient atmosphere surrounding the sample. Then incident laser pulse may ablate those particles hovering above the sample and create an unwanted spark/discharge. The creation of such presparks may be avoided when the sample is analysed within an interaction chamber with continuous flow of inert gas (or simply drift of air) above the sample [4]. Nevertheless, LIBS measurements are usually run in ambient atmospheric conditions which are of academic and industrial interest due to the *in-situ* applications. For more special requirements such as measurement of various gases, well-defined atmospheric conditions, detection of elements with emission lines in UV region the interaction or vacuum chamber should be employed. Though, the interaction chamber is not indispensable for LIBS measurement. Moreover, in the case of solid samples, prepulses are recommended to clean the surface from the unwanted contaminations which are not representative for the bulk of the sample. Those prepulses are not used for further spectrochemical analysis.

The radiation of the plasma is usually collected utilizing simple optical system (consisting of one or several lenses, ellipsoidal/spherical mirror or lens/mirror collector) from above or perpendicularly from the side of the plasma plume. Collected radiation is then introduced via optical cables and on the entrance slit of the spectrometer. Spectrometers in the echelle, Czerny-Turner and Paschen-Runge configuration are widely-spread. Each spectrometer offers different positives and negatives compared to the other, such as spectral range, spectral resolution, and optical throughput. The most suitable spectrometer should be chosen based on the application, *e.g.* the spectrum of a steel sample is very dense which requires a good spectral resolution for reliable analysis.

The spectrometer spectrally disperses the collected LIP radiation to obtain spectrally resolved lines. The main parameters of a spectrometer are the focal length f_S and the resolution R_S:

$$R_S = \frac{\lambda}{\Delta\lambda}, \tag{1.2}$$

where $\Delta\lambda$ is the least distance between two resolvable wavelengths in the neighborhood of the wavelength λ, and the angular dispersion D_S:

$$D_S = \frac{d\delta}{d\lambda}, \tag{1.3}$$

where $d\delta$ is the angle separating two light beams with the wavelength difference of $d\lambda$.

Dispersed LIP radiation is then detected by a detector, such as photomultiplier tube (PMT), CCD, ICCD (*i.e.* CCD combined with microchannel plate (MCP)), EMCCD, etc. The detectors differ in their properties and parameters and should be selected according to the application and selected spectrometer. Among the main parameters belong number of sensitive elements, quantum efficiency, spectral range, and possibility of temporal gating. The detectors and spectrometers are commonly used in following configurations [4]: (a) Czerny–Turner spectrometer equipped with an intensified PMT or CCD-line detector, (b) echelle spectrometer and mostly the ICCD detector, and (c) Paschen–Runge spectrometer and PMT or CCD line detectors.

It should be noted that, despite the simplicity and obvious advantages of LIBS the analysis is strongly dependent on the homogeneity of the sample when smaller laser spots are utilized without averaging. No need for sample preparation results in probable partial contamination of the measured surface, when solids are considered. Though, any side effects can be avoided with proper design of the experiment. The reproducibility, repeatability, and reliability of the LIBS measurement are the key parameters of further classification and quantitative analysis.

The laser and its properties such as collection optics and utilized spectrometer with detector play their role in further spectral analysis. This makes every LIBS device unique and results have to be compared with attention to those variables. Plasma dynamics, plasma properties and analysis of LIP is described in detail in following chapters.

1.3. LASER-INDUCED PLASMA

A typical LIP is formed utilizing short high energy laser beam where the target is exposed to the irradiance of up to 10^9 W.cm^{-2}. This irradiance is common in LIBS experiments (utilizing ns lasers) and it has been found that consequent laser ablation is stoichiometric [2, 33]. However, Miziolek *et al.* [2] further suggest checking the occurrence of stoichiometric ablation *a posteriori* because a LIP formation differs with employed LIBS system and samples under study.

In general, the plasma produced by a short Q-switched laser pulse is different from the LIP produced by a continuous source. The main difference is in the short persistence (usually units of µs) and rapid evolution of the LIP induced by Q-switched lasers. The spatial and temporal evolution of the LIP formation is depicted in figure 6. The luminous micro-plasma is produced when the high-energy laser pulse is focused on the sample surface (1). The nanosecond laser pulse[2] transmits high power on the small area of the sample (in the range of GW/cm^2) during the impact, *i.e.* laser/matter interaction. Sample begins to heat, melt and evaporate even during the laser pulse duration (2). Small amount of the sample with the ambient gas is ablated and a LIP is formed accompanied with characteristic sound of the ultrasonic shockwave (3). The shockwave is produced by an ultrasonic expansion of the plasma plume. Consequently, expansion causes changes in the pressure of the ambient gas surrounding the expanding plasma plume. A pressure shockwave and a plasma plume spread into the space in all directions resulting in a negative pressure gradient in the centre of the plume, the expansion is inverted and the whole process collapses back to the initial state.

Plasma starts to spread outward into the space (4) (against the direction of the laser pulse when the surface is considered) with ultrasonic velocity (~10^5 cm/s) even during the laser beam impact on the surface of the sample (4). Based on this expansion the plume appears pear- or cigar-shaped. Due to inverse Bremsstrahlung, the plasma is opaque concerning the NIR laser radiation. Thus the laser is absorbed by the plasma which shields the sample surface from the laser. However, the cascade absorption of multiples subsequent photons by the plasma leads to a local heating to several tens of kK. Plasma plume shields the surface of the sample against the laser radiation and therefore no more mass of the sample can be ablated and the laser pulse transmits the rest of its energy only into the plasma, this phenomenon is called plasma shielding.

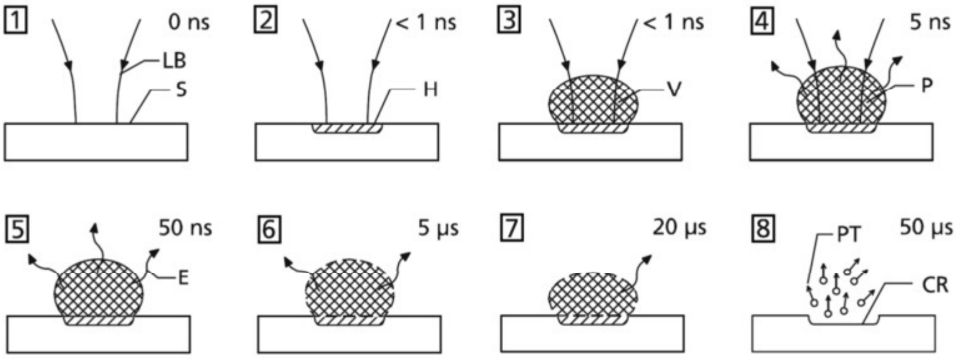


Figure 6, scheme of the temporal and spatial LIP evolution **[4]**.

---

[2] Q-switched lasers producing nanosecond pulses are the most commonly used by the LIBS scientific community due to the market availability and the micro-plasma formed employing those laser sources is well investigated [31].

In early stages of its temporal evolution the LIP is highly ionized and detected radiation of the plasma suffers from the Bremsstrahlung and electron-ion recombination. This radiation carries no meaningful spectral information and is usually gated-out during the detection. As plasma plume (highly ionized) expands into the space then loses its energy and starts to atomize. Particles forming the plasma are excited to higher quantum states because of the high temperature and particle interactions. While relaxing the ions and atoms emit the radiation (in the sense of characteristic wavelengths) which is detected and can serve to the purpose of spectroscopic analysis (4-8). The plasma temperature dynamically changes in the course of persistence of a LIP, when in the beginning of the plasma formation the temperature reaches tens of kK. The energy of the plasma is dissipated due to the emission of the characteristic radiation, recombination of the species within the plasma and the expansion of the plasma plume. Consequently, the plasma temperature rapidly decreases [57].

There are no restrictions for the state of matter (solid, liquid or gaseous), from which LIP could be formed. Therefore, laser pulse can be focused onto the surface of solids/liquids or into the bulk of liquids/gases to form a LIP. The limit irradiance of the laser pulse, which is needed to generate a LIP, is defined as the breakdown threshold. This limit differs with a substance to be analysed as well as with parameters of employed laser pulse [2, 56]. The creation of the plasma and its properties has convoluted dependency on the following variables: laser irradiance, laser wavelength, duration of the laser pulse, laser-matter interaction (different threshold energy for breakdown and a plasma formation is needed depending on the state of the matter), ambient gas (air, He, Ar …) and its pressure, vacuum conditions, etc.

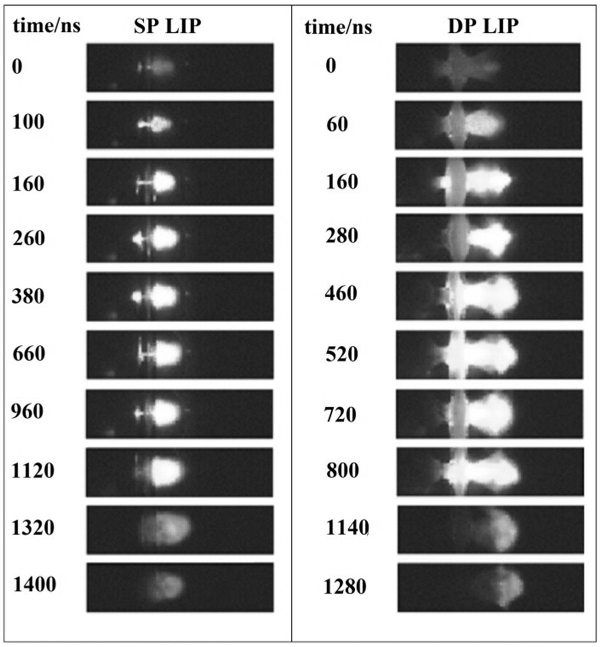


Figure 7, spatial and temporal evolution of single pulse (SP) and double pulse (DP) LIP formed on a vertical steady thin flow of liquid, measured with high speed camera Phantom v12.1 at Brno University of Technology. Experimental settings: energy of the first laser pulse was 40 mJ and 120 mJ of the second one, interpulse delay was set to 1 µs. The exposition of the high speed camera was 1 µs.

The spatial and temporal evolution of the single pulse and double pulse LIP formed on a thin laminar flow of liquid is depicted in the figure 7, where laser beams were introduced from the right hand side and liquid jet was flowing from the top to the bottom. A LIP was in both cases propagating in the direction against the incoming laser light. It is obvious, that in a DP LIBS experiment the plasma is more intense and the persistence of a DP LIP is longer compared to an SP LIP.

The emission of a LIP evolves rapidly in time, as is depicted in figure 8. Each spectrum in this diagram was obtained with a constant exposition time (gate width G_w) of 5 μs. The measurement was temporally resolved, $i.e.$ a time gap (gate delay G_d) between the ignition of a LIP and the beginning of the exposition was varied with a step of 100 ns. In early stages of the LIP emission, strong continuum emission is observed. This emission is caused by free-free transitions, Bremsstrahlung (a deceleration radiation caused by the interactions among free electrons and larger particles) and recombination radiation (coming from the recombination of ions and free electrons in LIP plume). The continuum emission carries no valuable information and is disturbing in further spectrochemical analysis. Therefore, it is beneficial to temporally gate the detection of plasma radiation in LIBS applications.

With increasing gate delay the electrons and ions recombine while a LIP cools down. As a consequence, the continuum radiation is less intense. The bold line in figure 8 denotes the gate delay for which the best signal-to-noise ratio (SNR) is observed. The gate delay and gate width have to be optimized for each LIBS experiment while the properties of a LIP differ, as was already mentioned. It is apparent that with an increasing gate delay intensities of atomic and ionic lines are gradually dropping.

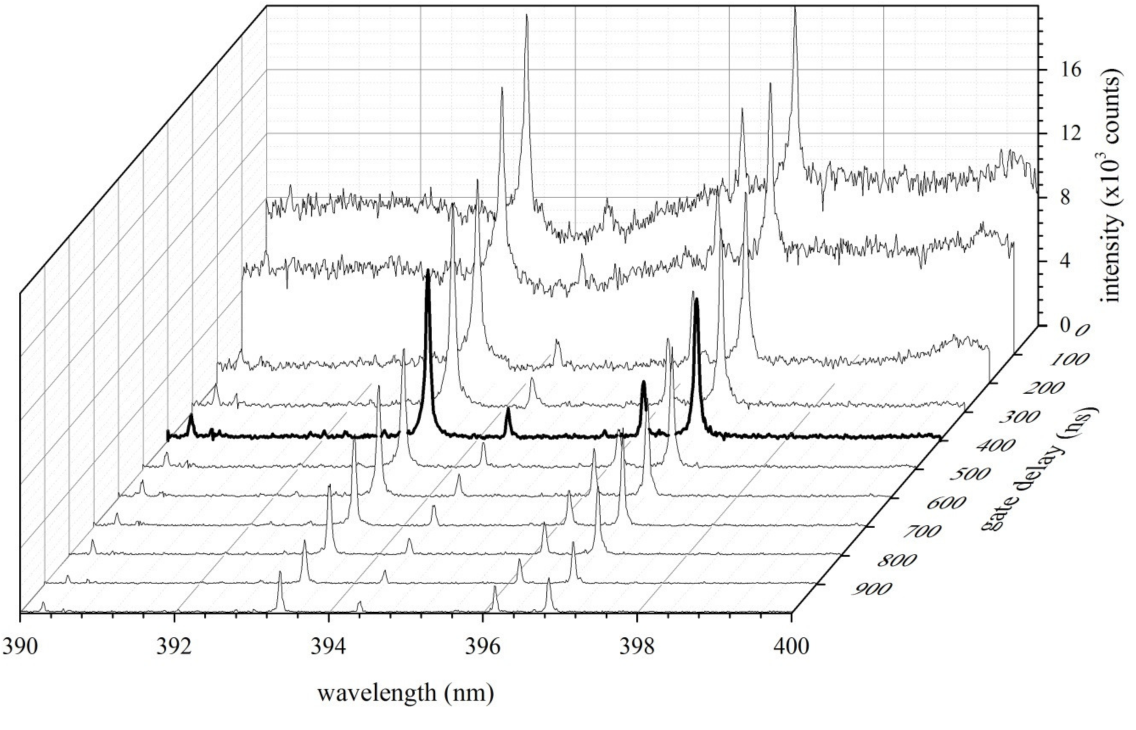


Figure 8, temporally resolved LIP measured of the mixture of soil standard Oreas 504 with distilled water (in a ratio of 2:1). Bold line marks the gate delay used in LIBS experiments optimized to the best signal-to-noise ratio.

While observing spectral lines and analysing their intensities qualitative and quantitative information is obtained. Nevertheless, quantitative analysis is not a trivial task. The intensity of the line depends on the amount of the element within the plasma as well as on the properties of a LIP. Considering a solid sample, the dependence of the detected LIP signal S_{el} (counts) of one selected line was derived in [2] as follows:

$$S_{el} = A_{ij} f_{int} f_{exc} f_{det}, \tag{1.4}$$

where A_{ij} is spontaneous transition probability of the quantum transition representing the chosen spectral line (so called Einstein coefficient). The function f_{int} is related to the ablation/vaporization of solid material, f_{exc} refers to the excitation/ionization mechanism leading to atomic (ionic) emission, f_{det} is the function characterizing the radiation environment. It is noteworthy that the three functions describing the initial laser/sample interaction are interrelated.

The shape of a spectral line can be approximated with a Voigt profile function, which is a convolution of Gaussian and Lorentzian profile functions. The Voigt profile of a spectral line (equations 1.5) is the consequence of two main broadening mechanisms (Doppler Effect and Stark effect) in the LIP plume. The estimation of the Voigt profile, *i.e.* the convolution of the Gaussian and Lorentzian profiles, is severe therefore a simplified pseudo-Voigt profile of the spectral line is usually considered. The Doppler broadening is caused by the expansion of the plasma plume and disorderly movement of the light emitting particles with respect to the detection direction. The Stark effect is a consequence of local electric fields formed by charged particles in the neighbourhood of emitting particles.

The shape of a spectral line $\Gamma_V(\lambda_e)$ is described by the Voigt profile as follows [4]:

$$\Gamma_V(\lambda_e) = \frac{2\sqrt{\ln 2/\pi}}{\Delta\lambda_D} K(u,a), \tag{1.5a}$$

$$K(u,a) = \frac{a}{\pi}\int_{-\infty}^{\infty} \frac{e^{-t^2}}{(u-t)^2+a^2}\,dt, \tag{1.5a}$$

$$u = \frac{2\sqrt{\ln 2}}{\Delta\lambda_D}(\lambda_e - \lambda_0), \tag{1.5a}$$

$$a = \sqrt{\ln 2}\,\frac{\Delta\lambda_N + \Delta\lambda_L}{\Delta\lambda_D}, \tag{1.5a}$$

where $K(u,a)$ is the Voigt profile function, u and a are variables, λ_e is the wavelength of the emitted characteristic radiation and λ_0 is the central wavelength of the spectral line. $\Delta\lambda_N$ is the natural line width, $\Delta\lambda_L$ is the Lorentz line width, and $\Delta\lambda_D$ is the Doppler line width.

The generation of a LIP becomes very complex and its properties (discussed in chapter 1.4) strongly depend on the employed instrumentation and the matrix of the sample (chapter 1.6). Moreover, there exist theoretical models capable to simulate the postbreakdown dynamic evolution of plasma and its properties. The theoretically imitated spectrum is in a good agreement with the experimentally measured one [57]. This model can help to understand the dynamics of plasma evolution and lead to the improvement in the spectral analysis.

1.4. LASER-INDUCED PLASMA PROPERTIES

Small amount of the sample is ablated after the interaction of short high energy laser beam with the surface of the sample (in the case of solid and liquid phase) and luminous plasma plume is formed. The properties of a LIP are dependent on the matrix of the sample, laser beam properties (its duration, energy, and wavelength) and ambient gas surrounding the sample. Produced by a high energy pulsed laser beam, LIP undergoes rapid changes in chemical composition, its temperature and density. The persistence of a LIP can be in the range of units of microseconds. Typical LIP is spatially inhomogeneous and its characteristic parameters evolve with time [14]. Temporal and spatial evolution of a LIP can be described *e.g.* by electron number density and electron temperature. The electron number density can be directly evaluated from the broadening of a spectral line (for instance H_{α} or Fe (I) 538.34 nm). The plasma temperature is usually estimated from the Boltzmann plot whose construction requires particles of the plasma to obey Maxwell-Boltzmann distribution. For those purposes, optically thin plasma and local thermodynamic equilibrium (LTE) have to be considered for further evaluation of physical parameters [2, 4, 12, 14, 58].

The concept of optically thin plasma assumes that all the emitted photons escape the plasma plume [58]. When this condition is not fulfilled, then a LIP is optically thick and self-absorption takes place. Self-absorption describes the phenomenon when emitted photons are absorbed within the plasma. In this case, detected spectral lines do not have Voigt profile and further analysis based on those lines could be misleading and the results inaccurate. The non-resonant lines should be used to avoid possible self-absorption, *i.e.* lines, which do not originate from the quantum transition containing quantum ground state. In general, resonant lines should not be utilized in the computation of the plasma temperature and in quantitative analysis should be considered with care (see further in the text). Moreover, the self-absorption occurs mainly in the earlier stages of a LIP evolution when the density of LIP is higher.

The electrons, atoms, ions and molecules forming up a plasma plume are described by various forms of energies (*e.g.* kinetic, excitation and ionization). Furthermore, all processes within the plasma have to be balanced and characterized by a single temperature. Then we can assume that the plasma is found in the state of a complete thermodynamic equilibrium. However, this state does not occur in laser-induced plasmas due to its rapid temporal evolution. As Hahn and Omenetto [12] stated, the process of excitation of atoms by collisions with electrons is equal to the reverse deactivation process (collisions of the second kind), collisional ionization is equal to three-body collisional recombination, and radiation emitted is equal to the radiation absorbed. Complete local thermodynamic equilibrium occurs when collisions are dominant and the laws describing the full thermodynamic equilibrium are valid except for the existence of radiation disequilibrium. Resonant transitions, characterized by higher values of the Einstein coefficient of spontaneous emission, are depopulated much faster in contrast to the other transitions, which results in radiative disequilibrium. When we consider only quantum transitions above certain level for describing the achievement of the equilibrium then plasma is in the state of partial local thermodynamic equilibrium.

In the LIBS literature it is a standard to use the so-called McWhirter criterion that is necessary but however not sufficient to ensure complete and partial LTE within the LIP [12]. The McWhirter criterion is as follows [2]:

$$n_e \geq 1.6 \cdot 10^{12} {T_e}^{1/2} (\Delta \mathrm{E})^3, \tag{1.6}$$

where $n_e\ (cm^{-3})$ is electron density, $T_e\ (K)$ is electron temperature, and $\Delta\mathrm{E}\ (eV)$ is the highest energy of the quantum transition for which the condition is valid. This criterion implies that atomic and ionic quantum states are populated and depopulated mainly by collisions. When this equation holds, the non-equlibrium of radiative energy can be neglected [2] and for each part of the plasma plume it is possible to find a temperature that satisfies the Boltzmann, Saha and Maxwell distributions. It is noteworthy that McWhirter criterion is valid only in rudimentary stages of the LIP formation when the electrons and ions form the major part of the plasma plume composition.

When the condition for LTE is fulfilled the population of species in excited quantum states n_i^s obeys the Boltzmann distribution [2]:

$$n_i^s = \frac{g_i}{U^s(T_e)} n^s e^{-E_i/kT_e}. \tag{1.7}$$

Where k is the Boltzmann constant, g_i and E_i are the statistical weight and the excitation energy of the excited quantum state i. n^s is the total number density of the species s in the plasma and $U^s(T_e)$ is the internal partition function of the species at the temperature T_e.

The equation (1.6) requires the knowledge of the plasma parameters, electron temperature and electron density. Those parameters can be calculated from the spectral information observed by collecting the plasma radiation. There exist many kinds of temperatures and many ways how to obtain their values. A LIP is considered to be in LTE and to be optically thin. The computation of the electron temperature T_e is based on the Boltzmann plot method for selected elemental lines [2]. Those lines are of the same element found in the same ionization stage but originating from different upper energy levels. As can be derived from the Boltzmann distribution, the integrated line intensity I_{ij} (number of transitions per unit volume per unit time) is as follows [2]:

$$I_{ij} = n_i^s A_{ij} = \frac{A_{ij} g_i}{U^s(T_e)} n^s e^{-E_i/kT_e}. \tag{1.8}$$

The Boltzmann plot method requires the information about the intensity of more elemental lines of the same element. Then we can obtain a temperature, which is responsible for the distribution of the species population among those quantum states. Applying the natural logarithm the equation (1.8) can be transformed into a well-known formula [2]:

$$\ln\left(\frac{I_{ij}}{A_{ij} g_i}\right) = \ln\left(\frac{n^s}{U^s(T_e)}\right) - \frac{E_i}{kT_e}. \tag{1.9}$$

Further, the Boltzmann plot, see figure 9, can be depicted based on the information about the intensity of several spectral lines of the same element originating from different quantum transitions. The Boltzmann plot has a slope of $-1/kT_e$, therefore the plasma temperature can be evaluated utilizing linear regression, without the knowledge of n^s or $U^s(T_e)$.

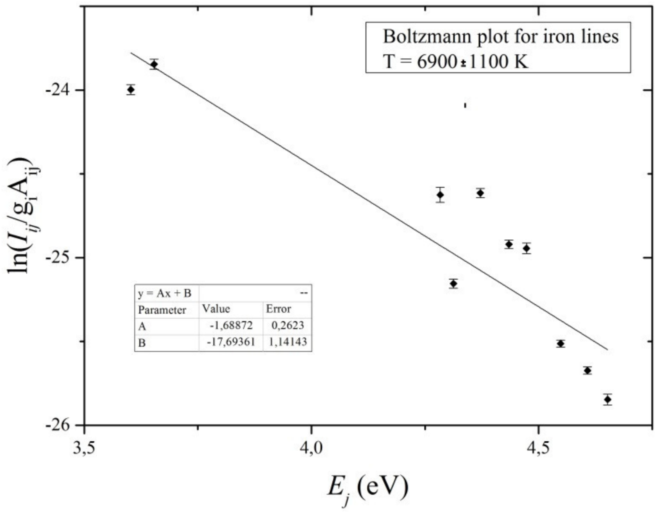


Figure 9, temperature estimated based on atomic iron lines, in the sample Oreas 111, via Boltzmann plot.

Nevertheless, the precision of the Boltzmann plot method can be in some cases very poor. To improve the reliability of the plasma temperature estimation the species of the same element in different ionization states are utilized together. This increases the range of upper energy states utilized for the computation. The Saha-Boltzmann distribution of the atomic and ionic species in excited state is described by an equation [2]:

$$\frac{I_{mn}^{II}}{I_{ij}^{I}} = \left(\frac{A_{mn}^{II} g_m^{II}}{A_{ij}^{I} g_i^{I}}\right) \left(\frac{2(2\pi m_e k T_e)^{3/2}}{n_e h^3}\right) e^{-\frac{\left(E_{ion} - \Delta E_{ion} + E_m^{II} - E_i^{I}\right)}{kT_e}}, \tag{1.10}$$

Where h is the Planck constant, the superscripts I and II refer to the parameters of atomic and singly ionized species, respectively. E_{ion} is the first ionization potential and ΔE_{ion} is the lowering correction parameter. Applying natural logarithm to the equation (1.10) as in the case of Boltzmann plot method, we can derive [2]:

$$\ln\left(\frac{I_{mn}^{II} A_{ij}^{I} g_i^{I}}{I_{ij}^{I} A_{mn}^{II} g_m^{II}}\right) = \ln\left(\frac{2(2\pi m_e k T_e)^{3/2}}{n_e h^3}\right) - \frac{\left(E_{ion} - \Delta E_{ion} + E_m^{II} - E_i^{I}\right)}{kT_e}. \tag{1.11}$$

As in the previous case, the logarithmic ratio of several atomic and ionic lines as the function of their energy differences results in a plot whose line slope depends on the electron temperature. The benefit of this method compared to the foregoing one is in the higher difference of upper energy states. Therefore, the slope is less sensitive to the measurement noise and inaccuracies coming from the computation. Moreover, the intercept refers to the electron density [2]:

$$n_e = \left(\frac{2(2\pi m_e k T_e)^{3/2}}{h^3}\right) \left(\frac{I_{ij}^{I} A_{mn}^{II} g_m^{II}}{I_{mn}^{II} A_{ij}^{I} g_i^{I}}\right) e^{-\frac{\left(E_{ion} - \Delta E_{ion} + E_m^{II} - E_i^{I}\right)}{kT_e}}. \tag{1.12}$$

The estimation of the temperature based on the Saha-Boltzmann equation is more complicated, but also more precise. Tognoni *et al.* [59] introduced upgraded so-called Calibration-Free LIBS approach, where the temperature is calculated from the Saha-Boltzmann equation as follows:

$$\ln\left(\frac{I_{mn}^{II}}{A_{mn}g_m}\right) = \ln\left(\frac{F^{abs}n^P C^I}{U^s(T_e)}\right) + \ln\left(\frac{2(2\pi m_e)^{3/2}(k_B T)^{3/2}}{n_e h^3}\right) - \frac{(E_{ion}+E_m^{II})}{k_B T}, \tag{1.13}$$

where F^{abs} is the wavelength-independent factor, n^P is the total particle density in plasma, C^I is the relative concentration of emitting species in the plasma, and the other symbols were already defined. However, this equation demands the knowledge of electron number density and plasma temperature prior to the calculation. In this case, the electron density is estimated independently and the value of temperature is guessed and then refined by iteration.

As it was already mentioned, the composition of the LIP is rapidly changing with time. The higher concentration of electrons and ionized species in the early stages of the plasma causes local electric fields. Consequently, the presence of the electric field affects the quantum transition of the particle from the excited state resulting in the shifted wavelength of the emitted photon. This phenomenon is called the Stark effect [2], its presence is observable by broadened spectral lines in the obtained spectrum. The full width at half maximum (FWHM) refers to the Stark broadening of the line $\Delta\lambda_{Stark}$ (nm) [2]:

$$\Delta\lambda_{Stark} = 2w\left(\frac{n_e}{10^{16}}\right) + 3.5A\left(\frac{n_e}{10^{16}}\right)^{1/4}\left[1 - BN_D^{-1/3}\right]w\left(\frac{n_e}{10^{16}}\right), \tag{1.14}$$

where B is a coefficient equal to 1.2 or 0.75 for ionic or neutral lines, respectively, w is the electron impact parameter (or electron impact half-width), and A is the ion broadening parameter and values of w can be found in literature [60]. The first term on the right side of the equation represents the electron interactions and the second one comes from the ion interactions, this term may be neglected in typical LIBS experiment and the equation (1.14) is then transformed to the following form [2]:

$$\Delta\lambda_{Stark} = 2w\left(\frac{n_e}{10^{16}}\right). \tag{1.15}$$

The broadening by other sources (natural, Doppler, instrumental, etc.) are usually omitted. The Stark width of the line is then directly estimated as the width of selected spectral line. Therefore the electron density can be directly evaluated from the Stark broadening. For instance, the electron number density was calculated from the broadening of atomic iron line Fe (I) 538.34 nm by Bengoechea *et al.* [61].

To conclude this part, LIP properties are strongly dependent on the energy of laser beam, its duration and wavelength, laser matter interaction, state of the matter and its composition, and ambient conditions. During the temporal evolution of the plasma the temperature can reach, in maximum, 6000 – 20000 K and the electron density 10^{17} – 10^{19} cm^{-3} [58]. The composition of the plasma is very inhomogeneous due to its rapid spatial expansion and the properties of the plasma (such as temperature and electron density) vary with the position along the plasma plume. The electron temperature T_e (K) and the electron density n_e (cm^{-3}) can be obtained from the plasma emission data. For the computation of the electron temperature the so called Boltzmann plot method is preferred in the field of LIBS. Assuming on those parameters, laser-induced plasma can be found in the state of local thermodynamic equilibrium for reliable quantitative analysis. [2, 4, 12, 14, 58, 62]

1.5. QUANTITATIVE ANALYSIS

A LIP consists of the elements forming the sample under study and ambient gas. The emission of a LIP is represented by a spectral line of characteristic wavelengths. Therefore, detected radiation of a LIP gives overall information about the elemental composition of a sample (so-called chemical fingerprint). Every sample is represented by the unique collection of spectrally resolved lines. The quantitative analysis can be done when inspecting the intensities of the lines in LIP spectra. However, theoretical assumptions have to be made to ensure the reliability of the analysis.

Quantitative analysis is one of the limitations of the LIBS technique [13]. The main limitation is the matrix effect, discussed in chapter 1.6. Further, LIBS lacks the repeatability of the measurement which is usually utilized in repetitive single spark mode. Moreover, the diversity of LIBS instrumentation makes the quantitative intra-comparison and reproducibility of the measurements among different LIBS groups complicated. Regardless of the limitations, the LIBS sensitivity is well under the ppm level for trace elements in solids or liquids. In general, small shot-to-shot fluctuations are obtained when ablating small amount of material, even from the surface of the homogenized standard sample. This uncertainty is described by a relative standard deviation (RSD) and usually yields in few percent. Signal fluctuations arise from the non-uniform conditions for LIP formation (such as the instability of the laser energy, sample roughness, and inhomogeneity of the sample composition). The latter is the main issue reducing the repeatability of the LIBS measurement.

In atomic absorption spectroscopy, the intensity of absorbed light can be related to the properties of the matter, where the light was absorbed (Beer-Lambert law) [63]. Analogically, the intensity of the detected spectral line can be related to the amount of corresponding element in the plasma plume. However, it is necessary to assume that the plasma is in LTE and detected lines are optically thin and that the laser ablation is stoichiometric, *i.e.* the composition of the plasma plume reflects the composition of the spot from which it was induced [2]. The detected intensity can be related to the amount of emitting particles in the plasma plume and consequently to the amount of particles ablated from the sample surface, *i.e.* representing the sample composition [12]. Therefore, the calibration curve, curve of growth, can be constructed via regression analysis where the intensity of the elemental line is plotted as a function of the concentration of a corresponding analyte. A set of standardized samples, with known concentrations of the element of interest, have to be measured to calibrate the system for further analysis of unknown samples as well as to establish the limits of detections of the LIBS system.

The parameters of the calibration curve are used for the computation of the limits of detection in parts per million (ppm). LOD signifies the lowest concentration at which we can decide whether the element is present or not [64]:

$$LOD = 3\frac{\sigma}{S}, \tag{1.16}$$

where σ (counts or a.u.) is the standard deviation of the background in the blank sample, the sample where the analyte is not present, or near the position of the peak selected for analysis when no blank sample is at hand. S is the slope of the linear part of the calibration curve. The above presented method the 3σ-IUPAC definition is the most spread in the LIBS field [2]. Furthermore, other concepts to calculate LOD are recommended to use in LIBS application [13]:

$$LOD = k_C RSD_B BEC, \tag{1.17a}$$

where $k_C = 3$ corresponds to 33.33% uncertainty, RSD_B is the relative standard deviation of the background. BEC is the background equivalent concentration. Equation (1.17a) can be further altered and then the equivalence with equation (1.16) is apparent [13]:

$$LOD = k_C \left(\frac{\sigma}{\bar{X}_B}\right)\left(\frac{\bar{X}_B}{S}\right), \tag{1.17b}$$

where $\bar{X}_B$ is the intensity of the background. LOD is the point, where we can distinguish signal of an analyte from the background not the point referring to the smallest concentration that can be measured. The limit of quantification (LOQ), or the limit of determination, is agreed to be:

$$LOQ = 3.3LOD \approx 10\frac{\sigma}{S}. \tag{1.18}$$

The slope of the calibration curve denotes the sensitivity of the measuring system. In general it is valid, the steeper the slope, the more sensitive the calibration curve. Linearity means that as the amount of an element increases, so does the related intensity, up to a point called limit of linearity (LOL). Then as the amount of the analyte is increased the intensity levels off and calibration curve loses its sensitivity. It is therefore crucial to work on the linear part of the curve to get reliable data for the unknown amount of the analyte present.

The flattening of the calibration curve is caused by the saturation of a selected elemental line. The linearity of the calibration curve breaks when the elemental line becomes saturated, *i.e.* the intensity of the elemental line is not linearly dependent on the content of the element in the samples. For this reason, a spectral line of the analyte has to be picked carefully for the calibration process. In typical LIBS plasma more lines of the same element are detected. Lines originating from non-resonant quantum transitions are recommended for the calibration curve formation, while resonant lines get easily saturated for lower concentrations. Resonant lines should be used warily for trace analysis because of their higher sensitivity. Beneficially, for sample sets with broader range of analyte concentrations several parts of calibration curve may be used to satisfy linearity.

It has to be stressed that the plasma properties differ with the laser-matter interaction. This complex phenomenon is called the matrix effect. The intensity of spectral lines of the same element (of the same amount) can differ significantly when measured in two samples with different matrices. Matrix effect is identical within each individual class of matrices. Thus, for each matrix class a unique calibration curve can be determined. This is then valid for all samples within this class.

So far only one spectral line was considered for construction of the calibration curve. Multivariate methods exist, in which more lines can be applied to improve the LOD, to partially compensate the matrix effect and/or to extend the dynamic range of the calibration curve, *i.e.* to linearize the calibration curve. Lines of the same element are usually applied to the advanced statistical algorithms, such as principal component regression (PCR) and partial least squares regression (PLSR), discussed further in chapter 2.

The LIP spectrum can give, under supervised conditions, qualitative and quantitative information on the sample under study. However, for the quantification process the matrix effect has to be taken into account. There are many ways how to compensate or to avoid the matrix effect. Another issue is the sensitivity of the LIBS technique, which can be improved, for instance by utilizing multiple pulses for plasma generation and/or by utilizing LIBS with laser-induced fluorescence spectroscopy (LIFS) [13].

Hahn and Omenetto [13] concluded the issue of quantitative analysis utilizing LIBS as follows:

> *"The problem is not that LIBS cannot perform quantitative analysis, but rather how to make LIBS become as accurate and reliable as the other spectroscopic methods. If quantitation is indeed the only vulnerable aspect of LIBS, this is where research should then be focused and vigorously continued."*

1.6. THE MATRIX EFFECT

It is well known fact that LIBS is strongly affected by a matrix effect [2, 4, 13, 65, 66, 67]. The composition of the sample, mainly in the sense of matrix or macro elements, is crucial in the laser/matter interaction and consecutive LIP formation and emission. It is common for emission spectroscopy techniques, detecting the analyte of the same amount in two samples with different matrices results in significantly different signal intensities. Consequently, each matrix class has its own calibration curve. Then general calibration of the system for various matrices is therefore not possible when univariate approach is considered. Therefore, it should be discussed how to compensate the matrix effect or even how to avoid its occurrence.

Matrix effect affects the properties of LIPs, whose spectrum may be further used in processing of obteained spectra. The goal of normalization is to unify the analyte signal from various samples for universal quantitative analysis. The most spread approach is to normalize the selected elemental line of an analyte to the intensity of the matrix element line or to the total plasma emission. Other parameters (such as T_e, n_e, acoustic signal from shockwave, the amount of ablated mass) can be advantageously utilized for normalization of analytical signal. Nevertheless, this thesis is focused on the *in-situ* and real-time measurements where the measurements of ablated mass and acoustic signal are whether complicated or even impossible. Moreover, the computation of plasma parameters (T_e, n_e) is tedious and with an uncertain output. Therefore, utilizing the plasma parameters for normalization purposes of LIP spectra is dropped from further consideration.

As is discussed in the following chapters of this thesis for the case of soil samples, each sample is described by the unique collection of elemental lines, so called chemical fingerprint. Then, samples can be distinguished on the basis of their matrix elements and assigned to appropriate groups. Individual calibration curves are created for those groups to partially avoid the matrix effect. This approach needs data libraries of related sample matrices prior the analysis of unknown samples [7].

Laser/matter interaction is the main issue causing the matrix effect. Ablation source with constant excitation conditions can supress the matrix effect in quantitative analysis employing LIP. The goal is to achieve matrix independence in a stoichiometric ablation. For instance, femtosecond lasers are capable of producing the LIP with comparable conditions for various materials. Zhang *et al.* [68] brought an extensive study comparing ns- and fs-lasers. Resulted calibration curves showed minimized matrix effect in the measurement of various matrices (soil, ore, metallic alloys) utilizing fs-LIBS. fs-lasers can be seen as ideal plasma sources, however, they are not as affordable as ns-lasers, which are predominantly spread within the LIBS community.

Laser ablation (LA) LIBS [13] can be used to decrease the matrix effect. The ablation step is separated from the analytical step by employing two different laser sources. The small amount of sample is ablated and in the form of aerosol carried to the interaction region, where it is reablated by another laser pulse. No matrix effect is then expected when the coupling of second laser remains consistent.

The calibration-free LIBS [3] is a method, which does not require measurement of the standardized samples prior to the quantitative analysis. The idea is to make each measurement self-consistent by applying LIP parameters to the spectral data and, in consequence, to compensate for the matrix effect. However, this promising technique is beyond the scope of this doctoral thesis.

The LIP spectrum containing atomic and ionic emission may provide qualitative and quantitative information about the elemental composition of the sample in real-time and *in-situ*. However, the occurrence of a matrix effect has to be considered. Therefore, LIBS device should be calibrated prior to the analysis according to the matrix of the interest when supervised standards are measured. (steel and other metallic alloys, soils, ceramic, etc.). Moreover, further analysis employing advanced statistical algorithms can improve the sensitivity and reduce the dependency of the signal on the matrices of various samples. The multivariate calibration and classification of the samples based on their characteristic spectral appearances prior the quantification is further discussed in chapter 4.

1.7. SIGNAL PREPROCESSING

In advance to the quantification analysis or classification, detected spectra should be pre-treated with a unified standardized algorithm. This chapter summarizes the basic and recommended steps in the data pre-treatment process prior a statistical analysis [69, 70, 71]. It is recommended to consider mean-centering and scaling of the data that can strongly affect the resulting model.

LIBS may be utilized in shot-to-shot regime when each consecutive laser pulse is utilized separately in the analysis, *e.g.* mapping [43]. However, in order to obtain robust statistical data set, each sample should be measured with considerably high number of repetitions. Conventionally, higher number of laser pulses per sample is employed to compensate for the fluctuations in LIBS measurement. In the first step, measured spectra are normalized (to a selected spectral line, to unity, or to the sum of overall intensity of a spectrum) in order to reduce the fluctuation in the intensities of spectral lines. Moreover, internal standardization of trace element lines to a matrix line leads to more a reliable comparison of samples with a similar composition. When choosing a matrix line, a transition with its upper quantum level energetically close to the upper level of the transition to normalize should be used in order to minimize uncertainties caused by temporal fluctuations within the plasmas. The smoothing is not necessary step in the pre-treatment. The background is subtracted for each line individually. Every element is represented by several lines originating from different quantum transitions. Selection of more elemental lines of the same element is profitable for classification and quantification analysis while the larger range of quantum transitions is more likely to describe plasma properties and a LIP temporal evolution. The shape of the spectral line can be fitted with pseudo-Voigt profile and the intensity is then calculated as the area under the peak. Spectra for each sample can be averaged, depending on the further data analysis.

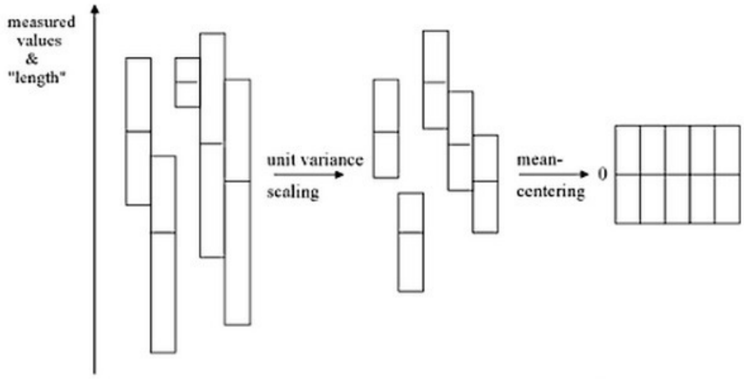


Figure 10, scaling and mean centering of the data matrix ***X*** **[69]**.

Individual samples are described by a set of variables, such as matrix elements or elements selected for a trace analysis. Variables are represented by elemental lines and are organized in the columns of the data matrix $\boldsymbol{X}\ (I \times J)$, where the rows are assigned to samples. In the case of regression analysis, matrix $\boldsymbol{Y}\ (I \times K)$ is constructed. Usually in the calibration process, there is only one variable representing the amount of the element selected for trace analysis. This matrix is then transformed to a vector $\boldsymbol{y}\ (I \times 1)$. Those data matrices are then further processed by chemometric algorithms described in following chapter.

Variables are often measured with different numerical ranges and it can happen that one variable over-shadows the others. In general, variables are scaled to a unit variance and their dependence of the model is equalized. This step should be done with care when scaling of the variable with a very low variance can introduce a high error into the computation. In the second step, data matrix is mean-centered (*i.e.* the vectors of interest are set around the origin of the coordinate system) for the improved interpretability of the model. Note, scaling and mean-centering of the data matrix, depicted in figure 10, are recommended in any case at least for the first step of the analysis, however, their contribution to the modelling of the data matrix does not have to be necessarily beneficial [69]. Advanced data pre-treament, such as transformation and data correction and compression, is not employed in the experimental part and therefore not discussed.

2. CHEMOMETRICS

In 1994, Svante Wold [72] tried to define the chemometric algorithms in general:

> *"A reasonable definition of chemometrics remains as: 'How to get chemically relevant information out of measured chemical data, how to represent and display this information.' "*

In other words, chemometrics are of great help when multivariate and complex data are acquired, as befits measurements employing spectroscopic techniques. Over past decades, chemometric algorithms have already proved their importance in the data mining and data analysis in various fields (analytical chemistry, economics, and biology). Chemometrics can serve for pattern recognition (classification) and quantification.

In general, the LIBS measurement results in bulky data sets, where each sample is represented by series of spectra (each spectrum possible to provide information on up to 60000 spectral features). Chemometric algorithms are established methods for classification and multivariate calibration of such complex and large datasets. It is expectable that the discrimination of the sample set represented by their LIBS spectra can emulate the distribution of the samples based on their physical and chemical properties. Utilizing chemometrics, unpredicted latent variables among apparently diverse samples could be as well found. Detail to data pretreatment and processing by the means of chemometric algorithms can be found in the chapter 1.7.

LIBS has already been used for quantitative analysis of mineral samples including the field-portable device [73, 74, 75, 76]. Multivariate statistical approaches for identification of different kinds of rocks and minerals were employed using both laboratory bench-top and stand-off LIBS systems [77]. Harmon *et al.* [77] proved the possibility of ascertaining the provenance of conflict minerals employing partial least square discriminant analysis of LIBS spectra. Bousquet *et al.* [78] tested the hypothesis that the most significant differences between soils come from the varying amounts of matrix elements.

In this chapter the main emphasis is given to the most spread chemometric algorithms, principal component analysis (PCA), partial least squares (PLS) regression and their variations, such as principal component regression (PCR), soft independent modelling of class analogies (SIMCA), and partial least squares discriminant analysis (PLS-DA). The discussion about one of the artificial neural network (ANN) algorithms is as well given, Kohonen's maps or self-organizing maps (SOM), which is seen as an alternative to PCA, however, its use can be beneficial due to a more advanced visualization. The main literature sources in following chapters are those books [70, 79, 69].

2.1. PRINCIPAL COMPONENT ANALYSIS AND RELATED ALGORITHMS

In 1901, Karl Pearson published the algorithm for transforming the data matrix with the aim to uncover the hidden correlated variables among the samples [80]. Early history of exploratory data analysis can be dated even to Cauchy and his studies in 1820s. Though this algorithm was then reinvented by Harold Hotelling [81] who called it the Principal Component Analysis, for the essential purpose of the algorithm to find the so-called principal components. PCA algorithm can be found under different names (factor analysis, eigenanalysis, eigendecomposition, and singular value decomposition (SVD)) hence the primary goal remains the same, to find latent variables in the data matrix. PCA discovers latent relationships between phenomenological results (in our case spectra) and intrinsic properties of the studied sample (here chemical composition). PCA is a projection method that creates new coordinate system formed by latent variables, *i.e.* the principal components, which describe major trends in data set in the least square sense.

TRANSFORMATION OF THE DATA MATRIX

The data matrix $\boldsymbol{X}$ consists of the experimental data, where each measurement is assigned a row and the relevant information, in our case wavelength, is organized into corresponding columns. In the PCA algorithm, the data matrix is decomposed and substituted by the product of two new matrices, the scores matrix $\boldsymbol{T}\ (I \times A)$ and loadings matrix $\boldsymbol{P}\ (A \times J)$, where A is the total number of estimated principal components. The projection of the points (samples, observations) in the X-space onto the principal component gives the related value, called a score. The scores represent the distribution of the points (samples) in the principal component space according to their variance and the loadings show the most important spectral variables, wavelengths, responsible for this variation. In this approach, the first created principal component describes the most of the variances in the data set. PCA reduces redundancy in the data set and creates a dataset, which can be visualized more easily. It is expected that samples from the same class will form one cluster, based on similarity in their spectra. [79, 82] Instead of the original multitude of variables that include this likeliness, however, the PCA transformed result confines all relevant dimensions that include spectral re-assemblance belonging to one class onto one single principle component. Thus, a multidimensional problem can get reduced to a single dimension, simply by neglecting irrelevant information and transforming the relevant information onto one dimension.

PCA involves an abstract mathematical transformation of the original data matrix and can be represented by the equation [79, 82]:

$$\boldsymbol{X} = \boldsymbol{TP} + \boldsymbol{E}, \qquad (2.1a)$$

$$x_{ij} = \sum_{a=1}^{A} t_{ia}\, p_{aj} + e_{ij}, \qquad (2.1b)$$

where $\boldsymbol{E}\ (I \times J)$ is the error matrix compensating the deviation of the model data to the original data set. The product of $\boldsymbol{TP}$ can be regarded as a model of the data that is an approximation to the original data set, when the error is represented by this error matrix $\boldsymbol{E}$. The index i describes the individual samples (observations), index j stands for the individual variables (wavelengths) and index a refers to the number of principal components. It is possible to calculate any number of PCs as desired (until the observed data matrix $\boldsymbol{X}$ is fully modelled within a given uncertainty) while the optimal number of PCs can be determined by various methods [82]. The maximal number of PCs is equal to the maximal number of input variables.

It is noteworthy that the equation (2.1a) is comparable with the equation used in SVD algorithm, which is used in a majority of available software for PCA [79, 82, 83]:

$$\boldsymbol{X} = \boldsymbol{UDV} + \boldsymbol{E}, \tag{2.2}$$

where $\boldsymbol{U}$ is identical with $\boldsymbol{T}$ and $\boldsymbol{V}$ is the same as $\boldsymbol{P}$, though the difference is in the scaling of the data for the individual matrices. $\boldsymbol{D}$ is a diagonal matrix which diagonal elements consist of the square roots of the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$.

The algorithm sets the new basis represented by equation (2.1) and schematically depicted in figure 11. The loadings matrix is a projection of the original data set onto a new basis. The scores give the new representation of the objects in the rotated coordinate space, figure 12. As can be seen, both created principal components are going through the origin of X-space, which is a consequence of the mean-centering of the data matrix. By projecting the points (samples, observations) onto a newly created basis (2 PCs space - a plane in the depicted case, or hyper plane for more PCs) it is possible to visualize the distribution of the points in data matrix. The PCA algorithm withdraws the highest variation among the data iteratively when creating the individual principal components. Always, the first principal component stands for the most variation among the data [83]. In mathematical terms, first principal component is the linear combination that best estimates the original variable in the sense of smallest least squares [84].

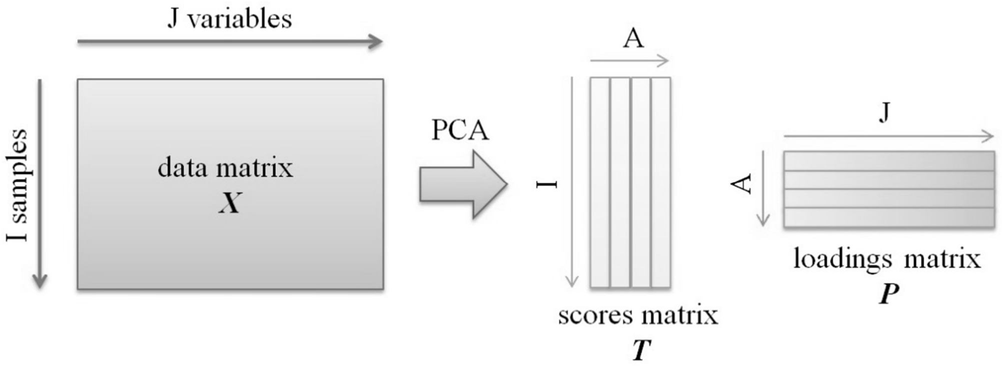


Figure 11, schematic of transformation of the data matrix by the means of PCA **[71]**.

To determine the importance of the PC, *i.e.* how much variance is described by the PC, is given by the number called eigenvalue. Principal components are organized in a descending manner according to their eigenvalues. The value λ_a of the *a*th eigenvalues are defined as the sum of squares [79, 82]:

$$\lambda_a = \sum_{i=1}^{I} t_{ia}^2. \tag{2.3}$$

It is usual to estimate the eigenvalues in percentages:

$$V_a = 100\frac{\lambda_a}{\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ia}^2}, \tag{2.4}$$

where an element in the denominator of the fraction refers to the sum of squares of the entire data matrix.

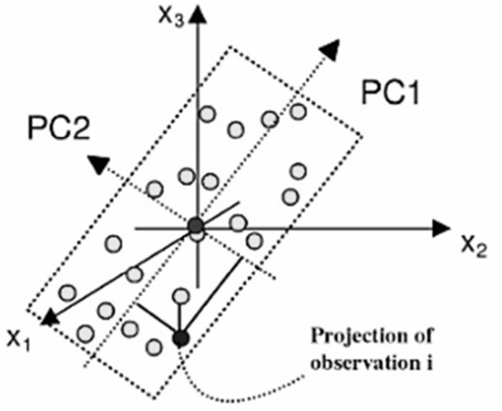


Figure 12, data matrix, represented by a three variables, projected into a new space by PCA **[82]**.

The number of used principal components is crucial for the resulted modelling of the data matrix. When not enough principal components are used, the data matrix is not well modelled. On the other hand, using too many principal components leads to an overfitting and to introduction of unwanted noise into the model. Therefore, the optimal number can be established via the cross validation (CV) algorithm [82]. Part of the data, individual observation, is kept out of the computation, model is created and the kept-out data are validated. Then, the predicted data are compared with the observed ones. The prediction sum of squares (PRESS), the sum of squared difference of predicted data $\hat{x}_{ij}$ and observed data x_{ij}, is obtained for every part of the data matrix, which is kept out of computation. The value of PRESS using *a* PC model is given by following equation [79]:

$$\mathrm{PRESS}_a = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\hat{x}_{ij} - x_{ij}\right)^2. \tag{2.5}$$

The error obtained from PRESS can be then compared to the residual sum of squares (RSS) error [79]:

$$\mathrm{RSS}_a = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}^2 - \sum_{a=1}^{A}\lambda_a. \tag{2.6}$$

The building of the model, *i.e.* computing new PCs, can be stopped when PRESS is not significantly smaller than RSS [82].

VISUALISATION

Individual variables can be cross-plotted in respect to each other in order to reveal possible latent variables and groups within the data set. However, this approach is not suitable when more variables are used for describing the data. PCA is then applied and individual principal components are used for visualization of the spectrochemical patterns within the data matrix [82].

Scores of the principal components are then cross-plotted while preserving the dominant patterns among the original data. PCA, as a least squares algorithm, is severely affected by the outliers, data strongly deviating from the rest of the data set. Therefore, it is beneficial to apply the PCA without any further analysis but the search for outliers. Due to the reduced dimensionality, those can be unambiguously identified simply by inspecting plots based on the first few PCs. Then, after removing the disturbing outlying data points, the PCA analysis can be repeated with an improved accuracy. Loading vectors can be inspected for giving the information about the importance of the individual variable. Less significant variables can be discarded from the computation without changing the resulting model. Nevertheless, this step has to be done carefully with respect to the case of the study.

REGRESSION

Principal components can be further used for quantification analysis, in the so-called principal component regression (PCR) algorithm [84]. In the first step PCA algorithm is used on the obtained data matrix to reduce the dimensionality and to take into account the structure of the data matrix, in other words to create the principal components. The vector $\boldsymbol{y}$, representing observed outcomes, is regressed onto the score matrix $\boldsymbol{T}$ [71]:

$$\boldsymbol{s} = \boldsymbol{T}\boldsymbol{y}, \tag{2.7}$$

where $\boldsymbol{s}$ is the vector of regression coefficients, whose dimensions are equal to the number of selected principal components. PCR model is then expressed by the regression coefficients vector $\boldsymbol{s}$ and loadings matrix $\boldsymbol{P}$, acquired in the first step. After transforming the vector $\boldsymbol{s}$ back using PCA loadings, the estimation of regression coefficients is possible with characterization of the original model.

CLUSTERING AND CLASSIFICATION

In the ideal case, the samples in one class have the same properties, their variables have similar values and there is no error coming from the measurement; moreover the classes are distinct. However, an ideal case will never occur and values in the variables describing individual samples are changing continuously; classes can overlap. Moreover, the distribution of the samples and individual observations is set to follow Gaussian distribution [79]. However, this is only an assumption based on no physical reasoning while in a real data space the data points do not necessarily obey the Gaussian distribution.

Based on the differences described by selected variables the samples can be discriminated into individual classes (groups, clusters). The classes can be known in advance or the samples can be assigned to groups based on PCA. The patterns among the data can be revealed by inspecting the score plots designed by a PCA algorithm. PC model can recognize any pattern which is present within a data set [85]. The possible problems which can occur during the classification process and the ways how to avoid those problems are further discussed on a real data, chapter 4.

For the purposes of supervised classification the algorithm based on disjoint PCA models was introduced [85]. The algorithm introduced by Svante Wold and co-workers is called Soft Independent Modelling of Class Analogies (SIMCA). In this algorithm, the model is designed by the means of

PCA for each class individually (referred to as ‘independent modelling’). The usefulness of this method is seen in the data reduction by PCA [86]. Essentially, the samples in each class should be very similar in a composition and hence different from the other groups. The total model of a ‘training’ data set consists of a collection of several disjoint class models. Then, ‘test’ samples are separately fitted to all of the class models and assigned to the model they fit best. It has to be kept in mind that the unknown sample, which fits into no known class might therefore create a new class. It can happen that one sample will be assigned to two classes (this feature is reflected in the name of the SIMCA algorithm by the term ‘soft’).

When the software is trained based on the training data, the classification power of the SIMCA algorithm can be estimated by assigning the unknown (validation) data. Leave-one-sample-out (LOSO) cross-validation can be applied when the validation data set does not exist [86]. Unknown samples are then classified by the means of their deviation to the individual PCA models. This deviation is given by the Euclidean distance of the unknown sample to the centre of the class in PC space. The Euclidean distance (orthogonal distance (OD)) is given by [86]:

$$\mathrm{OD}^c = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}^c\|, \tag{2.8}$$

where $\hat{\boldsymbol{y}}$ is the unknown sample and $\hat{\boldsymbol{y}}^c$ is the projection of this sample by the PCA model of class *c*. This procedure is repeated for all models of classes and the unknown sample is classified by comparing the square of the deviations $(\mathrm{OD}^c)^2$.

2.2. PARTIAL LEAST SQUARES ALGORITHMS

Multivariate analysis and calibration was strongly influenced by a contribution of Herman Wold, the Norwegian mathematician concerned about econometrics. H. Wold and co-workers introduced in the early 1970s the method called the partial least squares (PLS) model in latent variables (and/or projections to latent structures by means of partial least squares) [69, 87, 88]. H. Wold then developed the NIPALS (nonlinear iterative partial least squares) algorithm for estimating the values of parameters in PLS models. PLS algorithms have a long tradition in chemometrics, when some chemometricians apply PLS algorithms to nearly every kind of problem [79].

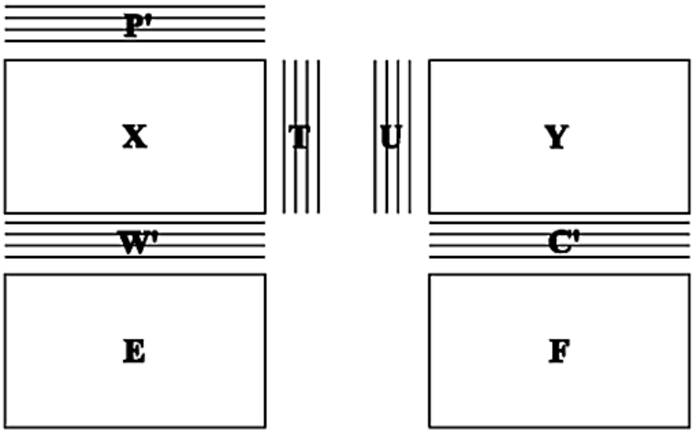


Figure 13, schematic of PLS **[69]**.

In other words, when PCA is extended to a data set divided into two or more matrices (containing dependent and independent variables) the algorithm is transformed to a so-called PLS analysis [82]. The main difference is that the PLS algorithm models as well the contribution in the variables in ***Y*** matrix, equations (2.9). This matrix can contain only one variable and then becomes a vector, equations (26). PLS then relates two matrices, the ***X*** matrix (containing the measured spectroscopic data, observations, and factors) and the ***Y*** matrix (consisting of classifiers, response, calibration data, and/or observed outcomes) [79, 89]:

$$\boldsymbol{X} = \boldsymbol{TP} + \boldsymbol{E}, \quad \boldsymbol{Y} = \boldsymbol{UC} + \boldsymbol{F}, \tag{2.9a}$$

$$x_{ij} = \sum_{a=1}^{A} t_{ia}p_{aj} + e_{ij}, \quad y_{ik} = \sum_{a=1}^{A} u_{ia}c_{ak} + f_{ik}. \tag{2.9b}$$

where the matrices $\boldsymbol{X}$, $\boldsymbol{T}$, $\boldsymbol{P}$ and $\boldsymbol{E}$ have the same meaning as in the equation (17a). The matrix $\boldsymbol{U}$ $(I \times A)$ is a score matrix in Y-space, and can be considered equal to the matrix $\boldsymbol{T}$ [69, 71]. The matrix $\boldsymbol{C}$ $(A \times K)$ is a weight matrix and the matrix $\boldsymbol{F}$ $(I \times K)$ is analogical to an error matrix in PCA. However, the matrices $\boldsymbol{T}$, $\boldsymbol{P}$ are different from those obtained via PCA. Moreover, the loadings obtained via PLS algorithm are not orthogonal as in the case of PCA [79]. PLS is schematically depicted in the figure 13.

To conclude the terminology [79], when there are multiple variables in the $\boldsymbol{Y}$ matrix, then the algorithm is in the literature called PLS2. If there is only one variable in the matrix $\boldsymbol{Y}$, then the equations (2.9) can be simplified to the form:

$$\boldsymbol{X} = \boldsymbol{TP} + \boldsymbol{E}, \qquad \boldsymbol{y} = \boldsymbol{Uc} + \boldsymbol{f}, \tag{2.10a}$$

$$x_{ij} = \sum_{a=1}^{A} t_{ia} p_{aj} + e_{ij}, \qquad y_i = \sum_{a=1}^{A} u_{ia} c_a + f_i. \tag{2.10b}$$

where $\boldsymbol{y}$, $\boldsymbol{c}$, and $\boldsymbol{f}$ are vectors. This algorithm is named PLS1 or PLS regression (PLSR). PLS-DA (partial least squares discriminant analysis) is used for data classification where the values in the vector $\boldsymbol{y}$ refer to classifiers or the numbers of a class membership of samples under investigation. PLS-DA is algorithm used for pattern recognition and classification. Then, the model prediction of $\boldsymbol{y}$ is a product of multiplication: $\hat{\boldsymbol{y}}_1 = \boldsymbol{t}_1 \boldsymbol{c}_1$. The additional principal component is computed from the residual vector: $\boldsymbol{f}_1 = \boldsymbol{y} - \hat{y}_1$. In the case of PLSR, the first PLS component is calculated similarly as in the PCA. However, the PLS component provides both, the approximation of points in X-space and correlation with the y-vector. The orientation of the first component is depicted in figure 14.

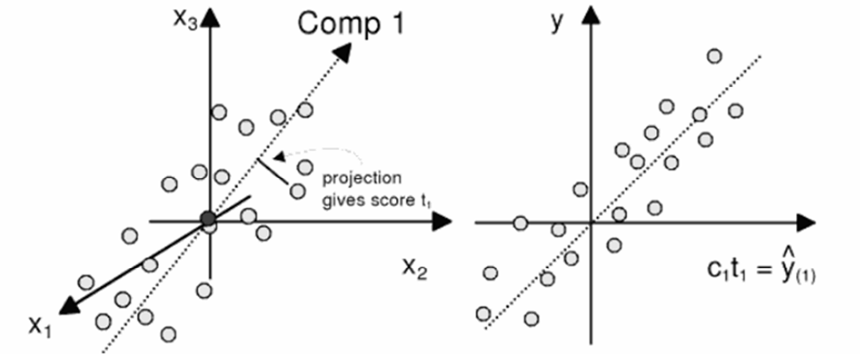


Figure 14, the first PLS component **[69]**.

Additionally the weight matrix $\boldsymbol{W}^*$ $(J \times A)$ is created as a consequence of PLS algorithm. The weight vectors for each component, latent variable, describe the way in which the variables in data matrix $\boldsymbol{X}$ are organized to form the score vectors [79, 90]:

$$\boldsymbol{T} = \boldsymbol{XW}^*. \tag{2.11a}$$

$$t_{ia} = \sum_{j=1}^{J} x_{ij} w_{ja}^*. \tag{2.11b}$$

Note, that the weight matrix $\boldsymbol{W}^*$ relates directly to the data matrix $\boldsymbol{X}$ and the weight matrix $\boldsymbol{W}$ relates to the residual matrix $\boldsymbol{E}$, computed for the previous dimension A [69].

The equation for a general solution of regression employing PLS algorithm is as follows [69, 70]:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{f}, \tag{2.12a}$$

$$y_i = \sum_{j=1}^{J} x_{ij} b_j + f_i, \tag{2.12b}$$

where the PLS regression coefficient vector $\boldsymbol{b}$ gives the relation between the PLS weights [69, 70]:

$$\boldsymbol{b} = \boldsymbol{W}^{*}\boldsymbol{c}, \tag{2.13a}$$

$$b_j = \sum_{a=1}^{A} w_{ja} c_a. \tag{2.13b}$$

It is valid for all of the PLS algorithms, once the PLS model is established with training data, unknown samples can be classified or quantified using the computed PLS weight matrices. The unknown concentration can be predicted by a dot product of the measured spectra of the unknown sample and the PLS regression coefficient vector.

The number of PLS components in the model as well as the prediction error of the model are estimated with the same algorithms as in the case of PCA. The size of each PLS component can be calculated, analogically as eigenvalues in PCA [79]:

$$\kappa_a = \left(\sum_{i=1}^{I} t_{ia}^2\right)\left(\sum_{j=1}^{J} p_{aj}^2\right). \tag{2.14}$$

However, the size of κ_a of successive components does not have to be in decreasing manner, as in the PCA case. This is an essential feature of the PLS algorithm in which both $\boldsymbol{X}$ and $\boldsymbol{c}$ blocks are modelled. The prediction power of PLSR and PCR is compared in the result section of this thesis.

Erikson et al. [69] concluded the difference between PCA and PLS as follows:

> *"PCA is a maximum variance least squares projection of* ***X****, whereas the PLS is a maximum covariance model of the relationship between* ***X*** *and* ***Y****."*

2.3. KOHONEN'S SELF-ORGANIZING MAPS

The self-organization is in the sense of neural origin, which invention was inspired by the processes in the brains of animals and their consequent behaviour excited by external stimuli. This algorithm was adapted by a Finnish researcher Teuvo Kohonen [91, 92] to formerly existing artificial neural network algorithms to be capable of learning and pattern recognition. It is noteworthy, that there exist two similar approaches how to construct self-organizing maps, however, in this chapter only the Kohonen's model will be discussed while the Wilshaw-van der Malsburg model will not be considered.

Kohonen's maps or Self-Organizing Maps (SOM) are an alternative approach (nonlinear generalization) to PCA, where the benefit is in advanced visualization of characteristic variables, specific samples and/or groups of samples. SOM can be effectively applied when large data sets or a vast number of variables are applied to the analysis. Such as PCA, SOM offers low-dimensional visualization of high-dimensional data. The algorithm is based on the unsupervised learning while following the initial topological distribution of the data in the input patterns. Processing of the data set and the application of the SOM is schematically depicted in figure 15.

INITIALIZATION AND TRAINING

In the data mining process, chapter 4, the software designed by Juha Vesanto [93] is applied to the data modelling by the means of SOM and PCA. For the purposes of the numerical computation and for the theoretical explanation, the data matrix **X** has to be organized in the manner that each of *I* rows is a j-dimensional data vector $\mathbf{x}_i = [x_1, \dots, x_j]$ corresponding to one sample. The intensities of individual spectral lines obtained from the measurement of the samples are organized in the number of j columns of the data matrix. Data in each column are normalized to unity in order to unify the measured data for more precise analysis.

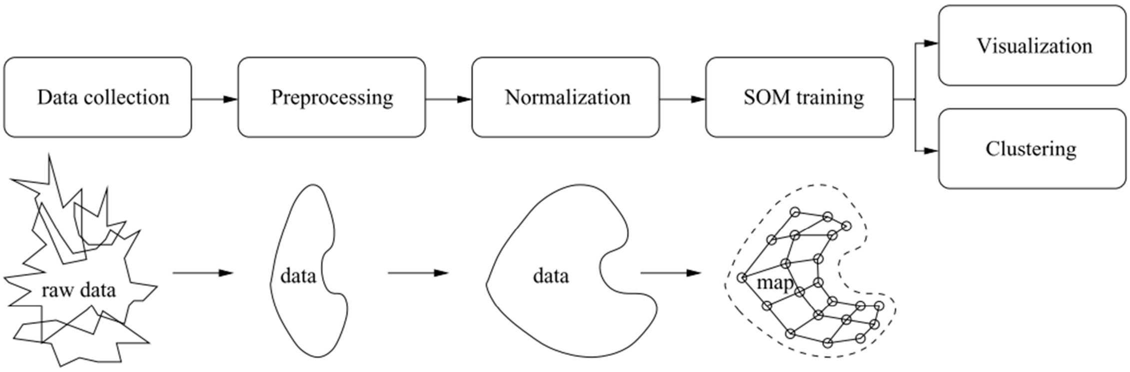
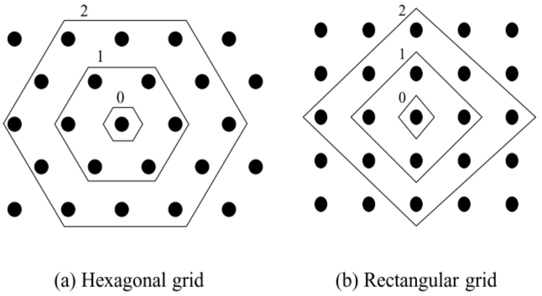


Figure 15, the application of the SOM in data mining. After data collection, the data is preprocessed, normalized, and a SOM is trained. This thesis is aimed on the SOM: visualization, clustering **[93]**.

SOM consists of L units, neurons, organized in low dimensional grid (usually 1D or 2D, but higher dimensions are as well possible, however, not used due to a problematic visualization). Every neuron is fully interconnected with the input. This connection is represented by a weight (prototype) vector $\mathbf{w}_l = [w_{l1}, \dots, w_{lj}]$ for each neuron $l = 1,2, \dots L$. The number of weights corresponds to the number of variables describing the sample. The unit positions $\mathbf{r}_l$ locates the position of the neuron l from the beginning of the grid.



(a) Hexagonal grid (b) Rectangular grid

Figure 16, SOMs with the size of 5x5 neurons. The numbers indicates the area of the neighborhood of each neuron **[93]**.

The neurons are organized in the maps with hexagonal pattern, which is the most common, or with rectangular pattern, figure 16. The number of neurons is selected prior to the analysis, their weight vectors are randomly generated but values are in the range of those in the data matrix. Because the initial weight vectors and positions of the neurons in the map are random such as the choice of the samples given to the system to be trained, therefore, the whole process is irreproducible. Nevertheless, the training process follows still the same algorithm and the results are highly correlated. Moreover, values of the neighbourhood function $h_{b_i l}(t)$ and the learning rate $\alpha(t)$ have to be selected. Both properties are time dependent and get smaller with increasing number of iterations *t*.

When the training of the map is started, the essential processes in the formation of SOM are competition and cooperation. Random sample $\mathbf{x}_i$ of a data matrix $\mathbf{X}$ is given to the system and the values of $\mathbf{x}_i$ are compared to the weight vectors $\mathbf{w}_l$ of each neuron. The neuron with the weight vector most similar to the randomly selected sample vector is then considered as a winner or best matching unit (BMU). In other words, the winner has the smallest Euclidean distance to the selected sample. The BMU b_i is a function of input vector $\mathbf{x}_i$ and its value is equal:

$$b_i(\mathbf{x}_i) = \arg\min_l \|\mathbf{x}_i - \mathbf{w}_l\|. \tag{2.15}$$

The winning neuron then learns from the sample, *i.e.* adapts its weight vector. The degree of adaptation is determined by the learning rate. The winning neuron affects as well the neurons in its surrounding, which area is determined by the neighbourhood function. When the weight vectors are adapted, the algorithm is repeated by applying another randomly selected sample $\mathbf{x}_i$ and so on for $t = 1,2,\dots,T$ iterations. The measure of adaptation is given by following equation:

$$\mathbf{w}_l(t+1) = \mathbf{w}_l(t) + \alpha(t) h_{b_i l}(t)[\mathbf{x}_i - \mathbf{w}_l(t)], \tag{2.16}$$

The neighbourhood function has its maximum in the position of the winning neuron b_i, it is monotonically decreasing and symmetric. The neighbourhood function is therefore considered to be the Gaussian function, but not necessarily:

$$h_{b_i l}(t) = e^{-\frac{\|\mathbf{r}_{b_i} - \mathbf{r}_l\|}{2\sigma^2(t)}}, \tag{2.17}$$

where $\mathbf{r}_{b_i}$ and $\mathbf{r}_l$ is the position of the winning neuron and its neighbour, respectively. Note that the neighbourhood function is translation invariant, does not depend on the position of the winner. The size of the neighbourhood is then given as follows:

$$\mathcal{N}_{b_i} = \{l \,|\, \|\mathbf{r}_{b_i} - \mathbf{r}_l\| \leq \sigma(t)\}. \tag{2.18}$$

The neighbourhood radius $\sigma(t)$ shrinks gradually with the number of iterations, while in the begging covers larger amount of the neurons surrounding the winner and in the end encircles only the winner. As the system is more organized, the property of cooperation among the neurons disappears. The SOM forms a flexible net which folds and stretches during the training. In each iteration the winner and its neighbourhood is pulled to the same direction in order to mimic the topological distribution of the initial data matrix, figure 17.

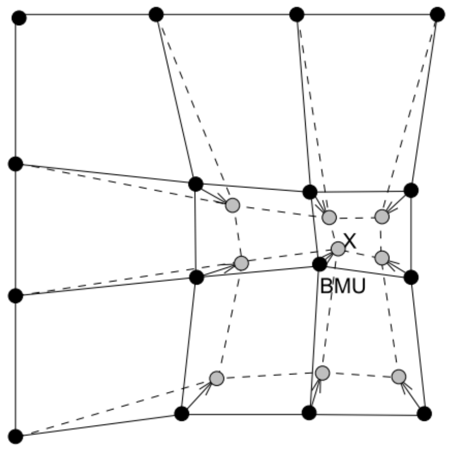


Figure 17, updating the best matching unit (BMU) and its neighbors toward the input sample marked with x.

VISUALIZATION

The unified distance matrix (U-matrix) shows the similarity of a neuron l to its neighbours. This approach of visualization leads to the revealing of potential clusters present in the map. The unified distance of each neuron is computed as the Euclidean distance among the weight vector of a neuron and weight vectors of neurons in the neighbourhood $\mathcal{N}_l$:

$$u_l = \sum_{h=1}^{H_l} \|\mathbf{w}_i - \mathbf{w}_l\| \qquad i \in \mathcal{N}_l; i \neq l, \tag{2.19}$$

where H_l is the number of neurons surrounding the neuron l in the given neighbourhood. Then neurons with similar weight vectors have lower value of the Euclidean distance. It is generally valid, the bigger the Euclidean distance, the more different the neurons. The U-matrix has similar properties to PC scores, where the clustering/grouping of the samples in the data set can be observed.

Another feature of SOM are so called component planes, they show how each variable affects the map and which variable is most associated with individual samples. Analogically to the U-matrix, a component plane is computed for each variable separately. The weights are converted according to the importance of the selected variable for describing the regions of the map. Component planes show whether and how the variable describes the class. The component planes can be used as well for clustering, however, the result has to be considered with care.

Concluding SOM, a high dimensional input space, data matrix, of activation patterns is mapped onto a discrete output space of neurons by the process of competitive learning. The resulted map is topologically similar to the distribution within the data matrix. Moreover, the unified distance matrix shows possible clustering among the data.

2.4. CHEMOMETRICS IN SPECTROCHEMICAL ANALYSIS

First of all, this discourse is concentrated on the application of LIBS to mining industry. For this reason, more detailed literature survey about the use of chemometrics for pattern recognition among the LIP spectra was done.

The group of A. W. Miziolek from the U.S. Army Research Laboratory has long-term experience in combining the outcomes of LIBS with the analysing power of chemometrics. One of the aims of this group is the detection of chemical and biological warfare agent stimulants (bacteria, explosives) and their further discrimination based on their spectral features [1, 94, 95, 96]. The stand-off detection of biological samples and their discrimination was as well tested [97]. The LIBS analysis on the biological samples can be found elsewhere [98, 99].

Group of M. Z. Martin coupled multivariate statistical analysis techniques with laser-induced breakdown spectroscopy to identify and classify various samples, such as preservative-treated wood [100], nuclear materials [101]. In several other groups were chemometrics used to distinguishing among the samples based on their spectral features; identification of biominerals among different kinds of samples [102], wheat grains [103]. Chemometrics were used with LIBS as well for improvement of the quantitative analysis [104].

The group of D. Cremers introduced as well the new algorithm for discrimination of biological samples into more groups by step-by-step approach [105]. Usually, the PCA algorithm struggle when the data set consist of more different groups. Then, the individual samples are not well resolved into the distinct groups and these groups tend to overlap when plotted into the PCA score plot. In this step-by-step approach, any cluster is removed from the computation when it is successfully assigned by the clustering algorithm. Then the PCA algorithm is applied again on the reduced dataset. This leads to simplification of the dataset and the PCA is not affected by too much variation. This process is repeated until all of the samples are successfully classified. This approach was employed for the analysis in other work [106].

Since PCA is unsupervised, it seeks to describe the overall variation in the data. This may not be useful for differentiating between different classes of samples. Therefore classification techniques such as SIMCA are more useful for describing data sets that consists of many measurements of several samples or classes. A SIMCA [86] model consists of a collection of PCA models. Each PCA model within the SIMCA model describes a particular sample type or class from the data set. SIMCA incorporates the properties of PCA models with information about the types of classes incorporated in the sample data set. The SIMCA model is then used to determine the nearest class for unknown test samples [39]. SIMCA algorithm was used to determine the class membership of unknown samples [1].

3. EXPERIMENTAL

LIBS SETUP

The experimental LIBS setup constructed in the laboratory at BAM in Berlin, Germany is depicted on the figure 18. The setup was utilized for measurements of all soil samples (certified reference materials and real samples collected directly at copper mines in north-west region of Iran, near Sungun village).

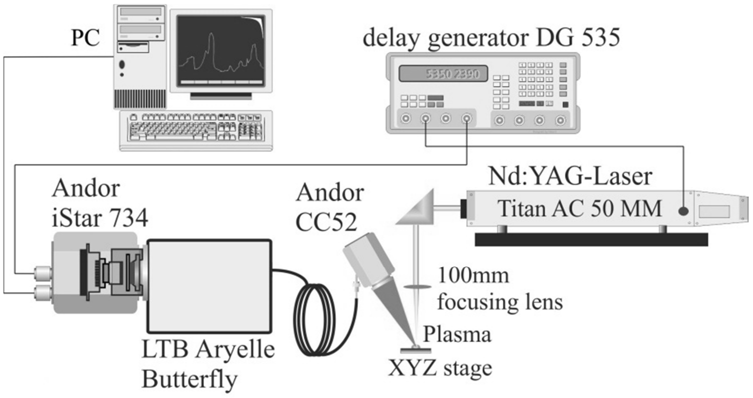


Figure 18, experimental LIBS setup constructed at BAM in Berlin, Germany.

In the first case study, high energy Nd:YAG laser Continuum (Surelite SLI-10, operated at 10 Hz, 1064 nm, 6 ns) was utilized for material ablation and plasma formation. However the laser was changed during the measurement process for the second and third case study, the rest of the setup was kept unchanged. The LIBS setup consisted of a diode pumped solid state (DPSS) Nd:YAG laser (Titan AC 50 MM, Atum lasers, Germany; operated at 10 Hz, 1064 nm, 10 ns). Laser pulse was focused from above onto the sample using a planoconvex lens (f_L= 100 mm) into a tight spot ~0.2 mm in diameter (0.25 mm when laser Continuum was utilized). After the impact of short laser pulse to the surface of a sample a luminous LIP was created. The radiation from a LIP was collected with a large aperture collector-collimator (Andor CC52, f/2) placed 250 mm form the interaction region at a 30° angle with respect to the incident laser-beam-path. The collector was coupled to a 400 µm optical fibre which delivered light at the entrance slit of an echelle spectrometer. Two wings echelle spectrometer Aryelle Butterfly (LTB, Germany; resolving power of 25000 in the first wing and 15000 in the second one) dispersed the plasma light in the range from 190 to 420 nm in the first wing and from 300 to 950 nm in the other wing.

The spectral information was recorded by an ICCD (Andor iStar 734, 1024x1024 pixels with the effective pixel size 13x13 µm). The ICCD was operated at 120x gain and 1x1 pixel binning. Due to a lower light sensitivity of the camera above the 600 nm the range of a complete detected spectrum was 190 – 600 nm. The timing of the experiment was optimized for each case study individually. The optimization process was aimed to obtain the best value of SNR. Representative number of laser pulses per sample was chosen for statistical lowering of possible differences among the spectra caused by the sample inhomogeneity. The values of laser pulse irradiance, temporal gating, the length of exposition, and number of spectra per sample (such as number of accumulations per spectrum) are listed in table 1. The whole LIBS setup was triggered by a delay generator (DG535, Stanford

Research Systems). Samples were placed on a motorized translational stage to provide a fresh spot for every laser pulse.

Table 1, experimental setting of the LIBS system for individual case studies.

Case study	Sample pretreatment	Pulse energy (mJ)	Irradiance (GW.cm^{-2})	Gate delay (µs)	Gate width (µs)	no. of accum.	no. of spectra	analysis
I	DST	100	33	0.5	10	10	20	univariate calibration, normalization
	pellet	150	51	1.5	10	20	10	
	slurry	150	51	1	10	20	10	
II	DST	35	11	0.4	5	10	20	univar. calib. PCA, SOM, PCR, PLSR
III	DST	35	11	0.4	5	10	20	univar. calib. SIMCA

The echelle spectrometer utilized for spectral analysis consisted of two wings with partly overlapping spectral regions. Obviously, it is advantageous to observe the spectra with broader spectral range. For this reason, continuous light emission from reference sources (tungsten and deuterium lamps) was observed and the correction function was estimated, figure 19. The correction factor provides the possibility to analyse any sample using both wings of the spectrometer. Spectra observed by both wings of the spectrometer can be combined when the spectra multiplied by the corresponding correction function.

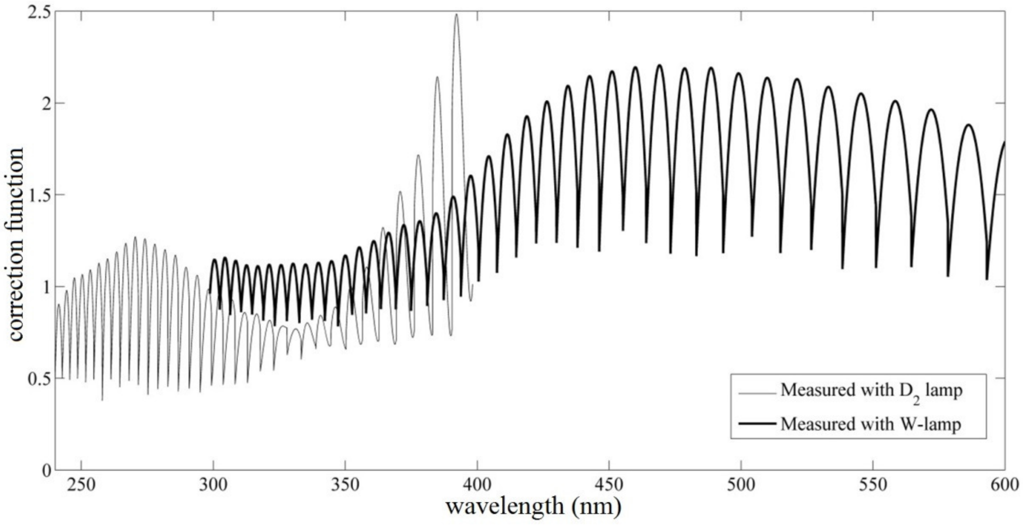


Figure 19, correction function estimated from the detected spectra of reference light sources.

SPECTRAL LINES SELECTION AND DATA PREPROCESSING

As it is explained further in the text, the igneous rocks can be completely discriminated by observing the differences in the composition of matrix elements (Al, Ca, Na, K, and Si) [107]. This means that the contribution of matrix elements is supposed to be the most influential in the classification process. To test this hypothesis, the PCA algorithm was applied on the data set of Oreas samples measured in the case study II and complete spectra were analysed, *i.e.* without any inclination to the matrix elements. Loadings of the first principal component which describes the most amount of variation among data are depicted in figure 20. This figure shows only a part of the whole observable spectral range, though the most eventful one. In general it is valid, the higher the loading value of a spectral line the more influential the spectral line for rock type discrimination. In the figure the highest peaks which belong to the matrix elements Al, Ca, and Si are highlighted. Mg lines were omitted while they are not considered to be a mean for general classification of rocks (QAPF diagram) [107]. Though, the contribution of Mg seems to be the most influential in this particular case.

Spectral lines of matrix elements (Al, Ca, Na, and Si), listed in table 2, were selected for further analysis. However, no potassium, as one of the proposed matrix elements, was observed in the LIP spectra due to the poor sensitivity of the ICCD camera over 600 nm. During the preprocessing, the spectra of individual samples were firstly treated with the Z-test to discard any possible outliers resulting from the sample inhomogeneity. Afterwards, spectra were normalized to the sum of overall intensities and then averaged. Selected spectral lines were fitted with pseudo-Voigt profile and their intensities were calculated as the area under the peak with background subtraction using custom Matlab (version R2012a) software.

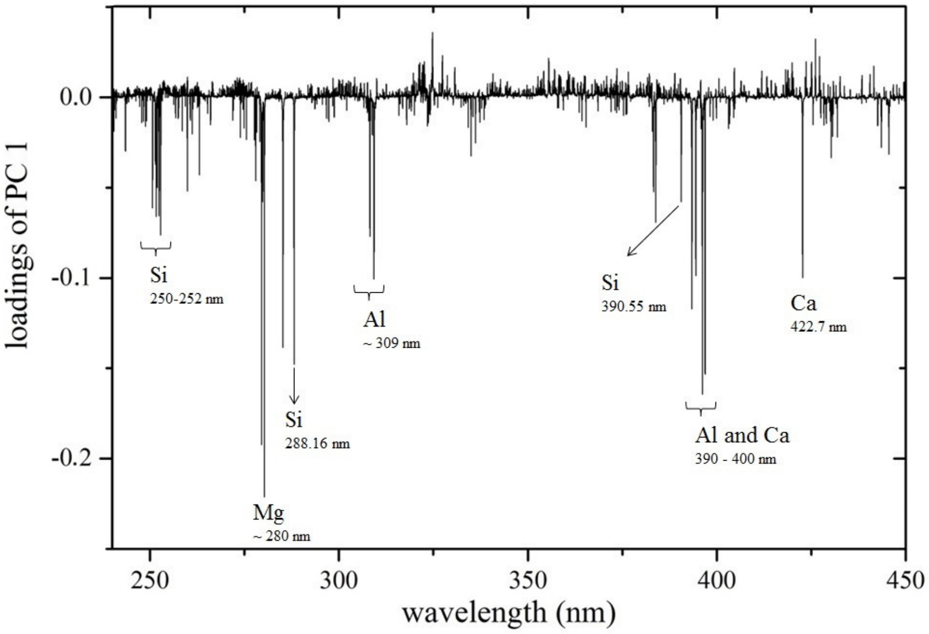


Figure 20, loadings of the first PC of Oreas without the selection of any regions of interest.

Obtained data were organized to a data matrix in the way that rows refer to the samples/measurements and selected spectral lines/variables refer to columns. The variation of the data in each column was normalized to unity and then mean-centered. The whole data matrix was analysed with chemometric algorithms to reveal possible latent variables among the data set and to provide the classification of the samples. The statistical toolbox in Matlab (PCA, SIMCA, PCR, and PLSR) was applied on the data matrix. Kohonen maps were created utilizing Somtoolbox in Matlab (Helsinki University of Technology, Finland) [108], free software for non-commercial use accessible online. The data analysis was partly focused on the uni- and multivariate calibration of the LIBS system. LIBS as an *in-situ* detector capable of fast and reliable quantitative analysis is of increasing interest as well in mining industry. In our case the main emphasis is given to the quantitative analysis of copper in the soil samples. Copper lines selected for uni- and multivariate analysis are listed in table 3.

Table 2, list of elemental lines selected for analysis.

Elemental line	**λ** nm	**E_i** eV	**E_j** eV	**Einstein coeff.** 10^8 s^{-1}
Al (I)	308.22	0	4.02	0.63
Al (I)	309.27	0.014	4.02	0.75
Al (I)	394.4	0	3.14	0.51
Al (I)	396.12	0.014	3.14	1.01
Ca (II)	393.37	0	3.15	1.47
Ca (II)	396.85	0	3.12	1.44
Ca (I)	422.7	0	2.93	2.18
Ca (I)	428.3	1.89	4.78	0.43
Ca (I)	558.87	2.53	4.74	0.49
Na (I)	589	0	2.104	0.62
Na (I)	589.59	0	2.102	0.61
Si (I)	251.43	0	4.93	0.74
Si (I)	251.61	0.03	4.95	1.68
Si (I)	251.92	0.01	4.93	0.55
Si (I)	288.16	0.78	5.08	2.17
Si (I)	390.55	1.91	5.08	0.13

Table 3, elemental lines of copper used for the formation of calibration curves.

Elemental line	**λ** nm	**E_i** cm^{-1}	**E_j** cm^{-1}	**Einstein coeff.** 10^8 s^{-1}
Cu (I)	324.75	0	3.82	1.37
Cu (I)	327.39	0	3.79	1.36
Cu (I)	515.32	3.79	6.19	0.6
Cu (I)	521.82	3.82	6.19	1.22

SAMPLE SETS

One of the main goals of this thesis is to validate the performance of chemometric algorithms for advanced classification and quantitative analysis of soil samples measured by a LIBS device. Chemometric algorithms were applied in various case studies and the results were compared and thoroughly explained.

Case study I: Cu assays of the monzonite rock type (obtained from University of Clausthal) were analysed in the form of pressed pellets, as slurries (mixture of fine dust with water) and as fine dust pressed to the surface of double sided adhesive tape (DST). The approaches were compared to each other in order to select the most convenient and non-demanding way of the sample pretreatment. The univariate calibration curves were depicted and further normalized utilizing the intensity of matrix elements lines to suppress the matrix effect, *i.e.* to improve their linearity.

Case study II: Certified reference samples Oreas (ORE[3], Australia) were analysed with an X-ray Fluorescence (XRF) device and LIBS. Then basic chemometric algorithms were applied, for classification and the detection of outliers (PCA) and for quantitative multivariate analysis (PCR, PLSR). Moreover, the PCA analysis was then emulated with Kohonen maps. The performance of selected chemometric algorithms was compared to each other while the positives and negatives of each algorithm were discussed in detail. Algorithms for sufficient and successful analysis of soil samples were proposed.

This research was extended by the analysis of 27 igneous rocks. Those results were submitted to the special issue of Spectrochimica Acta B and the article is attached to this thesis, see appendix.

Case study III: Robust data set consisting of 52 andesite and 28 diorite samples collected at a copper mine in north-west region in Iran (provided by University of Clausthal) were measured and classified using SIMCA. No ‘test’ data set was at hand and therefore the leave-one-sample-out (LOSO) approach was adapted to the analysis.

Those results were presented at the thematic conference EMSLIBS 2013 in Bari, Italy.

ROCK TYPE CLASSIFICATION

The observed LIBS spectrum reflects the chemical composition (*i.e.* the chemical fingerprint) of the sample. The composition of mineral ores differs with their locations and as well with the way of their alterations [107, 109, 110]. Different mineral ores and their alterations can be discriminated by examining their chemical fingerprints and the proportions of matrix elements (Al, Ca, Na, K, and Si). Mineral ores can be classified according to their content into several groups, forming a QAPF (quartz, alkali feldspar, plagioclase, feldspathoid) diagram (figure 21). The borders between these groups reflect natural relationships; each rock type displays a certain continuous variation of a mineral content and therefore the borders cannot be considered strictly as a line dividing individual rock types. The exact classification can thus not be done by knowledge of the concentration of only one or two characteristic elements but has to be derived from the complete elemental fingerprint. An established method for classification of such complex and large datasets is the principal components analysis (PCA). It is expectable that the PCA discrimination of the sample set represented by their LIBS spectra can in some way emulate the distribution of the rock types in QAPF diagram.

An alternative visualization of rocks classification to QAPF diagram is the Total Alkali Silica (TAS) diagram, figure 22; while QAPF classification of volcanic rocks remains the primary possibility [111]. The classification of rocks employing TAS diagram is purely descriptive and should be used for volcanic rocks only; furthermore TAS is independent of field location. In TAS diagram the

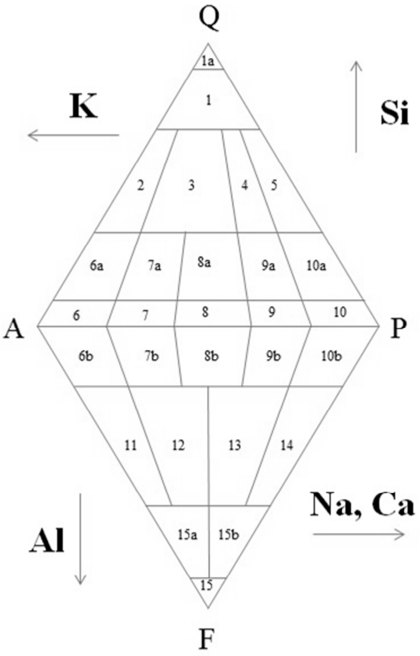
---

[3] ORE; Ore Research and Exploration Pty Ltd, Melbourne, Australia is a leading producer of certified reference materials for the mining, exploration and analytical industries [116].

concentrations of SiO_2 versus Na_2O and K_2O are plotted to discriminate the rocks of interest. As in the case of QAPF diagram, only the matrix elements are used for the necessary rock type classification. Despite of the strict rule for utilizing of the TAS diagram to classify the volcanic rocks, TAS was utilized for the schematic classification of sedimentary rocks, the certified reference materials used in this preliminary study.

The composition of mineral ore differs with the location of its occurrence/provenance as well as with the way of its alteration. Alteration is the property of a rock which explains its chemical and mineralogical changes in the course of time. In geology, the alteration is important because it may have an effect on grades of elements (*e.g.* copper), therefore the rocks are studied in individual alteration types. However, more detailed description of the rock type classification is beyond the scope of this thesis. Moreover, proposed discrimination tools are strictly restricted to igneous rocks (QAPF diagram), or to igneous volcanic rocks respectively (TAS diagram).

Plutonic Rocks	
1	Quartz-Rich Granitoid
1a	Quartzolite
2	Alkali-Feldspar Granite
3	Granite
4	Granodiorite
5	Tonalite
6	Alkali-Feldspar Syenite
6a	Quartz-Bearing Alkali-Feldspar Syenite
6b	Foid-Bearing Alkali-Feldspar Syenite
7	Syenite
7a	Quartz-Bearing Syenite
7b	Foid-Bearing Syenite
8	Monzonite
8a	Quartz-Monzonite
8b	Foid-Monzonite
9	Monzodiorite / Monzogabbro
9a	Quartz-Monzodiorite / Quartz-Monzogabbro
9b	Foid-Monzodiorite / Foid-Monzogabbro
10	Diorite / Gabbro / Anorthosite
10a	Quartz-Diorite / Quartz-Gabbro
10b	Foid-Bearing Diorite / Foid-Bearing Gabbro
11	Foid-Syenite (Foyaite)
12	Foid-Monzosyenite
13	Foid-Monzodiorite (Essexite)
14	Foid-Diorite
15	Foidolite
15a	-
15b	-



Volcanic Rocks	
1	-
1a	-
2	Alkali-Feldspar Rhyolite
3	Rhyolite
4	Dacite
5	Plagidacite
6	Alkali-Feldspar Trachyte
6a	Quartz-Bearing Alkali-Feldspar Trachyte
6b	Foid-Bearing Alkali-Feldspar Trachyte
7	Trachyte
7a	Quartz-Trachyte
7b	Foid-Bearing Trachyte
8	Latite
8a	Quartz-Latite
8b	Foid-Bearing Latite
9	Andesite
9a	Quartz-Bearing Andesite
9b	Foid-Bearing Andesite
10	Basalt
10a	Quartz-Bearing Basalt
10b	Foid-Bearing Basalt
11	Phonolite
12	Tephritic Phonolite
13	Phonolitic Tephrite / Phonolitic Basanite
14	Tephrite / Basanite
15	Foidite
15a	Phonolitic Foidite
15b	Tephritic Foidite

Figure 21, QAPF diagram **[107]**.

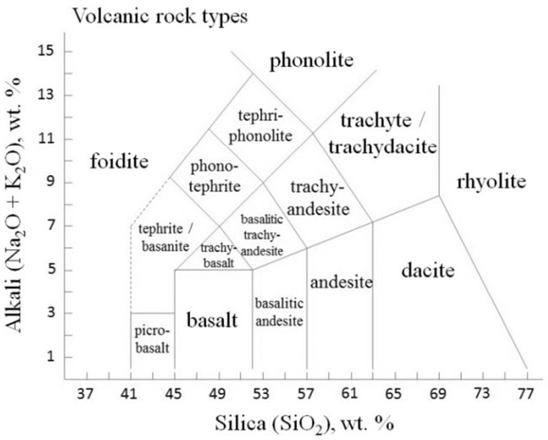


Figure 22, TAS diagram **[111]**.

4. RESULTS

4.1.CASE STUDY I

In this case study the main emphasis is given to the sample preparation prior to a LIBS measurement and consequent univariate quantitative analysis. For those purposes 10 samples of the same igneous rock type, monzonite, in the same state of alteration, potassic, were selected. Any monzonite rock is composed from similar amount of alkali-feldspar and plagioclase, figure 21. Its composition consists mainly from Al, Ca and alkali elements and contains less than 5% of Si. Therefore, Al and Ca are considered to be the main matrix elements. The samples for this case study were collected in the north-west region of Iran, near Sungun village, where are several copper mines. Samples were delivered in the form of homogenized fine dust. The Cu content, table 4, was estimated by ICP-MS after four-acid digestion. This analysis was done at the University in Clausthal, Germany. The rock type and alteration of each sample were confirmed by an experienced geologist in the same place.

Table 4, copper content in selected monzonite samples.

Sample	Cu (wt. %)	Sample	Cu (wt. %)
Cu1	0.2	**Cu6**	0.66
Cu2	0.26	**Cu7**	0.74
Cu3	0.39	**Cu8**	0.86
Cu4	0.44	**Cu9**	0.96
Cu5	0.53	**Cu10**	1.38

It is noteworthy that flashlamp pumped laser Continuum SLI-10 was employed to create a LIP. In the other two case studies more compact (in the sense of maximum generated pulse energy) DPSS laser was used. Thermodynamic properties of laser-induced plasmas such as temporal and spatial evolution are significantly different when generated by laser pulses of various energies. Though analysis applied on such diverse LIP spectra is comparable to some extent. Nevertheless, the reproducibility of LIBS measurements is proved in the article submitted to Spectrochimica Acta Part B, see appendix, where the laser Continuum SLI-10 and DPSS laser (Quantel Ultra 100) were utilized to produce a LIP.

Samples in the form of homogenized fine dust were prepared in three different ways: 1) in the form of fine dust mixed with water (henceforth referred to as slurry), 2) pressed to the surface of the double sided adhesive tape (DST), and 3) pressed to pellets. The approach of mixing sample with water was chosen to emulate the real shape of the sample which is measured on-site. The problem with mixing any sample (in the form of homogenized fine dust) with water results in the partial heterogeneity of a sample. Moreover, the ratio of water to soil dust affects the matrix of the sample and consequently the plasma properties [112]. Samples were prepared with the ratio of one part of water to two parts of soil dust. In another approach, samples were as soil dust pressed to the surface of the DST [113]. The sample was evenly deposited on the surface of a DST. Loose and redundant sample was then blown away from the surface of DST. The DST was measured alone to ensure that no disturptive spectral lines interfering with lines of interest can be detected. Finally in the case of pressed pellets, homogenized fine dust was mixed with micropowder Hoechst wax C ($C_{38}H_{76}N_2O_2$) in the ratio 2:1. The micropowder is conventionally used as tabletting aid for XRF analysis. The mixture was then pressed in hydraulic press with 20 tons for 30 seconds.

The calibration curves were constructed for each way of sample preparation, figure 23. The copper content in the samples varies from 0.2 to 1.38 wt. %. In this range of concentrations, selected copper lines are not suspected from self-absorption. Therefore, resonant atomic lines of copper observed at 324.75 and 327.39 nm may be used along with Cu (I) at 515.32 and 521.82 nm for the quantitative analysis. Measured spectra were averaged and lines fitted with pseudo-Voigt profile. Depicted calibration curves for both selected copper lines (figure 23a) 324.75 and b) 521.82 nm) have very poor coefficients of determination, R^2. Such calibration curves are naturally not satisfactory for any further use. This poor fit of the linear regression may be caused by a matrix effect. Though, the samples were selected from only one group in order to avoid the matrix effect affecting the intensity of the measurement. The trends of individual calibration curves (for DST, pellet and slurry approach) are similar and suggest similar dependency of the copper lines intensities on the matrix of relevant sample.

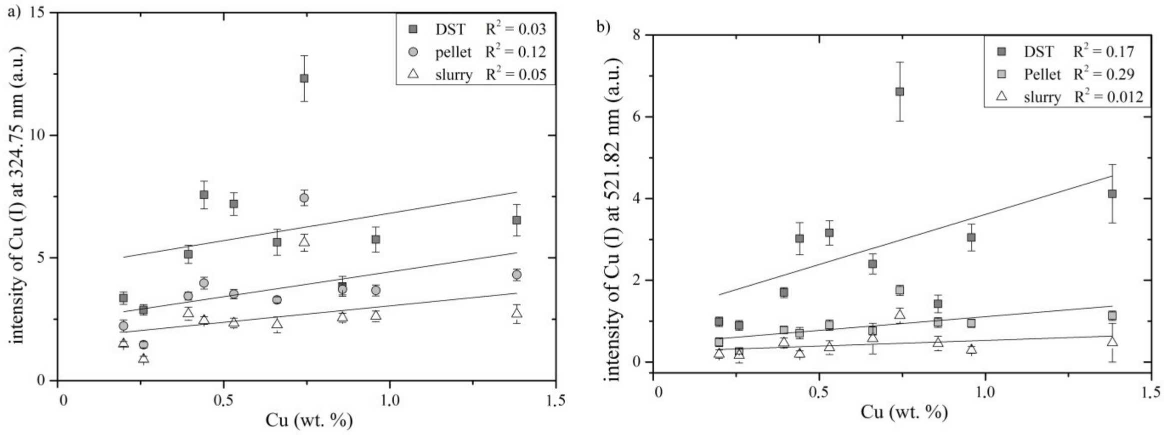


Figure 23, calibration curve for Cu (I) line at a) 324.75 nm and b) 521.82 nm.

The intensities of matrix lines Ca (I) at 422.7 nm and Al (I) at 309.27 nm were cross-plotted to reveal possible outliers in the data set, figure 24. It can be observed that two samples (Cu4 and Cu7) strongly deviate from the rest of the data set. Those samples were assigned as outliers and were discarded from further analysis. The simple cross-plot was chosen because only two matrix elements (Al and Ca) were assigned as dominant. It has to be kept on mind that the discrimination was based on the QAPF diagram, *i.e.* other spectral lines of other elements (Na, K, and Si) should be utilized in the normalization process.

The discrimination based on QAPF diagram allows assigning two samples with diverse content of matrix elements into one group. This diversity may vastly affect the LIBS measurement, *i.e.* the bigger the diversity in a sense of the matrix elements content the more influential the matrix effect. This hypothesis fits to our experimental results where the samples, Cu4 and Cu7, are strong outliers in respect to the main cluster of samples, figure 25. Consequently, the matrices of those outlying samples appear to influence the intensity of the copper line in different manner, *i.e.* the intensity of copper line is significantly higher than it should be expected in the context of other samples, figure 24. The intensity of Al line is at least two times higher and the intensity of Ca line is two times lower for samples Cu4 and Cu7 in respect to the rest of the group. The degree to which the matrix effect affects the intensity of trace elements line may be an objective for further study.

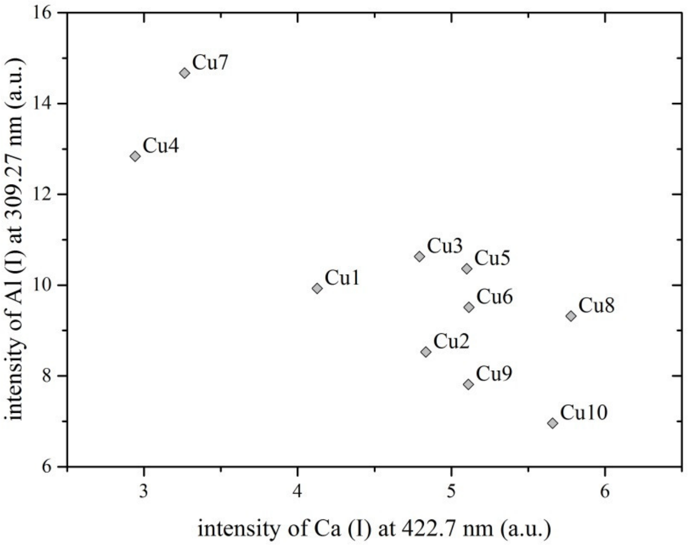


Figure 24, cross-plot of matrix elements Ca and Al to reveal possible outliers.

The calibration curves were normalized in order to supress the matrix effect. Figure 25, for those purposes spectral lines of Al and Ca were selected, a) Al (I) at 396.12 nm and b) Ca (I) at 422.7 nm. The calibration curves for Cu (I) lines at 324.75 nm and at 521.82 nm (results considering this line are not shown) were normalized. Moreover, normalization of the calibration curve did not lead to any significant improvement in the sense of linearity (R^2) when elemental lines of Na and Si were utilized, results not shown. The normalization utilizing the aluminium line improved the figures of merit (R^2 is above 0.84) more significantly than the calcium line (moderate value of R^2 up to 0.62). This is proved by the comparison of coefficients of determination in both plots. However, the normalization of the whole data set to aluminium line, figure 26 (plotted only for the DST measurement), is not sufficient while the two outliers still devaluate the linearity of resulted calibration curve ($R^2 = 0.71$) compared to the one without outliers ($R^2 = 0.96$). The regression line of the calibration curve containing the outliers is depicted as dotted line. Therefore, normalization of the calibration curve by the spectral line of a matrix element should be done with care. This normalization, based on presented results, is not able to improve the figures of merit when samples with more diverse composition are analysed. Discrimination of the data set to smaller parts with similar composition of matrix elements may be beneficial. It is advisable to believe that multivariate analysis may be of help when samples with more complex matrices are analysed at once. In follow up case studies the multivariate algorithms are used for classification, outlier detection and quantification.

In this case study, different approaches of sample preparations were examined. The calibration curve obtained from the measurement with samples spread on the surface of DST showed the steepest trend, figure 25a, and therefore may be considered as the most sensitive one. Though, the measurements for individual approaches were utilized under various conditions, table 1. The approach with the deposition of the sample on the surface of DST was selected for further analysis for its fast and non-demanding preparation. The preparation of pellets was abandoned for its difficult applicability to *in-situ* analysis. Both approaches showed good repeatability in the sense of low RSD. In a real application of mining the sample is obtained from the drill in the form of mixture with water. The amount of water is not stable which leads to a non-repetitive measurement. As it was already mentioned [112], the amount of water affects the matrix of the sample and consequently the intensity of measured spectral lines. Therefore, the amount of water has to be monitored properly to avoid unwanted matrix effects resulting from the water to sample ratio. Nevertheless, there is no limitation in drying the sample and depositing the sample on the surface of a DST.

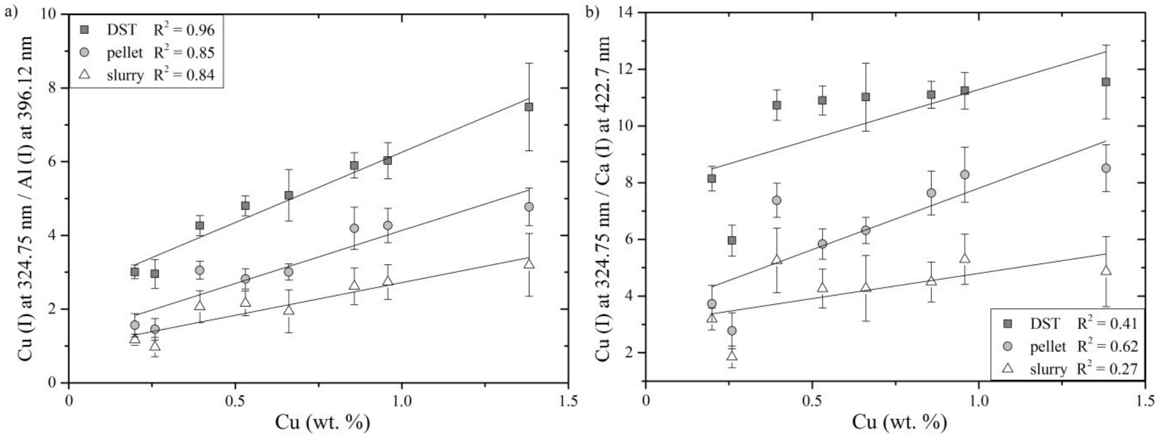


Figure 25, calibration curves for copper normalized to a) Al (I) at 369.12 nm and b) Ca (I) at 422.7 nm.

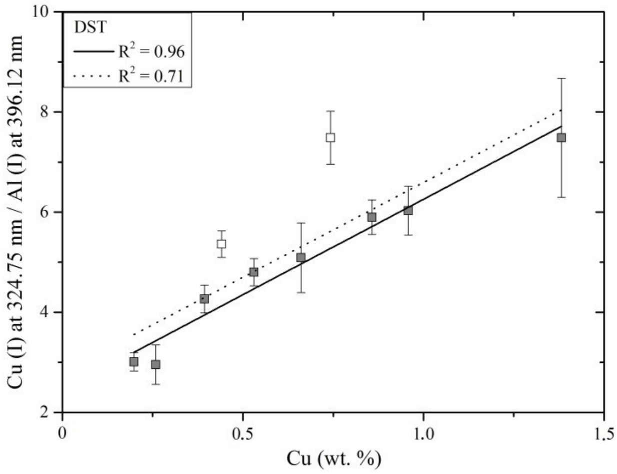


Figure 26, calibration curve containing outliers Cu4 and Cu7 (dotted line) normalized to Al line.

4.2. CASE STUDY II

Certified reference materials were chosen to test the constructed LIBS setup, its performance, and especially the proposed algorithms for multivariate data analysis. 18 certified samples (ORE, Australia) were selected for their wide range of copper amount, while the quantitative analysis of copper was one of the main concerns in the mining project, and the diversity in the composition of matrix elements. Samples are divided into three groups according to their matrices which vary not only in the composition of formerly proposed matrix elements (Ca, Al, Si, K, and Na). Consequently, the LIBS measurement is affected by the matrix effect as it was proved in the foregoing case study. The datasheet of ore standards provided by producer, table 5, is not satisfactory to construct the TAS diagram. For this reason, whole sample set was measured with handheld XRF device (NITON XL3t series manufactured by Thermo Fisher Scientific, USA) in the lab at BAM, table 6. The comparison of both tables, the first provided by a producer and the second measured with XRF device, showed their relatively good compliance. In the case of LIBS measurement, the system (laser energy, temporal gating) was set according to the data listed in table 1. Samples were measured in the form of fine dust spread into a thin film on the surface of DST.

Table 5, list of certified reference materials, OREAS ore standards.

Item no.	matrix	Cu / ppm	Fe / ppm	Al_2O_3 / ppm	CaO / ppm	SiO_2 / ppm
O91	siltstone	265	43400	147000	7800	675000
O92	siltstone	2294	46450	-	-	-
O95	mineralised siltstone	25900	8960	133000	4000	587000
O96	mineralised siltstone	39300	10010	127000	4600	558000
O97	mineralised siltstone	63100	121800	115000	3600	508000
O98	mineralised siltstone	148000	196000	75300	2600	334000
O110	massive sulphide	1620	251000	-	-	-
O111	massive sulphide	23700	352000	-	-	-
O111b	massive sulphide	24700	361000	-	-	-
O112	massive sulphide	51000	341000	-	-	-
O113	massive sulphide	135000	282000	-	-	-
O160	shale/siliceous dolomite	13	11500	50400	1460	887000
O161	shale/siliceous dolomite	4090	42600	24900	1580	842000
O162	siliceous dolomite	7720	85700	17000	132000	404000
O163	siliceous dolomite	17600	110700	32400	8600	634000
O164	siliceous dolomite	22500	68000	20700	3930	779000
O165	siliceous dolomite	32000	88600	25900	810	720000
O166	siliceous dolomite	88200	113800	13800	9800	614000

The data set under investigation is described by variables which may be correlated in some way. While comparing the variables in respect to each other the possible latent correlations are revealed. Data set can be divided into various parts, *i.e.* data points are clustered together. Most straight forward way how to examine the data set is to select several most representative variables and cross-plot them against each other. This is done for instance in figure 27 where a TAS diagram for selected certified ores is depicted by plotting the contributions of sodium oxide and potassium oxide against silicon dioxide. This approach is well evaluated in geochemistry for the classification of igneous rocks. It is noteworthy, that in this case study the TAS diagram is plotted from only silicon dioxide and potassium oxide obtained utilizing a handheld XRF device. Moreover, this TAS diagram serves only for demonstrative purposes and should be accepted with care. Based on the TAS diagram, depicted in figure 27, the data matrix is disintegrated into several distinct fragments. Group outliers are visible, namely samples Oreas110 and Oreas160, and should be discarded from the analysis. Moreover, sample Oreas110 can be considered as the member of the dolomite group. The elemental composition of the members of the siltstone and dolomite groups varies strongly especially in the amount of silicon.

Table 6, list of OREAS certified reference materials measured with hand-held XRF device.

Item no.	Cu / ppm	Fe / ppm	Al / ppm	Ca / ppm	K / ppm	Si / ppm
O91	280	42810	74810	5320	26410	292720
O92	2620	45820	78420	4810	27420	289860
O95	26830	87170	70580	2670	19530	237630
O96	40960	99480	77740	3290	18780	227930
O97	66610	121300	72310	2580	16910	193750
O98	150710	185870	55990	1890	10750	112810
O110	1740	250480	7970	19030	1000	258330
O111	24490	324490	-	1200	1140	72190
O111b	23050	317270	11370	1950	900	75170
O112	50580	305630	14100	1990	1100	75660
O113	132380	253770	24830	4000	1160	92690
O160	-	11370	26160	1040	15350	437570
O161	4550	42260	13270	1080	280	397600
O162	7950	82890	7630	101730	1810	196100
O163	18130	105300	20120	6230	690	291650
O164	24010	66220	11080	2850	320	362580
O165	33820	85090	16640	560	1520	326750
O166	90660	109580	9530	6910	2240	279550

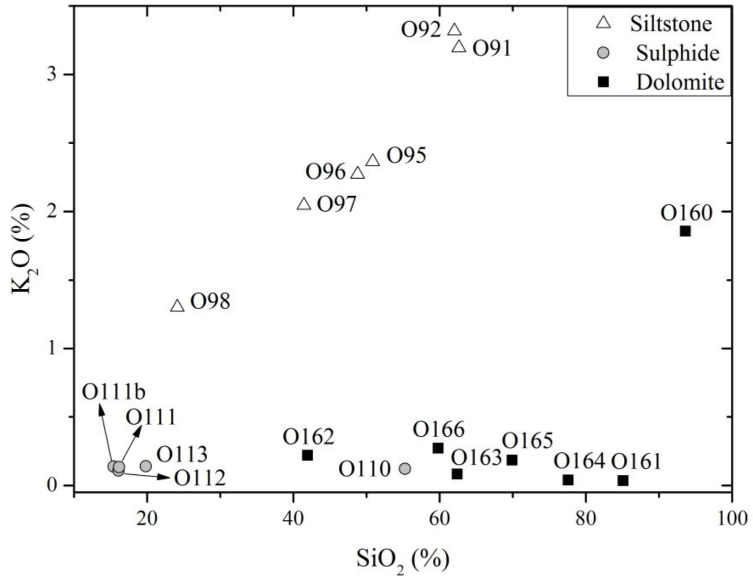


Figure 27, TAS diagram of OREAS standards.

CLASSIFICATION

The cross-plot approach is more difficult to be interpreted when more uncorrelated variables are used to describe the system. For those purposes the PCA algorithm is employed in the data analysis. PCA emulates the topographical distribution of the samples in the data set while only important and most distinctive information is used and the redundant information is truncated. PCA rotates the coordinate system and projects the data points in the least square sense. Firstly, PCA was applied on the data set obtained from the XRF measurement. This data set consists only from the Al, Ca, and Si contributions, while XRF device is not capable of detecting light elements such as Na. The intention was to compare the performance of LIBS and XRF by PCA. Therefore, potassium was omitted from the analysis while no spectral lines of potassium were detected utilizing LIBS.

PCA was applied on the data matrix, which was composed from the XRF, respectively LIBS, measurements. Spectra obtained from LIBS measurement were averaged and organized into the data matrix as rows. Selected variables/lines of matrix elements were fitted with pseudo-Voigt profile and assigned to columns, which were then normalized to unity and mean-centered. PCA scores of XRF data, depicted on the figure 28a, show three distinct groups and one strong outlier, Oreas 160. The score plots are depicted as 2D projections of a multidimensional space, where first two principal components cover the most of the variation among the data, 97.8% (65.2% and 32.6%). This result is similar to the topological distribution of the data points in the TAS diagram. However, contributions from different elements were utilized instead of the contribution of potassium. It may be suggested that silicon is the most influential element in the classification process. Furthermore, the dependency of the discrimination on the individual elements is described by a loadings plot constructed by plotting the loadings of related PC against each other. Investigating the loadings depicted on figure 28b the contributions of Si and Ca are dominant in the first principal component. While in the second principal component the most significant element is Al. This implies that the variation along the first PC in the scores plot, figure 28a, is closely related to the composition of the sample in the sense of Ca and Si content. Similarly this is valid for the second PC in the scores plot and the amount of Al in the sample. However, it should be repeated that individual principal components have in general no direct physical meaning, *i.e.* one PC cannot be connected to only one variable in multivariate case. Each principal component describes variation among the samples where the most variation within the data set is assigned to the first principal component, in this case 65.2%. It is obvious from the scores plot that the data tend to cluster according to rock types, i.e. according to the composition of matrix elements. Classification is possible based on the results of PCA applied to the spectra. In praxis, the samples are usually classified on-site by a geologist based on their physical properties which may not be reliable and correct for all of the samples. As an initial guess, one would not have expected such clearly separated groups because there are no sharp natural boundaries between classes of rock types, *i.e.* the amounts of matrix elements vary continuously.

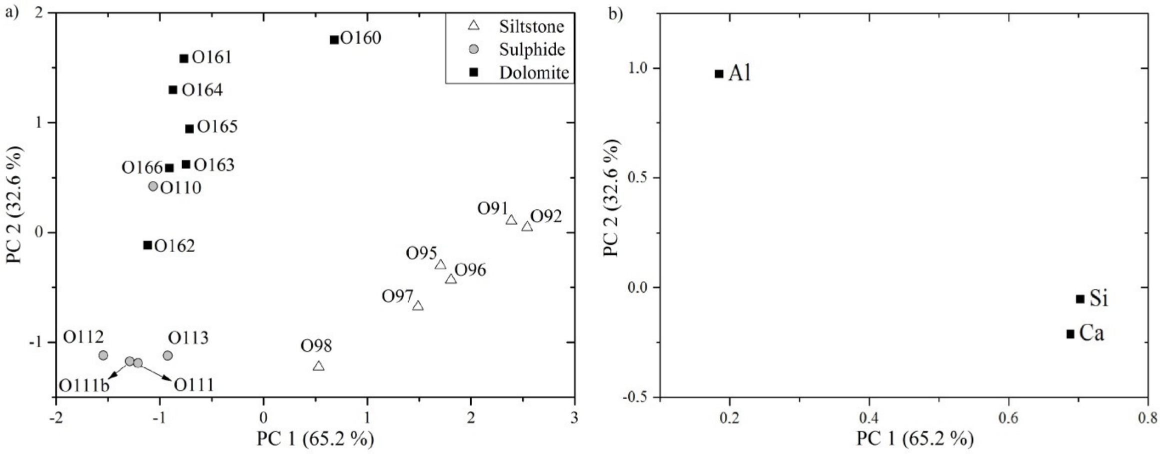


Figure 28, PCA of XRF data, a) scores, b) loadings.

PCA was then applied to the data matrix of LIBS measurement of Oreas samples, where spectra of each sample were averaged and then organized to rows of this data matrix. The data matrix consisted of 14 variables listed in table 2. The variables were represented by integral of the area under the peak of selected spectral lines (Al, Ca, and Si). Only those elements were picked up for the analysis in order to compare the results with the PCA analysis of XRF data. First three principal components covering 96.5% of total variation (60.5%, 30%, and 6%) were cross-plotted to investigate the classification of the samples, figure 29. Expected grouping is over-shadowed by extremely outlying data points, Oreas160 and Oreas162). PCA algorithm is essentially sensitive to outliers due to its least squares property. For this reason, it is beneficial to perform the PCA in the first step of the analysis only to detect outliers. Moreover, it may be concluded that LIBS measurement is significantly affected by the matrix effect.

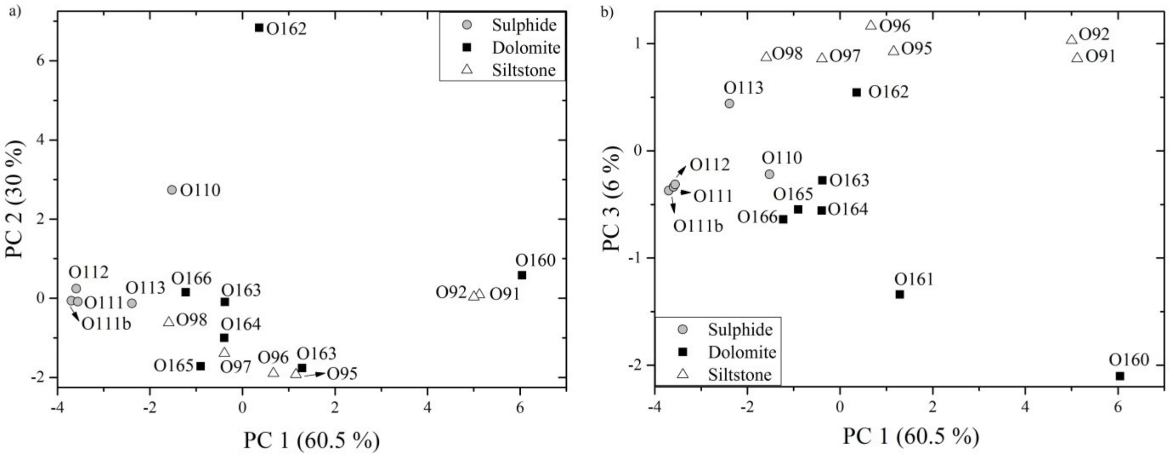


Figure 29, PCA scores of LIBS data, a) 1st and 2nd principal component, b) 1st and 3rd principal component.

The influence of individual matrix lines to the discrimination of standardized samples is schematically depicted in loadings plots of the first three PCs, figure 30. The main discrimination power may be assigned to Al and Si and secondary to Ca, which is supreme in the second PC. It was mentioned in the section 1.7, that the scaling of a variable with a lower concentration range results in higher computational error. This may be the case of Ca, however further study should be done. The loadings plots constructed from PCA of LIBS and XRF do not correspond. This may be a consequence of outliers disturbing the estimation of PC in the case of LIBS data.

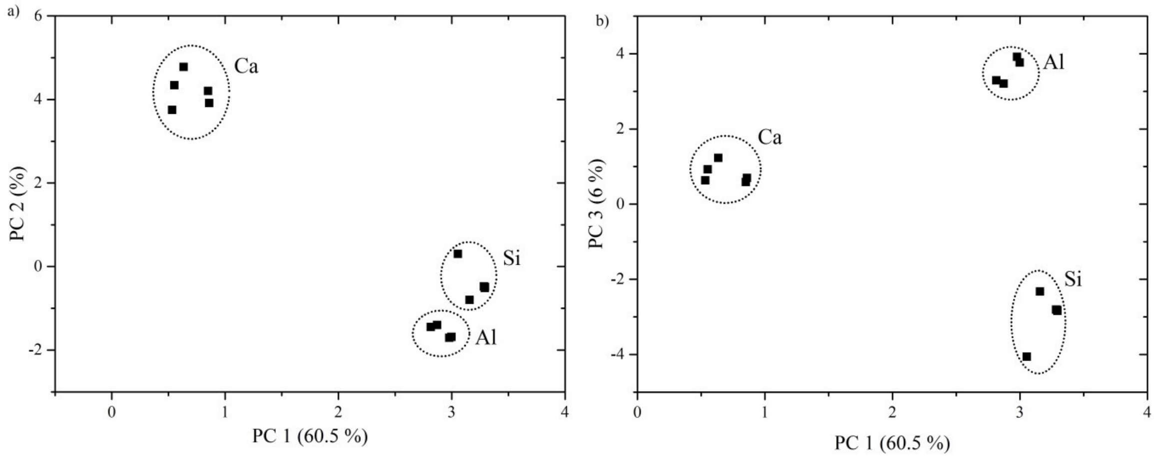


Figure 30, PCA loadings of LIBS data, a) 1st and 2nd principal component, b) 1st and 3rd principal component.

It is worth of a repetition that the outliers severely affect the estimation of PCA model. Therefore they should be discarded from the further analysis (classification and quantification). The results of classification and quantification of a data set without any outliers are shown at the end of this case study where classification is beneficially connected with quantitative analysis, figure 39.

It has been already mentioned that Kohonen maps are considered to be an alternative to PCA analysis, therefore, both algorithms were compared. Kohonen maps were trained utilizing unsupervised learning to create a two-dimensional representation of the input data set. During the training the neurons adapt their properties and preserve the topological distribution of the information in the data set. The training is based on the competitive and cooperative learning when the winner (the closest matching neuron to the selected data point) changes its properties according to the selected data point. Moreover, neurons surrounding the winner are as well affected by the change up to an extent given by neighbourhood function. Neurons are connected with their closest neighbours and form a net which is stretching and shrinking over the data space during the computation.

In the first step of the analysis the number of neurons and the total number of iterations should be determined. In our case, total of 100 neurons and 1500 iterations were selected. The number of neurons and iterations were selected in order to minimize the quantization error [93]. The size of neighbourhood function and the measure of adaptation can be as well defined. Both magnitudes decrease their values with every step of iteration while in the end the algorithm has only competitive character and winning neurons do not affect their neighbours. When the Kohonen map is trained two basic means of its visualization are usually utilized, the unified distance matrix and the component planes.

For I number of samples and L number of neurons there will be I winning neurons directly representing samples in the constructed Kohonen map. Samples are scattered according to the information provided by the variables, the data matrix was constructed in the same way as in the case of PCA. The winner changes its properties after each iteration, so as the properties of the closest neighbours and shifts its position to preserve the topological distribution of the samples in the original data matrix. When the training is finished the winners adapted their position and the property of each neuron is then visualised in a series of 2D component planes, figure 31. Trends in the data set can be observed while examining three component planes of Al, Ca, and Si (XRF data). Neurons organize their positions to create clusters of similar samples, when the similarity is based on the elemental composition. The component planes are alternative of PCA loadings and show how individual variables affect the distribution of the samples. Investigating figure 28b and figure 31, the contribution of individual elements is comparable.

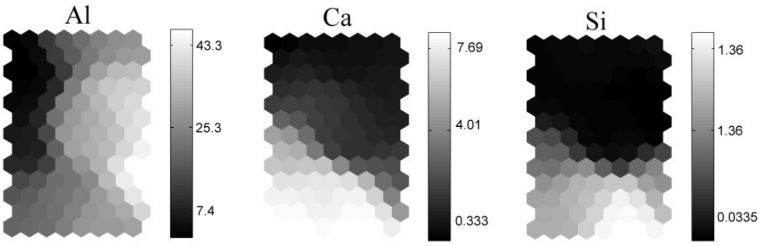


Figure 31, XRF measurement of Oreas samples: component planes

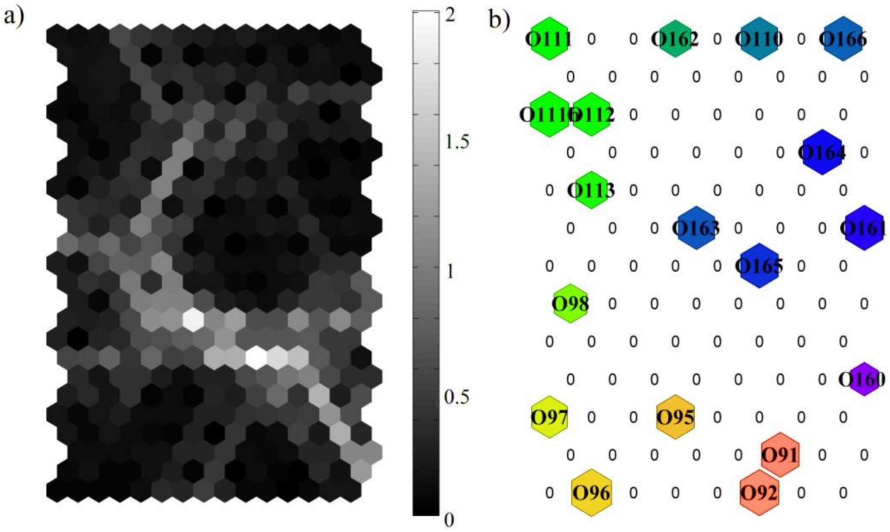


Figure 32, XRF measurement of Oreas samples: a) the unified distance matrix, b) samples assigned to neurons.

The visual comparison of the component planes gets more difficult when a higher number of variables is introduced into the computation. For this reason, the unified distance matrix is evaluated. The U-matrix is a representation of the component planes where the Euclidean distance between neighbouring vectors is visualized. When such distances are depicted in colour scale, the more white the segment of U-matrix the more distant the neighbouring neurons. As it is depicted in figure 32a, the newly constructed data space is divided into several segments where the colour of the segment stands for the distance of two consequent neurons. The whole map can be simplified into imaginations of valleys and mountains. It is therefore valid that the higher the mountains the more different the data points, neurons. The U-matrix can be seen as an alternative to PCA scores in a 2D projection. The exact positions of winning neurons representing the samples are visualized on figure 32b. The colour in this case signifies the similarity of the neighbouring samples and the size of each unit represents how frequently the corresponding neuron was assigned as a BMU during the training. Component planes and U-matrix should be, as well as PCA scores and loadings, examined in parallel to fully understand the newly constructed data space.

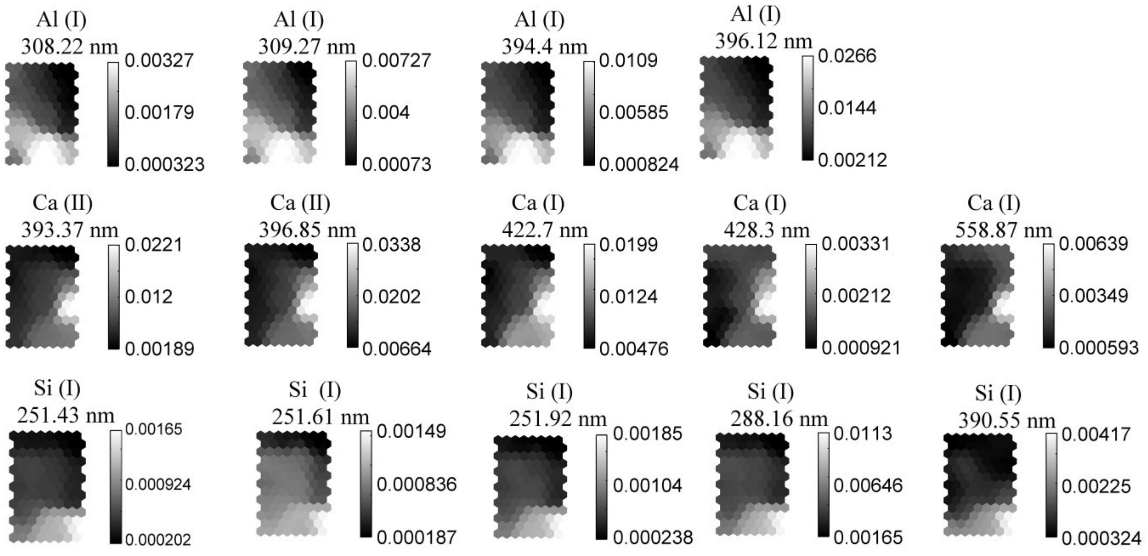


Figure 33, LIBS measurement of Oreas samples: component planes.

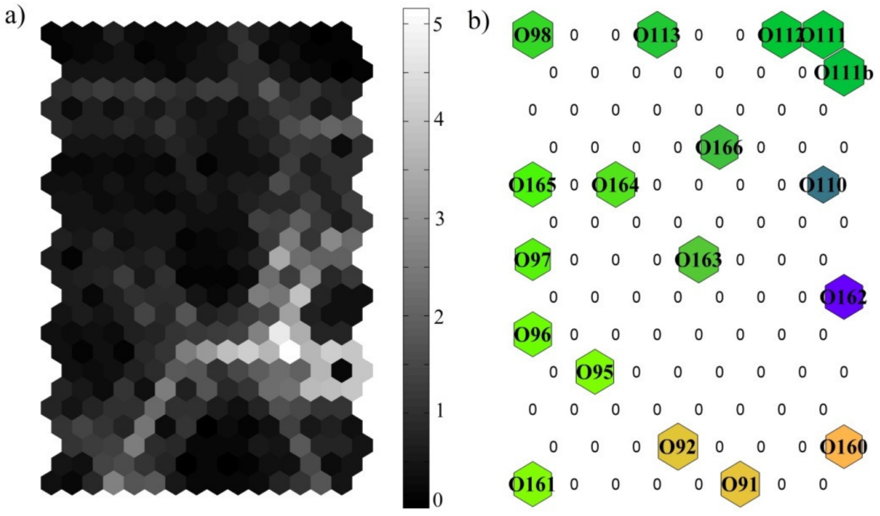


Figure 34, LIBS measurement of Oreas samples: a) unified distance matrix, b) samples assigned to neurons.

It was already stated that more spectral lines per elements were chosen in order to supervise spectral lines with each other. This is done to prevent the misuse of variable, while resonant lines may get saturated and the information they carry is therefore limited. On the other hand, less intense lines of an element may not be appropriate in the description of samples in which this element occurs in trace amount. Nevertheless, it is reasonable to expect the similar dependency of the model on various lines of the same element.

Kohonen maps were trained for LIBS data in another step of the analysis, where the data matrix was organized as in the previous case. Component planes of 14 variables, spectral lines of matrix elements (Al, Ca, and Si), are depicted on figure 33. The contributions of spectral lines of one element behave in similar manner as it was expected, *i.e.* individual component planes of one element follow similar trend in which the samples were organized in the map. The intensity of a spectral line is strongly dependent on the amount of the element in the sample. However, the total intensities of spectral lines of one element in a LIP differ according to their quantum properties. This is reflected in their contribution to the model. For instance the Al line at 309.27 nm is conventionally observed with significantly lower intensity compared to the Al line at 396.12 nm. This is evident from the component planes for both lines where the numerical range of the Al line at 309.27 nm is significantly lower than the one of the line at 396.12 nm. The component plane of the Al line at 309.27 nm appears to be shallower. This means that the Al line at 396.12 nm is more apt to reflect the variation among the samples.

Contributions from all of the component planes were then collected to form the U-matrix, figure 34a. This result is in a good agreement with the one from PCA algorithm. The outliers are evident from both parts of the figure, by a bigger distance discriminating the samples in the bottom right corner of the figure. The colouring assigns the majority of the samples into one cluster (green colour). This cluster is surrounded by strongly deviating outliers, namely O91, O92, O110, O160, and O162.

To conclude this section, the performances of PCA and Kohonen maps were compared. The main emphasis was given to the detection of outliers and sample clustering. In both cases, the outliers strongly influenced the resulted projection of the samples in the newly constructed coordination space. Clustering of the data set is possible based on similarities in spectral variables, *i.e.* matrix elements (Al, Ca, and Si). The data matrix can be in further steps disassembled to individual clusters, *i.e.* groups of samples with similar composition in matrix elements. For the very complex data with overlapping groups, this discrimination has to be done step by step, as suggested by Multari et al. [105]. In their study, the most distinct group in the PC space was taken out of the analysis and then PCA was applied again on the reduced data matrix. With this step, the variation within the data set is reduced and formerly overlapping groups may get separated. This is repeated until no more overlapping groups are present in the PC space. This approach was successfully utilized during the classification and quantitative analysis of 27 igneous rocks. This work was submitted to the thematic issue on EMSLIBS conference in Spectrochimica Acta B and the article is the part of this thesis, see appendix.

QUANTIFICATION

Special attention was paid to the amount of copper in the samples (ranging from 0.2 to 15 wt. %) for quantitative analysis. First step was the univariate quantitative analysis with emphasis to the concentration of copper via the formation of calibration curves. Samples of known composition were chosen and therefore the performance of the LIBS measurement can be foreseen.

Taking into account the linearity of the LIBS signal to trace element concentration (this assumption is valid and verified for lower concentrations), the calibration curve for copper was plotted, figure 35. The linearity can be observed for the intensity of the reference Cu (I) line at 521.82 nm obtained from LIBS measurement. When the copper reaches approximately 5 wt. % then the selected copper line is saturated and linearity in the calibration curve breaks.

Three different trends can be observed investigating the calibration plot constructed for Cu (I) line at 521.82 nm, figure 35. Note, that the outliers detected in the previous step were not omitted from the the qunatitative analysis. Different trends of the calibration curves reflect the strong dependency on the matrix effect and support the need for classification of the samples prior the quantitative analysis. Limits of detection for individual calibration curves were not computed due to the lack of more points forming those calibration curves. Nevertheless, constructing only one calibration curve for the whole data set may result in misleading further evaluation of the copper amount in unknown samples.

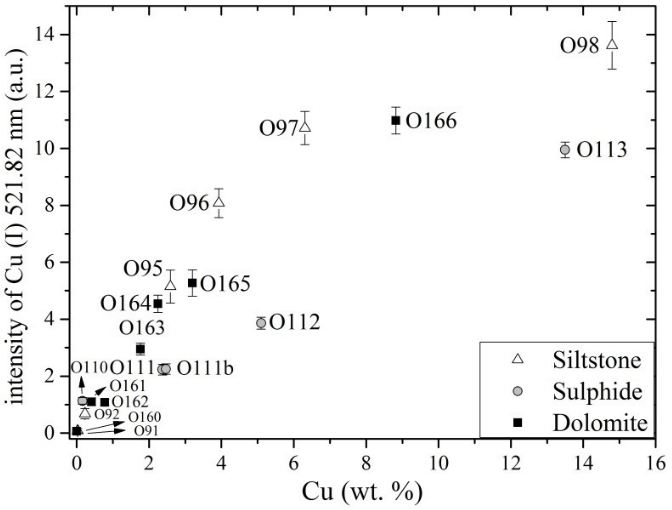


Figure 35, calibration curve for copper in Oreas standards, Cu (I) at 521.82 nm.

A further step for improving the calibration curve can be the introduction of a line of a matrix element for normalization of the calibration curve. The selection of a proper line of the matrix element can be difficult due to the varying composition of the samples. Lines of matrix elements were chosen to match energy of their higher excited state with the one of Cu line at 521.82 nm. The selection of an appropriate matrix line is challenging when complex matrices such as those of mineral ores are under investigation. Lines of matrix elements (Al, Ca, Na, and Si) were utilized for the calibration curve normalization but without any significant success, while the resulted linearity of the calibration curve was in the $R^2 \sim 0.6$ level. Plots of likewise normalized calibration curves are not shown in this work.

A typical LIBS spectrum carries abundant information content. This redundancy in spectral information may be conveniently used in multivariate analysis [2]. Consequently, the prediction power and precision of the LIBS measurement can be improved by applying more spectral lines at once. A data matrix was constructed from LIBS measurements with the emphasis on four copper lines, listed in table 3. Cu concentrations, from table 5, were organized in the calibration vector ***y***. Multivariate regression algorithms (PCR and PLSR) were then used against formerly constructed objects. Due to the linear property of multivariate algorithms for quantitative analysis, it has to be stressed that only linear data set can be introduced to PCR and PLCR for their linear property [114]. Therefore, the samples with copper content above 5.1 wt. % were omitted from the data matrix. The outliers detected with PCA and SOM in foregoing steps of the analysis were preserved.

It is noteworthy that both algorithms model the data set in a least square sense while preserving only the most valuable information and truncating the redundancy. The main difference between the algorithms is that the PLSR algorithm creates models in $\boldsymbol{X}$ and $\boldsymbol{y}$ while PCR models only the data in $\boldsymbol{X}$. Because of this feature the PLSR algorithm usually leads to a model that is able to fit response variable $\boldsymbol{y}$ with fewer components. The number of principal components used in modelling is essential and has to be chosen with care. Less PCs may not model the data matrix properly. On the other hand, too many PCs often result in overfitted models which are too optimistic and misleading.

For this reason the performance of PCR and PLSR algorithms were compared. To achieve this, all four principal components were gradually added to the model and the mean square error of prediction (MSEP) was estimated, figure 36a. In the plot the data points are a function of principal components when the optimal number PCs can be selected as its global minimum. There is no reason why both algorithms should be restricted to the same number of PCs. Though, in this case two PCs for PCR (describing 98.9% of total variation) and for PLSR (99.2%) were selected. PCR and PLSR models were created and copper contents were predicted based on selected number of principal components, figure 36b. From the inspection of the plot, the PCR model is comparable to the PLSR one. The R^2 values and MSEP confirm that, $\mathrm{R}^2_{\mathrm{PCR}} = 0.757$ and $\mathrm{MSEP}_{\mathrm{PCR}} = 0.016$ in respect to $\mathrm{R}^2_{\mathrm{PLSR}} = 0.758$ and $\mathrm{MSEP}_{\mathrm{PLSR}} = 0.015$. Though, those results are poor in the sense of coefficient of determination. This may be a consequence of vast variation in the content of matrix elements.

The prediction of PCR and PLSR can be further improved when matrix lines are added into the data matrix $\boldsymbol{X}$. This approach was suggested by Death *et al.* [115] for the modelling of ore sample measurements. The analysis then reflects as well the variance in the matrix elements. Therefore it is more likely to obtain better prediction from this kind of a model. Their theoretical assumption was then proved by the PCR prediction of Si, Al, and Fe contents. Therefore, Al, Ca, and Si lines were added to the Cu lines in the data matrix prior further analysis in this case study. Thus altered data matrix was then analysed with PCR and PLSR algorithms in the same steps as in foregoing case. The number of components was for both algorithms chosen after inspection of figure 37a. From the plot is obvious that PCR with 6 components may do the same job as PLSR with 5 components in the means of MSEP. This result proves the already stated fact that PLSR needs fewer components to create a reasonable regression model.

Copper amounts were predicted employing both algorithms with selected PCs, figure 37b. The PCR prediction is slightly better than the PLSR one. However more components were used to model the data matrix. Figures of merit confirm this statement, $\mathrm{R}^2_{\mathrm{PCR}} = 0.92$ and $\mathrm{MSEP}_{\mathrm{PCR}} = 0.0087$ in respect to $\mathrm{R}^2_{\mathrm{PLSR}} = 0.917$ and $\mathrm{MSEP}_{\mathrm{PLSR}} = 0.0062$. The approach stated by Death *et al.* results in better linearity of the calibration curve, *i.e.* in better coefficient of determination R^2, than in the case of the model with Cu lines only. The PLSR regression model does not require so many components as PCR. For this reason the PLSR model may be considered more parsimonious to the data. On the other hand, PCR is more apt for the modelling of the data matrix $\boldsymbol{X}$ than PLSR algorithm. As can be seen in figure 38, first few principal components of PCR describe more variance in the data matrix then those of PLSR. Therefore the selection of the more appropriate multivariate regression algorithm depends on the case of the study.

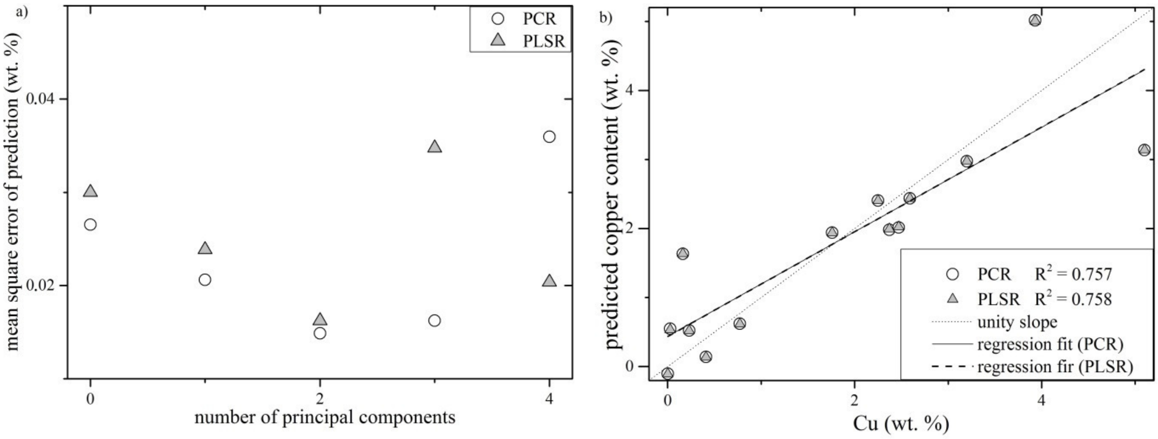


Figure 36, a) mean square prediction error for individual PCs, b) predicted copper contents utilizing PCR and PLSR.

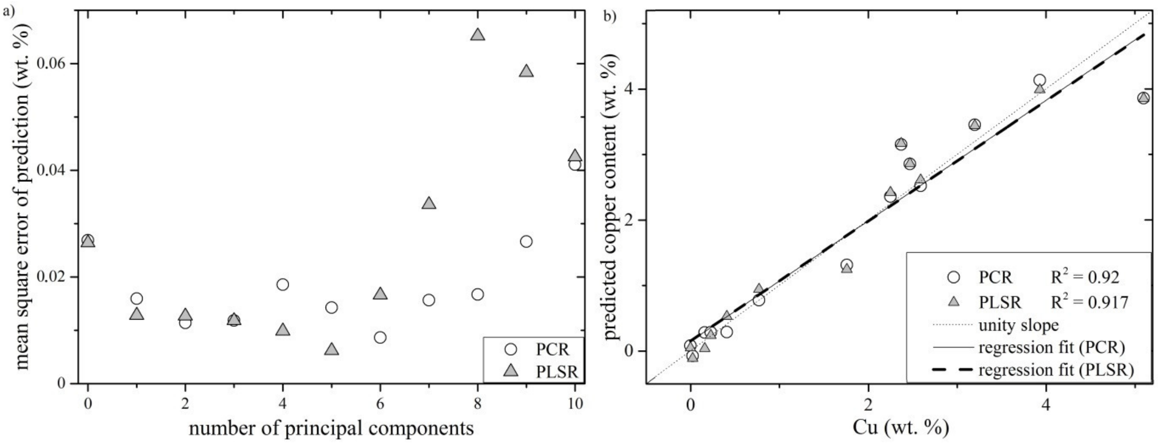


Figure 37, a) mean square prediction error for individual PCs, b) predicted copper contents utilizing PCR and PLSR.

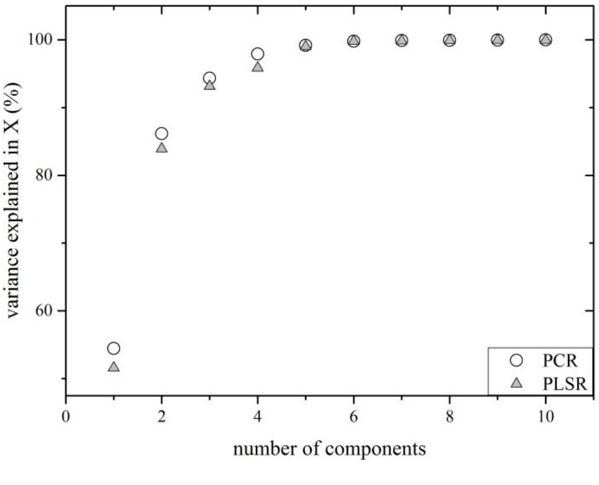


Figure 38, variance explained in the data matrix ***X*** by PCR and PLSR.

It the last step of the analysis, the classification and quantitative analysis are beneficially employed together. When the samples are isolated in distinct groups based on their matrices, partial calibration curves can be plotted. For those purposes, the data matrix was treated of the outliers. Then PCA was once more applied, figure 39a, where the reduced data matrix is decomposed into three distinct groups. First two PCs depicted in the plot sum up 93.9% (84.7% and 9.2%) of overall variation within the reduced data matrix. Afterwards, corresponding calibration curves are plotted, figure 39b. The copper content in an unknown sample may be obtained directly from corresponding calibration curve without any need for utilization of multivariate regression.

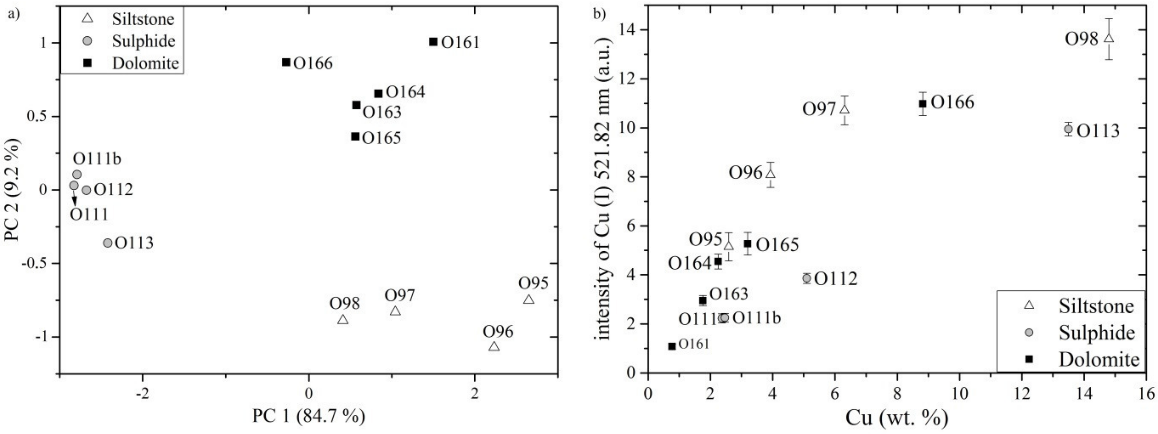


Figure 39, reduced data set, a) PCA analysis, b) calibration curve.

Concluding the quantification part of this case study, the univariate calibration falls apart when samples with varying matrices are analysed at once. The prediction power of such constructed calibration curve is poor. Therefore samples should be clustered prior the quantitative analysis, in order to avoid the uncertainties caused by the matrix effect, and partial calibration curves should be plotted to improve the prediction power of quantitative analysis.

4.3.CASE STUDY III

The sample set was compiled of 80 samples, 52 andesite and 28 diorite samples, collected in north-west region in Iran. As in the first case study, the samples were analysed using ICP-MS after four acid digestions at the University of Clausthal. Those samples were assigned to groups at the same place.

Investigating the QAPF diagram, the elemental composition of both igneous rocks is closely related [109, 110]. Diorite is one of the plutonic (intrusive) rocks which were formed from magma that cooled down under the surface of the Earth. Their composition was influenced by the pre-existing rocks into which they intruded. Andesite is a member of the volcanic (extrusive) rocks which were formed from the magma on the surface of the planet. The magma which formed plutonic rocks cooled down much faster than in the case of volcanic rocks. Therefore, volcanic and plutonic rocks may be distinguished by the size and the shape of their grains. The QAPF diagram is equivalent for both types of igneous rocks and therefore the chemical composition in the sense of Al, Ca, Na, and Si may not be the only decisive factors for the classification process.

The samples were measured in the form of fine dust pressed onto the surface of DST. The LIBS system was set as listed in table 1. Representative 200 laser pulses per sample (corresponding to 200 laser-induced plasmas) were observed. The exposed spectra were averaged and analysed utilizing the algorithm which was already described. The data matrix was composed as in foregoing cases, rows refer to averaged spectra of a sample and columns refer to variables. The typical spectra of both igneous rock types are depicted on figure 40. Despite the fact that both rock types are close neighbours in the QAPF diagram the differences are noticeable in all of marked matrix lines. Therefore, lines of Al, Ca, Na, and Si were utilized in the computation of principal components, lines listed in table 2.

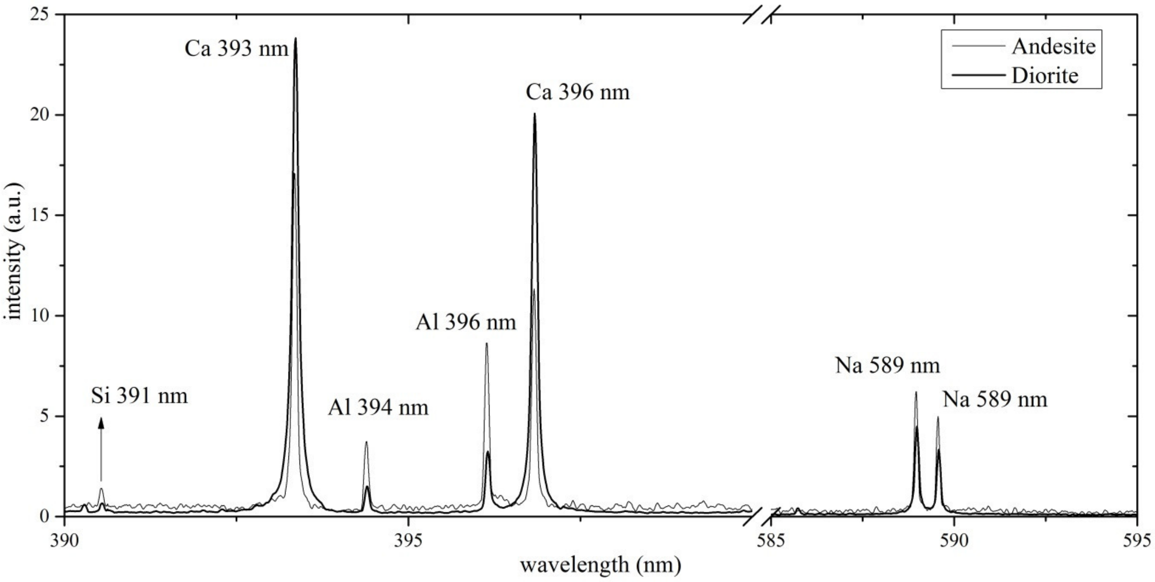


Figure 40, typical LIBS spectrum of andesite and diorite (bold line) rocks.

CLASSIFICATION

PCA algorithm was applied to the data matrix in order to detect outliers and primarily to investigate possible clustering of measured samples to appropriate groups, figure 41. First three principal components covering 97.3% of overall variation (64.9%, 19.9% and 12.5%) are cross-plotted. Andesite and diorite clusters are amply separated along the first principal component. The discrimination of andesite and diorite is dependent mainly on the contribution of Al, Ca and Si as can be seen in the loadings plot, figure 42. Those matrix elements are scattered along the first principal component. However it cannot be stated that one of those elements is superior to the others. Therefore, the principal component should be associated to the contribution of all matrix elements at once in a way which is depicted on both parts of the loading plot. On the other hand, the contribution of Na to the discrimination is rather secondary and do not have direct impact on the sample classification. However, it can be stated that diorite rocks are composed from higher amount of Al and Si and lower amount of Ca compared to andesite rocks, as suggest intensities of relevant spectral lines in figure 40 and values of related loading factors depicted on figure 42.

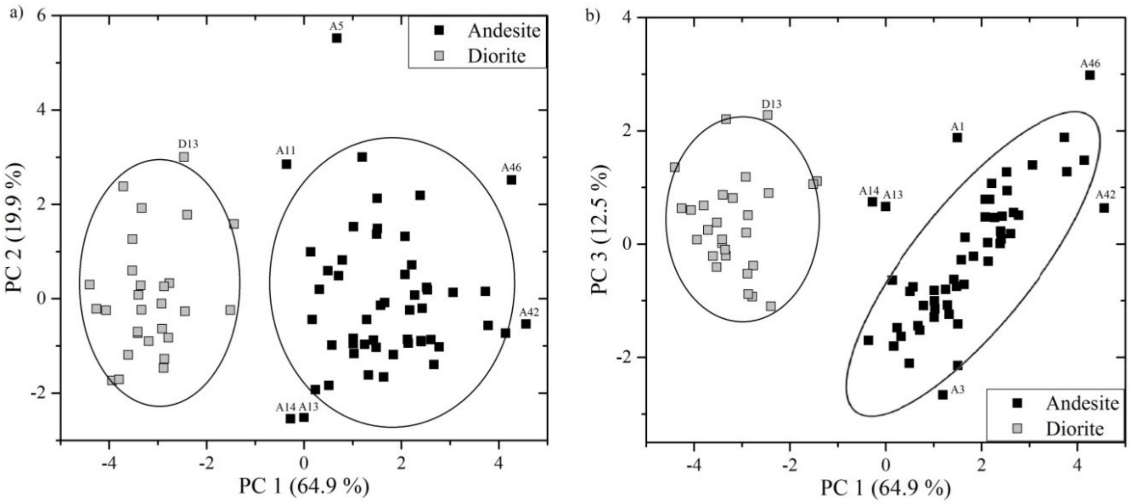


Figure 41, PCA scores, a) 1st and 2nd principal component, b) 1st and 3rd principal component.

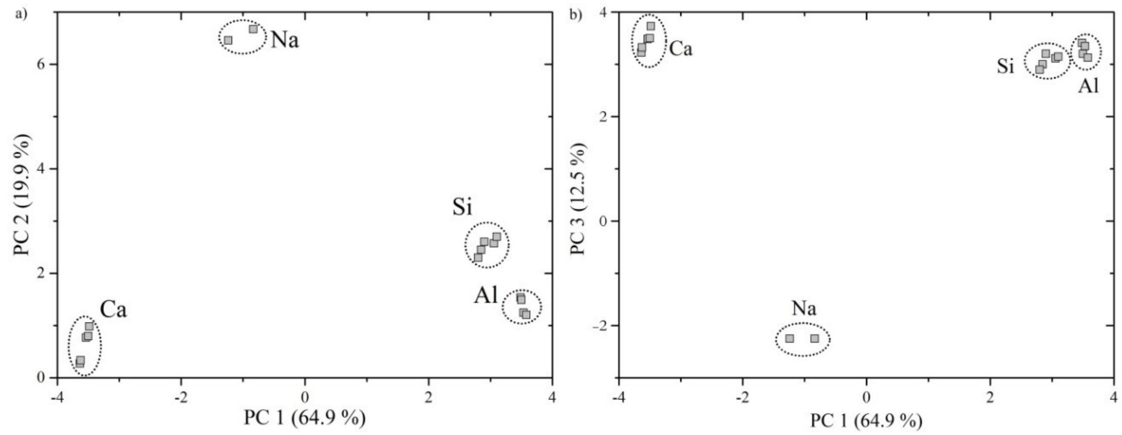


Figure 42, PCA loadings, a) 1st and 2nd principal component, b) 1st and 3rd principal component.

No misclassified samples were detected but several samples outlying from the appropriate clusters are present. Those samples (namely D13, A1, A3, A5, A11, A13, A14, A42, and A46) may be discarded from the further analysis, but this was not done in this case study. All of the samples were remained in the data set to test the robustness of the analysis. Furthermore, it could be expected that samples should be clustered more closely in the PC space. However, considering vast possible diversity in the composition of the samples collected in-field the resulted discrimination of the samples may be accepted as satisfactory.

The Gaussian clustering algorithm, built-in to the MATLAB Statistics Toolbox, was used to discriminate individual groups of samples according to their distribution in the PC space. Sufficient number of PCs was selected to describe 95% of overall variation within the data, in this case three PCs. Figure 41 shows the fitting result projected onto first two PCs. The ellipsoids in the projection plot mark the area that covers 99% of the volume of the particular Gaussian sub-distribution. Some clusters can overlap when discriminating samples with similar composition. One should note that Gaussian clustering was used only for enclosing the points into separate groups. For this purpose spectra for one sample were not averaged but plotted individually, this leads to increase the number of points in the PC space.

It was proved by the means of PCA that the samples are not misclassified. In the next step of the analysis SIMCA algorithm was utilized for classification of ‘unknown’ samples. No testing data set was present and therefore LOSO algorithm was utilized. In this case, spectra of each sample were not averaged to build up a data matrix with more data points/rows (1600 data points in total). Step-by-step was each individual sample (represented by 20 spectra) taken out of the data matrix. For each rock type a model was established with SIMCA. Then the omitted sample was classified based on the similarities with SIMCA models. A total of 13 spectra were mismatched from total 1600 spectra applied to the SIMCA, table 7. After such analysis the resulted uncertainty of a sample being misclassified is below 1%.

Table 7, the results of quantification utilizing SIMCA.

Rock type	Total no. of samples	Total no. of spectra	Mismatched spectra
andesite	52	1040	7
diorite	28	560	6

QUANTIFICATION

The reliable quantification of copper content is necessary for the purposes of the mining industry. Calibration curve was constructed utilizing copper line at 521.82 for all the samples, figure 43a. This calibration line gets saturated when the amount of copper in the sample reaches approximately 1 wt. %. The linear part of this curve was approximated by a line with coefficient of determination $R^2 = 0.86$. This may suggest that the matrix effect is not that significant as in foregoing case studies. Samples were divided into two groups according to their class membership and then an individual calibration curve was constructed for each class, figure 43b. However, this attempt improved the prediction power only moderately in the case of andesite samples, $R^2 = 0.88$. But the prediction power of the calibration curve for diorite samples is lower than the one estimated for all samples, $R^2 = 0.81$.

Spectral lines of Al, Ca, Na, and Si were utilized to normalize those calibration curves. However, normalization of the calibration curve to one of the matrix lines was not successful. Resulted coefficients of determination were significantly lower than those already presented. This may be a consequence of a complex dependence on multiple elements as proved by a PCA loadings plot, figure 42. There the first principal component responsible for the main sample classification was related to the contribution of Al, Ca, and Si. Normalized calibration curves are not shown for this case study.

The last step of this analysis was the prediction of a copper content by the means of PCR and PLSR, figure 44b. The range of copper content in the samples was reduced so that only the linear part of the calibration curve, figure 43a. The data matrix consisted of all matrix elements formerly selected for PCA analysis and copper lines, table 3. The algorithm was applied as in the foregoing case study, optimal number of principal components was estimated from the MSEP plot, figure 44a. There is a slight improvement in the prediction power when PCR and PLSR are utilized, $R^2 = 0.9$ for both algorithms. PLSR with four components does a comparable job as PCR with five components.

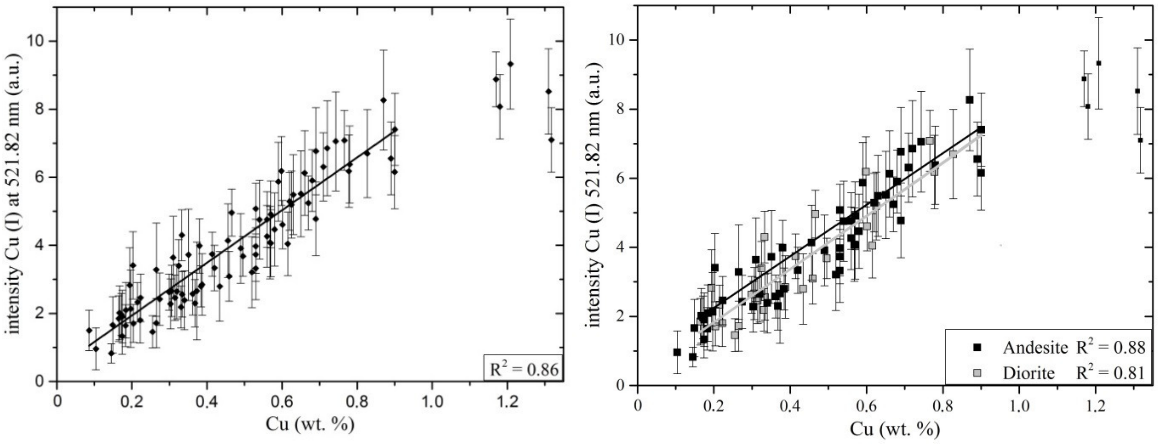


Figure 43, calibration curve for copper line Cu (I) at 521.82 nm utilizing a) all samples at once, b) samples discriminated with PCA analysis.

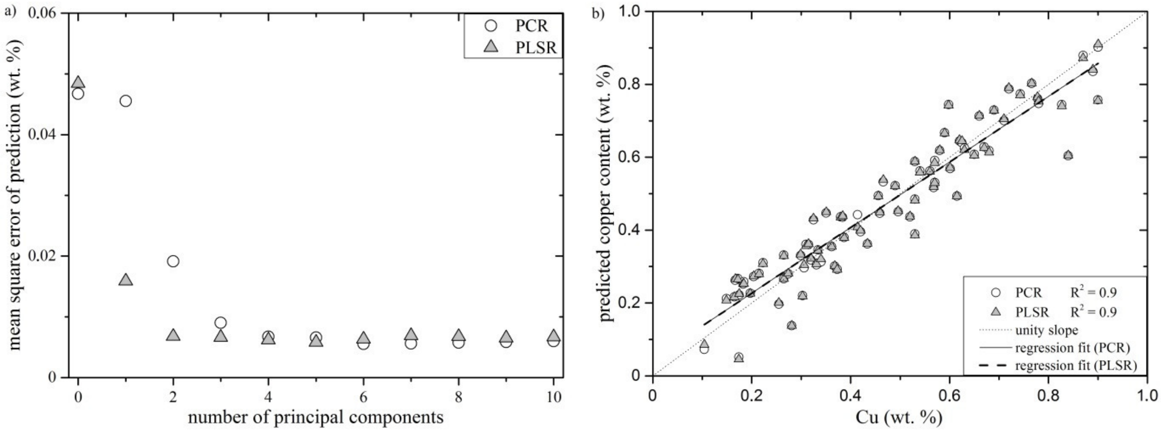


Figure 44, a) mean square prediction error for individual PCs, b) predicted copper contents utilizing PCR and PLSR.

The data matrix was divided into two parts according to the class membership, figure 45a) andesite and b) diorite. For each class separate models based on PCR and PLSR were estimated. In both cases the separate modelling lead to a slight improvement in the means of coefficient of determination. Figures of merit listed for the andesite model, $\mathrm{R}^2_{\mathrm{PCR}} = 0.914$ and $\mathrm{MSEP}_{\mathrm{PCR}} = 0.0059$ for 7PCs in respect to $\mathrm{R}^2_{\mathrm{PLSR}} = 0.917$ and $\mathrm{MSEP}_{\mathrm{PLSR}} = 0.006$ for 5PCs and for the diorite model $\mathrm{R}^2_{\mathrm{PCR}} = 0.93$ and $\mathrm{MSEP}_{\mathrm{PCR}} = 0.0031$ with 6PCs in respect to $\mathrm{R}^2_{\mathrm{PLSR}} = 0.93$ and $\mathrm{MSEP}_{\mathrm{PLSR}} = 0.0038$ for 5 PCs. This step in the analysis lowers the diversity in the data set and consequently eases the computational severity and inaccuracy of the model. However, this last step may be seen as redundant in the quantification analysis. While to the best of my knowledge based on the literature research multivariate algorithms for classification and quantitative analysis are not used together.

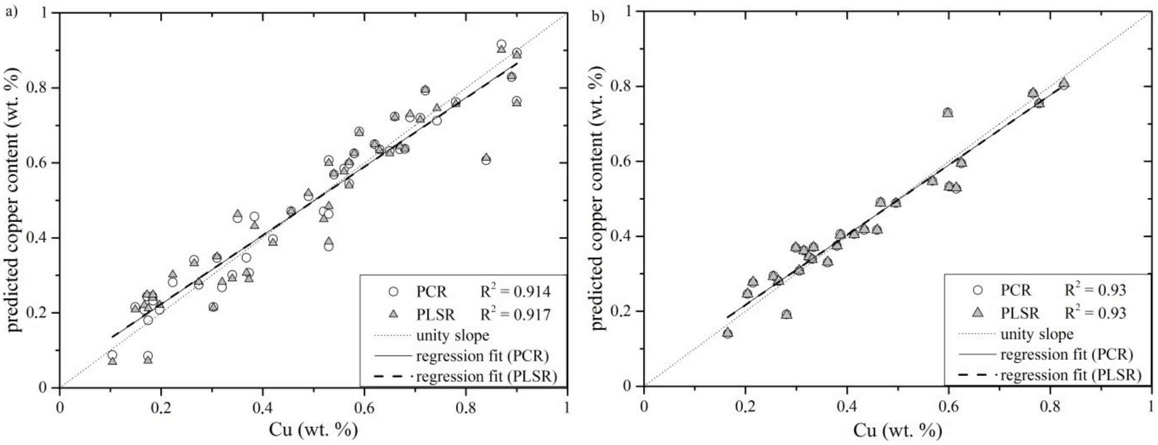


Figure 45, PCR and PLSR prediction of copper content for a) andesite samples, b) diorite samples.

PCA was successfully utilized to classify the samples of two different rock types. This classification was then emulated with the SIMCA algorithm. Moreover, it was proved that multivariate regression (PCR and PLSR) may compensate the matrix effect to a certain extent when data matrix is composed from spectral lines of an analyte as well as lines of matrix elements. It may be further stated based on the improved linearity of predicted analyte content from PCR and PLSR models and on the results given in the article, see appendix. The content of analyte predicted from PCR and PLSR models are observed with improved bias, *i.e.* predicted analyte content is closer to the certified value.

CONCLUSION

The results shown in this thesis proved the pertinence of the LIBS in mining industry. The primary aim was the determination of copper content in various igneous rocks. It was observed that the matrix effect may be significant even for smaller data sets consisting of samples of the same igneous rock. Therefore, it is recommended to discriminate the sample set with the emphasis on the similarity in the composition of matrix elements. Consequently, more reliable quantitative analysis of trace elements may be achieved following this recommendation.

In the frame of three case studies the various ways of sample pretreatment were tested. The univariate calibration curve was successfully utilized with the normalization to selected matrix element for limited number of samples. For this reason, the data sets, which are more complex, were treated with multivariate regression algorithms. The multivariate quantitative analysis was provided by the means of PCR and PLSR and the performance of both algorithms was compared in detail. PCR and PLSR may compensate the matrix effect to a certain extent. Though it was proved in general case (or for bigger sample set) is such analysis unsatisfactory. Therefore, PCA was employed on the data set to reveal possible outliers and clusters of data/samples when expecting the newly constructed PC space. Subsequently, the performance of PCA was then emulated by Kohonen Maps, which are considered analogical to PCA. Finally, a robust data set was classified by the means of SIMCA with the emphasis on matrix elements.

As for the experimental part of this thesis, an algorithm for a reliable data analysis was proposed. It is recommended in the first step of the analysis to apply PCA algorithm to detect outliers in the data set with any further intention. Then PCA should be applied once more to reveal possible clusters based on the similarities of the samples in the composition of their matrix elements. Moreover, it is beneficial to create partial calibration curve for each data cluster individually. In that way the occurrence of the matrix effect can be avoided and reliable and satisfactorily accurate calibration curves may be plotted. Besides, the matrix effect may be suppressed to a certain extent when multivariate calibration is applied. It is noteworthy that the data matrix may be constructed from lines (variables) of matrix elements only to sufficiently more accurately reflect the variance among the sample matrices.

Chemometric algorithms were successfully utilized for the multivariate classification and quantitative analysis of LIBS data. Further work using the stand-off LIBS device is therefore scheduled with the aim to create necessary data libraries which are irreplaceable part of successful *in-situ* multivariate analysis.

SHRNUTÍ

Výsledky uvedené v této disertační práci potvrzují vhodnost použití metody LIBS v těžebním průmyslu. Primárním cílem bylo stanovit množství mědi v různých druzích vyvřelých hornin. Bylo zjištěno, že matriční jev může výrazně ovlivnit měření metodou LIBS, přestože byly naměřeny pouze malé soubory vzorků stejné horniny. Proto je doporučeno roztřídit vzorky s ohledem na podobnosti ve složení matric, respektive poměry množství matričních prvků. Následně je možné určit množství analytu ve vzorku s větší přesností.

V rámci třech případových studií byly naměřeny vzorky různých magmatických hornin a způsoby jejich přípravy před vlastním meřením metodou LIBS. Kalibrační křivky vytvořené pomocí jedné spektrální čáry analytu byly úspěšně použity (spolu s linearizací vzhledem ke spektrální čáře matričního prvků), ovšem pouze pro omezený rozsah vzorků. Z tohoto důvodu byly komplexnější soubory dat analyzovány pomocí multivariačních regresních metod, PCR and PLSR, jejichž výsledky byly podrobně srovnány. Multivariační kvantitativní metody mohou vyvážit matriční jev, avšak pouze do určité míry. Bylo potvrzeno, že v obecném případě jsou tyto metody nedostačující. Proto byla metoda PCA aplikována na soubor dat se záměrem odhalit odlehlé body/vzorky (tj. vzorky výrazně se lišící od zbytku dané skupiny nebo celého souboru dat) a skupiny (klastry/shluky) dat/vzorků prozkoumáním nově vytvořeného prostoru hlavních komponent. Výsledné zpracovaní dat pomocí metody PCA bylo napodobeno pomocí Kohonenových map, které jsou analogické k PCA. Na závěr experimentální části byl obsahlý soubor dat úspěšně roztříděn metodou SIMCA s důrazem na matriční prvky.

Na základě experimentálních měření byl ustanoven algoritmus pro spolehlivou analýzu dat. V prvním kroku je tedy doporučeno použít metodu PCA na soubor dat výhradně k detekci odlehlých dat/vzorků. Poté je vhodné použít metodu PCA znovu se záměrem roztřídit soubor dat na jednotlivé skupiny podle podobnosti ve složení jejich matric. Následně je možné vytvořit kalibrační křivky samostatně pro jednotlivé skupiny vzorků. Takto vytvořené kalibrační křivky jsou dostatečně přesné pro použití v těžebním průmyslu. Tímto způsobem je možné dostatečně předejít matričnímu jevu a výrazně tak omezit jeho negativní účinek na LIBS měření. Navíc, matriční efekt lze do určité míry eliminovat i pomocí multivariačních metod PCR and PLSR. Je vhodné zmínit, že jednotlivé vzorky mohou být dostatečně reprezentovány pouze použitím spektrálních čar matričních prvku.

Chemometrické algoritmy byly úspěšně použity při multivariační kvantifikaci a klasifikaci magmatických hornin naměřených metodou LIBS. Dalším cílem stavějícím na základech položených touto disertační prací je vytvoření knihoven dat, které následně budou nenahraditelnou součástí úspěšné *in-situ* multivariační analýzy.

ZUSAMMENFASSUNG

Die Ergebnisse zeigen die Bedeutung von LIBS in der Bergbauindustrie, wobei ein Hauptaugenmerk auf der Bestimmung von Kupfer in vulkanischen Gesteinsproben lag. Es konnte beobachtet werden, dass der Matrixeffekt einen signifikanten Einfluss auf selbst kleine Datensätze bestehend aus Proben des gleichen Gesteins hat. Anhand dieser Erkenntnisse ist es empfehlenswert Proben schwerpunktmäßig nach der Zusammensetzung der Matrixelemente zu klassifizieren. Unter Berücksichtigung dieses Hinweises kann eine verlässliche Quantifizierung von Spurenelementen erhalten werden.

Im Rahmen dreier Fallstudien wurden unterschiedliche Probenvorbereitungen getestet. Eine univariate Kalibriergerade konnte nur für eine begrenzte Anzahl an Proben unter Normalisierung, bezogen auf ausgewählte Matrixelemente, eingeführt werden. Aus diesem Grund wurden Komplexere Datensätzen mittels multivariater Regression ausgewertet. Die multivariate quantitative Analyse wurde mit Mittelwerten von PCR und PLSR ausgeführt und jeweilige Ergebnisse verglichen. PCR und PLSR können den Matrixeffekt nur bis zu einer bestimmten Größenordnung kompensieren und um dies zu bestätigen wurde eine umfangreiche Studie, mit den erwartet-unbefriedigenden, Ergebnissen erstellt. Um mögliche Ausreißer und somit Cluster besser zu erkennen wurde die PCA eingeführt. Die Leistung der PCA wurde im nächsten Schritt mittels Kohonenkarte überprüft, welche gleiche Annahmen macht wie die PCA. Schließlich konnte eine Klassifizierung der Proben, mit Mittelwerten der SIMCA-Analyse und der hohen Gewichtung von Matrixelementen, realisiert werden.

Für eine zuverlässige Datenanalyse konnte ein Algorithmus entwickelt werden. Zunächst wird im ersten Schritt eine PCA durchgeführt werden um Ausreißer zu erkennen und zu entfernen. Im zweiten Schritt wird die PCA erneut durchgeführt um Cluster aufgrund der jeweiligen Ähnlichkeit der Matrices von Gesteinsproben zu erkennen. Darüber hinaus ist es von Vorteil für jeden Cluster eigene Kalibrierkurven zu erstellen um den Einfluss von Matrixeffekten zu minimieren und so verlässliche, zufriedenstellende Daten zu erzeugen. Zusätzlich wird die multivariate Datenanalyse angewendet um weitere Auswirkungen der Matrices auf Analysedaten zu unterdrücken. Schließlich soll noch erwähnt werden, dass die Analyse von Matrixelement Signalen (Variablen) sogar ausreichend ist um die Gesamtheit einer Gesteinsprobe zu beschreiben.

Chemometrische Algorithmen wurden erfolgreich eingesetzt um multivariate Klassifizierungen und quantitative Analysen mittels LIBS durchzuführen. Weitere Arbeiten unter Einsatz des stand-off LIBS Systems sind, mit dem Ziel benötigte Datenbanken aufzubauen, geplant. Diese stellen einen unerlässlichen Teil von multivariaten *in-situ* Analysen dar.

LIST OF ABBREVIATIONS

AES	Atomic Emission Spectroscopy
ANN	Artificial Neural Network
BMU	Best Matching Unit
CD	Collisional-Dominated
CF	Calibration-Free
CR	Collisional-Radiative
CV	Cross Validation
DP	Double Pulse
DPSS	Diode Pumped Solid State
DST	Double Sided adhesive Tape
FWHM	Full Width at Half Maximum
ICCD	Intensified Charge-Coupled Device
ICP	Inductively Coupled Plasma
LA	Laser Ablation
LIBS	Laser-Induced Breakdown Spectroscopy
LIFS	Laser-Induced Fluorescence Spectroscopy
LIP	Laser-Induced Plasma
LIPS	Laser-Induced Plasma Spectroscopy
LOD	Limit Of Detection
LOQ	Limit Of Quantification
LOL	Limit Of Linearity
LOSO	Leave One Sample Out
LTE	Local Thermodynamic Equilibrium
MS	Mass Spectrometry
Nd:YAG	Neodymium-doped Yttrium Aluminium Garnet
NIPALS	Nonlinear Iterative PArtial Least Squares
OD	Orthogonal Distance
OES	Optical Emission Spectrometry
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
PLSR	Partial Least Squares Regression
PMT	PhotoMultiplier Tube
PRESS	PREdiction Sum of Squares
QAPF	Quartz Alkali feldspar Plagioclase Feldspathoid
RSD	Relative Standard Deviation
RSS	Residual Sum of Squares
SIMCA	Soft Independent Modelling of Class Analogies
SNR	Signal to Noise Ratio
SOM	Self-Organizing Map

SP	Single Pulse
SVD	Singular Value Decomposition
TAS	Total Alkali Silica
UV	Ultra Violet
XRF	X-Ray Fluorescence

LIST OF SYMBOLS

1. LASER-INDUCED BREAKDOWN SPECTROSCOPY

a	the variable of the Voigt profile function
A	the ion broadening parameter
A_{ij}	the spontaneous transition probability of the quantum transition (Einstein coefficient)
B	the coefficient equal to 1.2 or 0.75 for ionic or neutral lines respectively
BEC	the background equivalent concentration
C^I	the relative concentration of emitting species in the plasma
D_L	the diameter of the spot on the lens
D_S	the angular dispersion of the spectrometer
$d\delta$	the angle separating two light beams
$d\lambda$	the wavelength difference
ΔE	the highest energy of quantum transition for which the McWhirter criterion is valid
ΔE_{ion}	the lowering correction parameter
$\Delta\lambda$	the least distance between two resolvable wavelengths
$\Delta\lambda_{Stark}$	the Stark broadening of the line
E_i	the excitation energy of the excited state *i*
E_{ion}	the first ionization potential energy
f	the frequency of the laser radiation
f_{det}	the function referring to the radiation environment
f_{exc}	the function referring to the excitation/ionization mechanisms leading to the atomic/ionic emission
f_{int}	the function referring to the ablation/vaporization of solid material
f_L	the focal length of the lens
f_S	the focal length of the spectrometer
F^{abs}	the wavelength independent factor
g_i	the statistical weight of the excited state *i*
G_d	the gate delay
G_w	the gate width
$\Gamma_V(\lambda_e)$	the shape of a spectral line described by the pseudo-Voigt profile
h	the Planck constant
I	the superscript referring to atomic species
II	the superscript referring to ionic species
i	the index of upper quantum state of an atom
I	the number of samples
I_{ij}	the integrated intensity of the elemental line
j	the index of lower quantum state of an atom
J	the number of variables
k	the Boltzmann constant
k_C	the constant of calibration curve
K	the number of observed outcomes
$K(u,a)$	the Voigt profile function

λ	the wavelength of the laser radiation
λ_e	the wavelength of the emitted characteristic radiation
λ_0	the central wavelength of the spectral line
m	the electron mass
m	the index referring to the upper state of singly ionized atom
n	the index referring to the lower state of singly ionized atom
n_e	the number density of electrons
n^p	the total particle density of the plasma
n^s	the total number density of the species s in the plasma
n_i^s	the population of the species s in the excited quantum state
R^2	the coefficient of determination
R_S	the resolution of the spectrometer
RSD_B	the relative standard deviation of the background
S	the slope of the linear part of the calibration curve
S_{el}	the signal of the elemental line
σ	the standard deviation of the background in the blank sample
T_e	the electron temperature
u	the variable of the Voigt profile function
$U^s(T_e)$	the internal partition function of the species at the temperature T_e
w	the electron impact parameter
w_0	the waist of the laser beam
$\boldsymbol{X}$	the data matrix
$\bar{X}_B$	the intensity of the background
$\boldsymbol{Y}$	the regression matrix

2. CHEMOMETRICS

a	the index referring to the individual principal component
A	the total number of principal components
$\alpha(t)$	the learning rate
b_i	the best matching unit
B	the PLSR coefficient
c	the index referring to the class in SIMCA model
$\boldsymbol{C}$	the weight matrix of $\boldsymbol{Y}$
$\boldsymbol{E}$	the error (residual) matrix
$\boldsymbol{F}$	the error (residual) matrix of $\boldsymbol{Y}$
$h_{b_i l}(t)$	the neighbourhood function
H_l	the number of neurons surrounding the neuron l
i	the index referring to the individual sample/measurement
I	the total number of samples/measurements
j	the index referring to the individual variable
J	the total number of variables
k	the index referring to the individual observed outcome
K	the total number of observed outcomes

κ_a	the size of the *a*th component in the PLS model
l	the index referring to the individual neuron
L	the total number of neurons
λ_a	the value of the *a*th eigenvalue
$\mathcal{N}_{b_i}$	the neighbourhood of the winning neuron
$\mathcal{N}_l$	the neighbourhood of the neuron *l*
OD^c	the orthogonal distance to class *c*
$\boldsymbol{P}$	the loadings matrix
$\boldsymbol{r}_{b_i}$	the position of the winning neuron in the map
$\boldsymbol{r}_l$	the position of the neuron *l* in the map
s	the vector of regression coefficients
$\sigma(t)$	the neighbourhood radius
t	the index referring to the individual iteration
T	the total number of iterations
$\boldsymbol{T}$	the scores matrix
u_l	the Euclidean distance among the weight vector of a neuron *l* and weight vectors of neurons in the neighbourhood $\mathcal{N}_l$
$\boldsymbol{U}$	the scores matrix of $\boldsymbol{Y}$
V_a	the value of the *a*th eigenvalue in percentage
w_l	the weight (prototype) vector of neuron *l*
$\boldsymbol{W}^*$	the weight matrix of $\boldsymbol{Y}$
x_{ij}	the observed data
$\hat{x}_{ij}$	the predicted data
$\boldsymbol{X}$	the data matrix
$\hat{y}$	the unknown sample
$\hat{y}^c$	the projection of the unknown sample by PCA model of class *c*
$\boldsymbol{Y}$	the regression matrix

REFERENCES

[1] C. A. Munson, F. C. De Lucia Jr., F. Piehler, K. L. McNesby and A. W. Miziolek, "Investigation of statistics strategies for improving the discriminating power of laser-induced breakdown warfare agent stimulants," *Spectrochimica Acta Part B,* vol. 60, pp. 1217-1224, 2005.

[2] A. W. Miziolek, V. Palleschi and I. Schechter, Laser-Induced Breakdown Spectroscopy (LIBS) Fundamentals and Applications, Cambridge, UK: Camridge University Press, 2006.

[3] D. A. Cremers and L. J. Radziemski, Handbook of Laser-Induced Breakdown spectroscopy, New York, USA: John Wiley & Sons, Ltd., 2006.

[4] R. Noll, Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications, Springer, 2012.

[5] F. J. Fortes, J. Moros, P. Lucerna, L. M. Cabalín and J. J. Laserna, "Laser-Induced Breakdown Spectroscopy," *Anal. Chem.,* vol. 85, pp. 640-669, 2013.

[6] S. J. Rehse, H. Salimnia and A. W. Miziolek, "Laser-induced breakdown spectroscopy (LIBS): an overview of recent progress and future potential for biomedical applications," *Journal of Medical Engineering & Technology,* vol. 36, pp. 77-89, 2012.

[7] R. S. Harmon, R. E. Russo and R. R. Hark, "Application of laser-induced breakdown spectroscopy for geochemical and environmental analysis: A comprehensive review," *Spectrochimica Acta Part B (Article in press),* 2013.

[8] R. Gaudiuso, M. Dell´Aglio, O. De Pascale, G. S. Senesi and A. De Giacomo, "Laser Induced Breakdown Spectroscopy for Elemental Analysis in Environmental, Culturar Heritage and Space Applications: A Review of Methods and Results," *Sensors,* vol. 10, pp. 7434-7468, 2010.

[9] D. Santos Jr., L. C. Nunes, G. G. Arantes de Carvalho, M. da Silva Gomes, P. F. de Souza, F. de Oliveira Leme, L. G. Cofani dos Santos and F. J. Krug, "Laser-induced breakdown spectroscopy for analysis of plant materials: A review," *Spectrochimica Acta Part B,* vol. 71, pp. 3-13, 2012.

[10] J. D. Winefordner, I. B. Gornushkin, T. Correll, E. Gibb, B. W. Smith and N. Omenetto, "Comparing several atomic spectrometric methods to the super stars: special emphasis on laser induced breakdown spectrometry, LIBS, a future superstar," *J. Anal. At. Spectrom.,* vol. 19, pp. 1061-1083, 2004.

[11] R. C. Wiens, S. Maurice, J. Lasue and a. et, "Pre-flight calibration and initial data processing for the ChemCam laser-induced breakdown spectroscopy instrument on the Mars Science

Laboratory rover," *Spectroschimica Acta Part B,* vol. 82, pp. 1-27, 2013.

[12] D. W. Hahn and 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," *Applied Spectroscopy,* vol. 64, pp. 335-366, 2010.

[13] D. W. Hahn and N. Omenetto, "Laser-Induced Breakdown Spectroscopy (LIBS), Part II: Review of Instrumental and Methodological Approaches to Material Analysis and Applications to Different Fields," *Applied spectroscopy,* vol. 66, pp. 347-419, 2012.

[14] C. Aragón and J. A. Aguilera, "Characterization of laser induced plasmas by optical emission spectroscopy: A review of experiments and methods," *Spectrochimica Acta Part B,* vol. 63, pp. 893-916, 2008.

[15] L. J. Radziemski, "From LASER to LIBS, the path of technology development," *Spectrochimica Acta Part B,* vol. 57, p. 1109–1113, 2002.

[16] G. C. Dacey, "Optical masers in science and technology," *Science,* vol. 135, pp. 71-74, 1962.

[17] F. Brech and L. Cross, "Optical Micro Emission Stimulated by a Ruby MASER," *Applied spectroscopy,* vol. 16, p. 59, 1962.

[18] J. A. Maxwell, "The laser - as a source in emission spectroscopy," *Chem. Can.,* pp. 10-11, 1963.

[19] R. C. Rosan, M. K. Healy and W. F. McNary, Jr., "Spectroscopic ultramicroanalysis with a laser," *Science,* Vols. 236-237, p. 142, 1963.

[20] R. C. Rosan, D. Glick and F. Brech, "Progress in laser microprobe emission spectroscopy," *Federation proceedings,* vol. 24, p. 542, 1965.

[21] E. F. Runge, R. W. Minck and F. R. Bryan, "Spectrochemical analysis using a pulsed laser source," *Spectrochimica Acta,* vol. 20, pp. 733-736, 1964.

[22] J. Debras-Guédon and N. Liodec, "De l'utilisation du faisceauissu d'un amplificateur á ondes lumineuses par émissioninduite par rayonnement (laser á rubis), comme source énergétique pour l'excitation des spectres d'émission des éléments," *Comptes rendus hebdomadaires des séances de l'Académie des sciences,* vol. 257, pp. 3336-3339, 1963.

[23] J. Debras-Guédon and N. Liodec, "Sur une extension despossibilités de l'analyse spectrale dans le domaine del'analyse ponctuelle par utilisation d'un laser," *Bull. Soc. Fr. Ceram.,* pp. 61-68, 1963.

[24] E. Archbold and T. P. Hughes, "Electron temperature in a laser heated plasma," *Nature,* vol. 204, p. 670, 1964.

[25] J. F. Ready, "Development of plume of material vaporized by giant-pulse laser," *Applied Physics Letters,* vol. 3, pp. 11-13, 1963.

[26] E. Archbold, D. W. Harper and T. P. Hughes, "Time resolved spectroscopy of laser-generated microplasmas," *British Journal of Applied Physics,* vol. 15, pp. 1321-1327, 1964.

[27] M. Baudelet and B. W. Smith, "The first years of laser-induced breakdown spectroscopy," *Journal of analytical atomic spectrometry,* vol. 28, pp. 624-629, 2013.

[28] T. R. Loree and L. J. Radziemski, "Laser-induced breakdown spectroscopy: time-integrated applications," *Plasma Chemistry and Plasma Processing,* vol. 1, pp. 271-279, 1981.

[29] F. J. Fortes and J. J. Laserna, "The development of fieldable laser-induced breakdown spectrometer: No limits on the horizon," *Spectrochimica Acta Part B,* vol. 65, p. 975–990, 2010.

[30] W. o. Knowledge, Thomson Reuters, 2014. [Online]. Available: http://apps.webofknowledge.com. [Accessed 24 March 2014].

[31] A. P. M. Michel, "Review: Application of single-shot laser-induced breakdown spectroscopy," *Spectrochimica Acta Part B,* vol. 65, pp. 185-191, 2010.

[32] F. Anabitarte, A. Cobo and J. M. Lopez-Higuera, "Laser-Induced Breakdown Spectroscopy: Fundamentals, Applications, and Challenges," *ISRN Spectroscopy,* pp. 1-12, 2012.

[33] R. E. Russo, X. Mao and S. S. Mao, "The physics of laser ablation in microchemical analysis," *Analytical chemistry,* vol. 74, pp. 70-77, 2002.

[34] I. B. Gornushkin and U. Panne, "Radiative models of laser-induced plasma and pump-probe diagnostics relevant to laser-induced breakdown spectroscopy," *Spectrochimica Acta Part B journal,* vol. 65, pp. 345-359, 2010.

[35] J. van Dijk, G. Kroesen and A. Bogaerts, "Plasma modelling and numerical simulation," *Journal of Physics D: Applied Physics,* vol. 42, pp. 190301-14, 2009.

[36] R. Noll, V. Sturm, Ü. Aydin, D. Eilers, C. Gehlen, M. Höhne, A. Lamott, J. Makowe and J. Vrenegor, "Laser-induced breakdown spectroscopy—From research to industry, new frontiers for process control," *Spectrochimica Acta Part B,* vol. 63, pp. 1159-1166, 2008.

[37] R. Noll, C. Fricke-Begemann, M. Brunk, S. Connemann, C. Meinhardt, M. Scharun, V. Sturm, J. Makowe and C. Gehlen, "Laser-induced breakdown spectroscopy expands into industrial applications," *Spectrochimica Acta Part B,* 2014.

[38] E. Tognoni, V. Palleschi, M. Corsi and G. Christoforetti, "Quantitative micro-analysis by laser-induced breakdown spectroscopy : a review of the experimental approaches," *Spectrochimica*

Acta Part B, vol. 57, pp. 1115-1130, 2002.

[39] F. C. De Lucia Jr. and J. L. Gottfried, “Rapid analysis of energetic and geo-materials using LIBS,” *Materials today,* vol. 14, pp. 274-281, 2011.

[40] R. S. Harmon, F. C. DeLucia, C. E. McManus, N. J. McMillan, T. F. Jenkins, M. E. Walsh and A. Miziolek, "Laser-induced breakdown spectroscopy – An emerging chemical sensor technology for real-time field-portable, geochemical, mineralogical, and environmental applications," *Applied Geochemistry,* vol. 21, pp. 730-747, 2006.

[41] R. Gaudiuso, M. Dell'aglio, O. De Pascale, G. S. Senesi and A. De Giacomo, "Laser induced breakdown spectroscopy for elemental analysis in environmental, cultural heritage and space applications: a review of methods and results.," *Sensors,* vol. 10, pp. 7434-7468, 2010.

[42] J. L. Gottfried, F. C. DeLucia, C. A. Munson and A. W. Miziolek, "Laser-induced breakdown spectroscopy for detection of explosives residues: a review of recent advances, challenges, and future prospects.," *Analytical and bioanalytical chemistry,* vol. 395, pp. 283-300, 2009.

[43] J. Kaiser, K. Novotný, M. Z. Martin, A. Hrdlička, R. Malina, M. Hartl, V. Adam and R. Kizek, "Trace elemental analysis by laser-induced breakdown spectroscopy — Biological applications," *Surface Science Reports,* vol. 67, pp. 233-243, 2012.

[44] P. Geladi, "Chemometrics in spectroscopy. Part 1. Classical chemometrics," *Spectrochimica Acta Part B,* vol. 58, pp. 767-782, 2003.

[45] V. I. Babushok, F. C. De Lucia, J. L. Gottfried, C. A. Munson and A. W. Miziolek, "Double pulse laser ablation and plasma: Laser induced breakdown spec- troscopy signal enhancement," *Spectrochimica Acta Part B,* vol. 61, pp. 999-1014, 2006.

[46] J. Scaffidi, S. M. Angel and D. A. Cremers, "Emission enhancement mechanisms in dual- pulse LIBS," *Analytical Chemistry,* vol. 78, pp. 24-32, 2006.

[47] B. Sallé, P. Mauchien and S. Maurice, "Laser-induced breakdown spectroscopy in open-path configuration for the analysis of distant objects," *Spectrochimica Acta Part B,* vol. 62, pp. 739-768, 2007.

[48] K. Y. Yamamoto, D. A. Cremers, M. J. Ferris and L. E. Foster, "Detection of metals in the environment using a portable laser-induced breakdown spectroscopy instrument," *Applied Spectroscopy,* vol. 50, pp. 222-233, 1996.

[49] B. C. Castle, A. K. Knight, K. Visser, B. W. Smith and J. D. Winefordner, "Battery powered laser-induced plasma spectrometer for elemental determinations," *Journal of Analitical Atomic Spectrometry,* vol. 13, pp. 589-595, 1998.

[50] C. M. Davies, H. H. Telle, D. J. Montgomery and R. E. Corbett, "Quantitative analysis using

remote laser-induced breakdown spectroscopy (LIBS)," *Spectrochimica Acta Part B,* vol. 50, pp. 1059-1075, 1995.

[51] D. A. Cremers, J. E. Barefield II and A. C. Koskelo, "Remote elemental analysis by laser-induced breakdown spectroscopy using a fiber-optic cable," *Applied Spectroscopy,* Vols. 857-860, p. 49, 1995.

[52] D. A. Cremers, "The analysis of metals at a distance using laser-induced breakdown spectroscopy," *Applied spectroscopy,* vol. 41, pp. 572-579, 1987.

[53] S. Maurice, R. C. Wiens, M. Saccoccio and et al., "The ChemCam Instrument Suite on theMars Science Laboratory (MSL) Rover: Science Objectives and Mast Unit Description," *Space Science Review,* vol. 170, pp. 95-166, 2012.

[54] G. B. Courréges-Lacoste, B. Ahlers and F. R. Pérez, "Combined Raman spectrometer/laser-induced breakdown spectrometer for the next ESA mission to Mars," *Spectrochimica Acta Part A,* vol. 68, p. 1023–1028, 2007.

[55] M. Born and E. Wolf, Principles of optics, Cambridge (UK): Cambridge University Press, 1999.

[56] J. R. Bettis, “Correlation among the laser-induced breakdown thresholds in solids, liquids, and gases,” *Applied Optics,* vol. 31, pp. 3448-3452, 1992.

[57] I. B. Gornushkin, A. Y. Kazakov, N. Omenetto, B. W. Smith and J. D. Winefordner, “Experimental verification of a radiative model of laser-induced plasma expanding into vacuum,” *Spectrochimica Acta Part B,* vol. 60, pp. 215-230, 2005.

[58] M. Capitelli, A. Casavola, G. Colonna and A. De Giacomo, “Laser-induced plasma expansion: theoretical and experimental aspects,” *Spectrochimica Acta Part B,* vol. 59, pp. 271-289, 2004.

[59] E. Tognoni, G. Cristoforetti, S. Legnaioli, V. Palleschi, A. Salvetti, M. Mueller, U. Panne and I. Gornushkin, “A numerical study of expected accurancy and precision in Calibration-Free Laser-Induced Breakdown Spectroscopy in the assumption of ideal analytical plasma,” *Spectrochimica Acta Part B,* vol. 62, pp. 1287-1302, 2007.

[60] H. R. Griem, Spectral line broadening by Plasmas, New York, USA: Academic Press, Inc., 1974.

[61] J. Bengoechea, C. Aragón and J. A. Aguilera, “Asymmetric Stark broadening of the Fe I 358.34 nm emission line in a laser induced plasma,” *Spectrochimica Acta Part B,* vol. 60, pp. 897-904, 2005.

[62] M. Capitelli, F. Capitelli and A. Eletskii, “Non-equilibrium and equilibrium problems in laser-

induced plasmas," *Spectrochimica Acta Part B,* vol. 55, pp. 559-574, 2000.

[63] J. D. Ingle Jr. and S. R. Crouch, Spectrochemical analysis, New Jersey, US: Prentice-Hall, Inc., 1988.

[64] V. Thomsen, D. Schatzlein and D. Mercuro, "Limits of detection in spectroscopy," *Spectroscopy,* vol. 18, pp. 112-114, 2003.

[65] H. E. Bauer, F. Leis and K. Niemax, "Laser induced breakdown spectrometry with an échelle spectrometer and intensified charge coupled device detection," *Spectrochimica Acta Part B,* vol. 53, pp. 1815-1825, 1998.

[66] D. W. Hahn, W. L. Flower and K. R. Hencken, "Discrete Particle Detection and Metal Emission Monitoring Using Laser-Induced Breakdown Spectroscopy," *Applied Spectroscopy,* vol. 51, pp. 1836-1844, 1997.

[67] R. E. Russo, X. L. Mao, O. V. Borisov and H. Liu, "Influence of wavelength on fraction in laser ablation ICP-MS," *J. Anal. At. Spectrom.,* vol. 15, pp. 1115-1120, 2000.

[68] B. Zhang, M. He and B. Huang, "Minimizing matrix effect by femtosecond laser ablation and ionization in elemental determination," *Analytical Chemistry,* vol. 85, pp. 4507-4511, 2013.

[69] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström and S. Wold, Multi- and megavariate data analysis. Part I: Basic principles and applications, Umeå, Sweden: Umetrics AB, 2006.

[70] H. Martens and T. Naes, Multivariate calibration, Chichester, UK: John Wiley & Sons Ltd, 1989.

[71] R. G. Brereton, Chemometrix for pattern recognition, Chichester, UK: John Wiley & Sons, Ltd, 2009.

[72] S. Wold, "Chemometrics; what do we mean with it, and what do we want from it?," *Chemometrics and Intelligent Laboratory Systems,* vol. 30, pp. 109-115 , 1995.

[73] N. J. McMillan, R. S. Harmon, F. C. DeLucia and A. W. Miziolek, "Laser-induced breakdown spectroscopy analysis of minerals: Carbonate and silicates," *Spectroschimica Acta Part B,* vol. 62, pp. 1528-1536, 2007.

[74] R. S. Harmon, J. Remus, N. J. McMillan, C. McManus, L. Colins, J. L. Gottfried, F. C. DeLucia and A. W. Miziolek, "LIBS analysis of geomaterials: geochemical fingerprinting for the rapid analysis and discrimination of minerals," *Applied Geochemistry,* vol. 24, pp. 1125-1141, 2009.

[75] R. T. Wainner, R. S. Harmon, A. W. Miziolek, K. L. McNesky and P. D. French, "Analysis of

environmental lead contamination: comprasion of LIBS field and laboratory instruments," *Spectrochimica Acta Part B,* vol. 56, pp. 777-793, 2001.

[76] J. L. Gottfried, R. S. Harmon, F. C. DeLucia Jr. and A. W. Miziolek, "Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification," *Spectrochimica Acta Part B,* vol. 64, pp. 1009-1019, 2008.

[77] R. S. Harmon, K. M. Shughue, J. J. Remus, M. A. Wise, L. J. East and R. R. Hark, "Can the provenance of the conflict minerals columbite and tantalite be ascertained by laser-induced breakdown spectroscopy?," *Anal Bioanal Chem,* vol. 400, pp. 3377-3382, 2011.

[78] B. Bousquet, J. B. Sirven and L. Canioni, "Towards quantitative laser-induced breakdown spectroscopy analysis of soil samples," *Spectrochimica Acta Part B,* vol. 62, pp. 1582-1589, 2007.

[79] A. Smilde, R. Bro and P. Geladi, Multi-way analysis: applications in the chemical sciences, Chichester, UK: John Wiley and Sons, Ltd., 2004.

[80] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine,* vol. 2, pp. 559-572, 1901.

[81] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology,* vol. 24, pp. 417-441, 1933.

[82] S. Wold, K. Esbensen and P. Geladi, "Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems,* vol. 2, pp. 37-52, 1987.

[83] S. D. Brown, "Chemical Systems Under Indirect Observation: Latent Properties and Chemometrics," *Applied Spectroscopy,* vol. 49, pp. 14A-31A, 1995.

[84] K. R. Beebe and B. R. Kowalski, "An Introduction to Multivariate Calibration and Analysis," *Analytical chemistry,* vol. 59, pp. 1007-1017, 1987.

[85] S. Wold, "Pattern recognition by means of disjoint principal components models," *Pattern Recognition,* vol. 8, pp. 127-139, 1976.

[86] K. Vanden Branden and M. Hubert, "Robust classification in high dimensions based on the SIMCA Method," *Chemometrics and Intelligent Laboratory Systems,* vol. 79, pp. 10-21, 2005.

[87] R. W. Gerlach, B. R. Kowalski and H. O. A. Wold, "Partial least-squares path modelling with latent variables," *Analytica Chimica Acta,* vol. 112, pp. 417-421, 1979.

[88] S. Wold, H. Martens and H. Wold, "The multivariate calibration problem in chemistry solved by the PLS method," in *Matrix Pencils*, Pite Havsbad, Sweden, Springer, 1983, pp. 286-293.

[89] M. Sjöström, S. Wold, W. Lindberg, J. Persson and H. Martens, “A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables,” *Analytica Chimica Acta,* vol. 150, pp. 61-70, 1983.

[90] S. Wold, M. Sjöström and L. Eriksson, “PLS-regression : a basic tool of chemometrics,” *Chemometrics and Intelligent Laboratory Systems,* vol. 58, pp. 109-130, 2001.

[91] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological Cybernetics,* vol. 69, pp. 59-69, 1982.

[92] T. Kohonen, “The self-organizing map,” *Neurocomputing,* vol. 21, pp. 1-6, 1988.

[93] J. Vesanto, *Data exploration process based on the self-organizing map,* Helsinki: Helsinki University of Technology, 2002, p. 96.

[94] F. C. De Lucia Jr., A. C. Samuels, R. S. Harmon, R. A. Walters, K. L. McNesby, A. LaPointe, R. J. Winkel Jr. and A. W. Miziolek, “Laser-induced breakdown spectroscopy (LIBS): a promising versatile chemical sensor technology for hazardous material detection,” *Sensors,* vol. 5, pp. 681-689, 2005.

[95] A. C. Samuels, F. C. De Lucia Jr., K. L. McNesby and A. W. Miziolek, “Laser-induced breakdown spectroscopy of bacterial pores, molds, pollens, and protein: initial studies of discrimination potential,” *Applied Optics,* vol. 42, pp. 6205-6209, 2003.

[96] E. G. Snyder, C. A. Munson, J. L. Gottfried, F. C. De Lucia Jr., B. Gullett and A. W. Miziolek, “Laser-induced breakdown spectroscopy for the classification of unknown powders,” *Applied Optics,* vol. 47, pp. G80-G87, 2008.

[97] J. L. Gottfried, F. C. De Lucia Jr., C. A. Munson and A. W. Miziolek, “Standoff Detection of Chemical and Biological Threats Using Laser-Induced Breakdown Spectroscopy,” *Applied Spectroscopy,* vol. 62, pp. 353-363, 2008.

[98] J. D. Hybl, G. G. Lithgow and S. G. Buckley, “Laser-Induced Breakdown Spectroscopy Detection and Classification of Biological Aerosols,” *Applied Spectroscopy,* vol. 57, pp. 1207-1215, 2003.

[99] S. J. Rehse, J. Dietrich and S. Palchaudhuri, “Identification and discrimination of Pseudomonas aeruginosa bacteria grown in blood and bile by laser-induced breakdown spectroscopy,” *Spectrochimica Acta Part B,* vol. 62, pp. 1169-1176, 2007.

[100] M. Z. Martin, N. Labbé, T. G. Rials and S. D. Wullschleger, “Analysis of preservative-treated wood by multivariate analysis of laser-induced breakdown spectroscopy spectra,” *Spectroschimica Acta Part B,* vol. 60, pp. 1179-1185, 2005.

[101] M. Z. Martin, S. Allman, D. J. Brice, R. C. Martin and N. O. Andre, “Exploring laser-induced

breakdown spectroscopy for nuclear materials analysis and in-situ applications," *Spectrochimica Acta Part B,* vol. 74, pp. 177-183, 2012.

[102] G. Vítková, K. Novotný, L. Prokeš, A. Hrdlička, J. Kaiser, J. Novotný, R. Malina and D. Prochazka, "Fast identification of biominerals by means of stand-off laser-induced breakdown spectroscopy using linear discriminant analysis and artificial neural networks," *Spectrochimica Acta Part B,* vol. 73, pp. 1-6, 2012.

[103] M. R. Martelli, F. Brygo, A. Sadoudi, P. Delaporte and C. Barron, "Laser-induced breakdown spectroscopy and chemometrics: a novel potential method to analyze wheat grains," *J. Agric Food Chem,* vol. 58, pp. 7126-7134, 2010.

[104] F. R. Doucet, T. F. Belliveau, J. L. Fortier and J. Hubert, "Use of Chemometrics and Laser-Induced Breakdown Spectroscopy for Quantitative Analysis of Major and Minor Elements in Aluminum Alloys," *Applied Spectroscopy,* vol. 61, pp. 327-332, 2007.

[105] R. A. Multari, D. A. Cremers, J. M. Dupre and J. E. Gustafson, "The Use of Laser-Induced Breakdown Spectroscopy for Distinguishing Between Bacterial Pathogen Species and Strains," *Applied Spectroscopy,* vol. 64, pp. 750-759, 2010.

[106] A. M. Ollila, J. Lasue, H. E. Newson, R. A. Multari, R. C. Wiens and S. M. Clegg, "Comparison of two partial least squares-discriminant analysis algorithms for identifying geological samples with the ChemCam laser-induced breakdown spectroscopy instrument.," *Applied Optics,* vol. 51, pp. 130-142, 2012.

[107] A. L. Streckeisen, "To each plutonic rock its proper name," *Earth Science Reviews,* vol. 12, pp. 1-33, 1976.

[108] K. Kiviluoto, J. Parviainen and M. Pollari, "Somtoolbox," Laboratory of Information and Computer Science, Helsinky University ov Technology, 23 March 2005. [Online]. Available: http://www.cis.hut.fi/projects/somtoolbox/. [Accessed 26 May 2014].

[109] A. L. Streckeisen, "Classification and nomenclature of plutonic rocks," *Geologische Rundschau,* vol. 63, pp. 773-785, 1974.

[110] A. L. Streckeisen, "Classification and nomenclature of volcanic rocks," *Neues Jahrbuch für Mineralogie,* vol. 141, pp. 1-14, 1978.

[111] M. J. Le Bas, R. W. Le Maitre, A. Streckeisen and B. Zanettin, "A Chemical Classification of Volcanic Rocks Based on the Total Alkali-Silica Diagram," *Journal of Petrology,* vol. 27, pp. 745-750, 1986.

[112] D. A. Rusak, M. Clara, E. E. Austin, K. Visser, R. Niessner, B. W. Smith and J. D. Winefordner, "Investigation of the Effect of Target Water Content on a Laser-Induced Plasma,"

Applied Spectroscopy, vol. 51, pp. 1628-1631, 1997.

[113] I. B. Gornushkin, A. Ruiz-Medina, J. M. Anzano, B. W. Smith and J. D. Winefordner, “Identification of particulate minerals by correlation analysis using a microscopic laser induced breakdown spectrometer,” *J. Anal. At. Spectrom,* vol. 15, pp. 581-586, 2000.

[114] S. Laville, M. Sabsabi and F. Doucet, “Multi-elemental analysis of solidified mineral melt samples by Laser-Induced Breakdown Spectroscopy coupled with a linear multivariate calibration,” *Spectrochimica Acta Part B,* vol. 62, pp. 1557-1566, 2007.

[115] D. Death, A. Cunningham and L. Pollard, “Multi-element and mineralogical analysis of mineral ores using laser induced breakdown spectroscopy and chemometric analysis,” *Spectrochimica Acta Part B,* vol. 64, pp. 1048-1058, 2009.

[116] ORE, “certified reference materials,” 18 Sep 2013. [Online]. Available: http://www.ore.com.au/. [Accessed 9 Oct 2013].

APPENDIX

Manuscript attached to this thesis reflects the recent work on the application of the LIBS method to the field of mining industry. This manuscript entitled: 'Laser-induced breakdown spectroscopy for *in-situ* qualitative and quantitative analysis of mineral ores' was submitted to thematic issue of Spectrochimica Acta Part B dedicated to the EMSLIBS conference held in Bari, Italy in September 2013 (ref. no. NO-EMSLIBS2013-25).

This work brings a comprehensive study on the classification and quantitative analysis of igneous rocks. As in the aforementioned case studies, the LIBS measurement is strongly affected by a matrix effect. Normalization of the calibration curve and multivariate regression algorithms (PCR and PLSR) do not lead to the improvement in the linearity of the formerly constructed calibration curve. Therefore, it is proposed to classify the samples with respect to the composition in the matrix elements of relevant samples. PCA was successfully used for the purposes of sample classification. Then, partial calibration curves were constructed for each individual cluster of samples. Such data pretreatment improved the precision of quantitative analysis utilizing partial calibration curves in comparison to quantitative analysis based on calibration curves constructed without any sample/data classification. The precision of quantitative analysis was estimated by the means of bias, which is an analytical measure evaluating the closeness of measured/predicted result and related certified value. Measurement of the whole data set (and consecutive data analysis) was utilized on two LIBS systems in two different laboratories. Thus, the classification of the samples leads to the same class membership. Moreover, the precision of the quantitative analysis of both data sets (system 1 and system 2) are comparable in the sense of the bias. This strengthens the reproducibility of the LIBS measurement done on various LIBS systems.

Laser-Induced Breakdown Spectroscopy for *in-situ* Qualitative and Quantitative Analysis of Mineral Ores

P. Pořízka[1,2], A. Demidov[1], J. Kaiser[2], J. Keivanian[3], I. Gornushkin[1], U. Panne[1,4], J. Riedel[1]*

[1]BAM, Federal Institute for Materials Research and Testing, Richard Willstätter-Straße 11, D-12489 Berlin, Germany.
[2]Institute of Physical Engineering, Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896/2, 61669 Brno, Czech Republic.
[3]Technical University Clausthal, Institute for Mining, Erzstraße 18, 38678 Clausthal-Zellerfeld, Germany.
[4]Humboldt Univerisät zu Berlin, Chemistry Department, Brook-Taylor-Straße 2, D-12489 Berlin, Germany.
**corresponding author: Dr. Jens Riedel, jens.riedel@bam.de*

Abstract

In this work, the potential of laser-induced breakdown spectroscopy (LIBS) for discrimination and analysis of geological materials was examined. The research was focused on preliminary classification of mineral ores using their LIBS spectra prior to quantitative determination of copper. 27 igneous rock samples were analyzed in the form of fine dust pressed onto the surface of a double-sided sticky tape. Two different LIBS setups in two laboratories were used to prove the reproducibility of classification and quantitative analysis. Quantitative analysis is not a trivial task in LIBS measurement because intensities of emission lines in laser-induced plasmas are strongly affected by the mineral matrix, i.e. the matrix effect. The principal component analysis (PCA) was applied on the measured data set to discriminate individual rocks according to their chemical composition. Partial calibration plots constructed from the clustered data displayed an improvement over the calibration plot constructed from all ore samples. The classification of mineral samples with complex matrices can thus be recommended prior to calibration analysis.

Key words: Laser-induced breakdown spectroscopy; LIBS; Chemometrics; Principal Component Analysis; geochemical analysis

Introduction

With steadily increasing consumption of mineral resources, it is crucial to optimize the exploration of new ore deposits and the mining process itself with reliable and rapid analytical methods for identification of minerals. However, reliable *in-situ* identification of minerals can be very challenging. Frequently used in-field identification of minerals is based on visual examination of their physical properties by an experienced geologist, which, however, can result in false positives [1]. More advanced laboratory-based techniques, such as X-Ray fluorescence (XRF) and inductively coupled plasma atomic emission or mass spectrometry (ICP-AES/MS) after wet digestion of samples [2] are not always suitable due to their long turn-around time and analysis cost per sample.

Laser-induced breakdown spectroscopy (LIBS) can meet the challenge as a field technique for identification and analysis of various minerals. LIBS became a popular technique due to its fast and non-demanding measurement routine, minor need for sample preparation, and low-cost instrumentation in comparison to other atomic emission spectroscopic techniques.

LIBS provides qualitative and quantitative information about samples under investigation in *real-time* and *in-situ* with an inherent multi-elemental capability [3, 4]. The observed LIBS spectrum reflects the chemical composition, i.e. chemical fingerprint, of the sample.

The composition of mineral ores differs with their provenance as well as with the way of their alterations [5]. Alteration is the property of a rock which explains its chemical and mineralogical changes in the course of time. In geology, the alteration is important because it may have an effect on grades of elements (e.g. copper), therefore the rocks are studied in individual alteration types. Different mineral ores and their alterations can be discriminated by examining their chemical fingerprints, especially the composition of matrix elements (e.g. Al, Ca, K, Na, and Si) [6]. Mineral ores can be classified into groups by forming the QAPF (Quartz, Alkali feldspar, Plagioclase, Feldspathoid) diagram [5]. Different areas in this diagram reflect natural relationships between rock types; each rock type displays a certain continuous variation of a mineral content [5, 7]. The exact classification of mineral types does not rely on the knowledge of concentrations of only one or two characteristic elements but has to be derived using the complete elemental fingerprint. The established method for classification of such complex and large datasets is the principal components analysis (PCA). It is expected that the PCA discrimination of samples represented by their LIBS spectra can emulate the distribution of the rock types in the QAPF diagram.

Earlier applications of LIBS in geology were comprehensively reviewed by Harmon *et al.* [8]. LIBS has already been used for quantitative analysis of mineral samples including field-portable devices [1, 9, 10, 11]. Multivariate statistical approaches for identification of different kinds of rocks and minerals were employed using both laboratory bench-top and stand-off LIBS systems [12]. Harmon *et al.* [13] demonstrated the possibility of ascertaining the provenance of conflict minerals employing the partial least squares discriminant analysis (PLS-DA) of LIBS spectra. Bousquet *et al.* [14] tested the hypothesis that the most significant differences between soils come from the varying amounts of matrix elements. Application of PCA to LIBS spectra can be found in [15, 16, 17, 18].

In this work, the potential of LIBS for determination of copper in mineral ores is investigated based on preliminary identification of ores according to their chemical compositions. This pre-discrimination into individual classes of minerals allows the reduction of the strong influence of sample compositions on intensities of copper emission lines induced by the mineral matrix. The principal component analysis (PCA) is used for unsupervised classification of the mineral ores into different groups prior to the quantitative determination of the copper content. Partial calibration plots are constructed from these groups of samples and compared to the calibration plot constructed from all certified ore samples simultaneously.

Experimental

Samples were collected at copper mines located in the north-west region of Iran, near to Sungun area. The sample set consists of 27 samples of three igneous rock types (andesite – ANS, diorite – DIO, and monzonite to quartz monzonite – KP) and alterations (table 1). The samples were received from the Clausthal University of Technology, Clausthal, Germany where they were analyzed by ICP-MS after four acid digestions. The amount of copper ranged between 0.06 and 0.79 weight percent. Apart from the chemical analysis, the samples were classified into groups and arranged in the QAPF diagram by an experienced geologist who tested samples by the visual inspection. The samples were analyzed with two different LIBS systems to ensure the independence of the classification results upon the individual LIBS system and, thus, to prove the applicability of a selected multivariate technique for unambiguous sample classification. Samples were measured in the form of a fine dust pressed onto the surface of the double-sided sticky tape; this approach was earlier suggested by Gornushkin *et al.* [19, 20]. Samples are placed on a motorized translational stage to provide a fresh spot for every laser pulse.

System 1 consists of a high energy Nd:YAG laser (Continuum Spitlight-10, 10 Hz, 1064 nm, 10 ns) focused onto the sample with an irradiance of 30 GW.cm^{-2} (the spot diameter is ≈ 250 μm) using a 100 mm focal length plano-convex lens. The radiation from the luminous plasma is collected using a large aperture collector-collimator (Andor CC52, f/2) placed 250 mm the interaction region at a 30° angle with respect to the laser axis. The collector is coupled to a 400 μm optical fiber which delivers light at the entrance slit of an echelle spectrometer (LTB Aryelle Butterfly). The resolving power and spectral range of the spectrometer are 15000 and 300 - 600 nm, respectively. The spectral information is recorded by an intensified CCD (Andor iStar 734, 1024 x1024 pixels with an effective pixel size of 13 x13 μm). The ICCD operates at 120x gain and 2 x 2 pixel binnings. The timing of the experiment was previously optimized and set to the 2 μs gate delay and 10 μs gate width. The whole LIBS setup is triggered by a delay generator (DG535, Stanford Research Systems).

The complementary LIBS system 2 consists of a diode pumped solid state (DPSS) laser (Quantel Ultra 100) operated at a fundamental wavelength 1064 nm and 5 Hz repetition rate. A short laser pulse (8 ns) is focused using a 450 mm focal length plano-convex lens creating a spot diameter ~ 200 μm on the target surface with the irradiance 20 GW cm^{-2}. The plasma radiation is collected top-on (collinearly with the laser beam) by a toroidal mirror of the 300 mm focal length and f/7. The radiation is detected by a CCD (Andor Newton, 1024 x 256 pixels with the effective pixel size 26 x 26 μm) attached to an echelle spectrometer (LTB Arryele 400) with the working range of 200 - 600 nm, resolving power 15000, and f/10. The spectrometer is equipped with an optomechanical chopper to cut off the plasma continuum radiation during first 0.3 μs of the plasma evolution, the gate width is set to 400 μs. Timing of the LIBS system is mastered by the chopper and triggered by the delay generator (DG535).

With system 1, one spectrum consists of 10 accumulations, 20 spectra per sample; giving overall 200 laser pulses per a sample. A similar measurement routine is employed with system 2, i.e. 20 accumulations and 10 spectra per a sample, giving also 200 pulses per a sample. 200 pulses per a sample are chosen for better statistics and smoothing effects of sample inhomogeneity.

The first step in data analysis was a construction of the calibration curve for copper. The spectral lines listed in table 2 were background-subtracted and fitted by the pseudo-Voigt profile; an analytical signal was the line integral. The custom software was written in MATLAB (version R2012a). Lines of matrix elements, table 2, were alternatively employed to normalize the signal of analytical lines and to linearize calibration curves.

During the second step, the classification by PCA and multivariate regression by PLSR and PCR were performed on the preprocessed data using the customized LIBRA software (KU Leuven, Belgium) [21]. The Z-test was carried out for spectra of each sample to remove possible outliers after which the spectra were averaged. The Z-test was utilized to ensure that no outlier spectra affected by local sample inhomogeneity are used in further data analysis. The averaged spectra were normalized to their integral intensities and mean-centered.

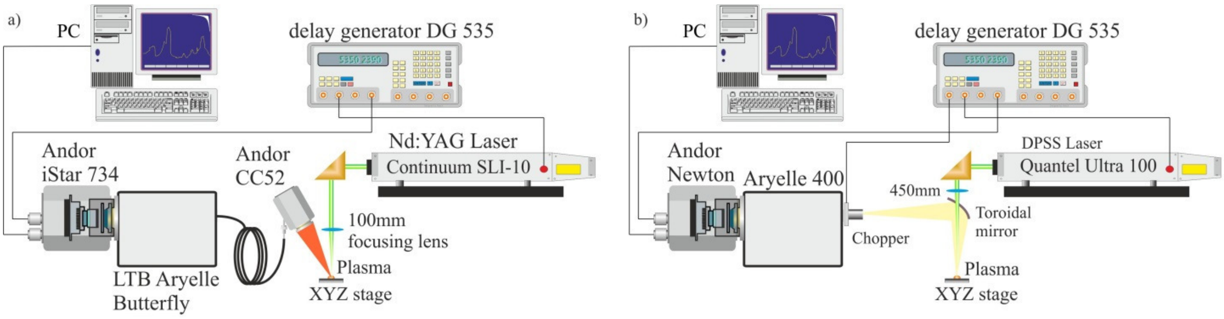


Figure 1, Experimental LIBS a) setup 1, b) setup 2.

Multivariate analysis

Principal component analysis (PCA) is commonly used for classification of LIBS data [12, 14, 18, 22]. The PCA is the mathematical coordinate transformation method that emphasizes the variance in the data set and reduces redundancy [23, 24, 25]. A data matrix ***X*** consists of individual spectra arranged in rows with the columns denoting the wavelengths. The PCA decomposes the data matrix into matrices of scores and loadings. The scores are analogous to coordinates of the data points (sample spectra) in a newly constructed PC space and the loadings are the transformation coefficients that are indicative of variances at specific wavelengths. It is expected that samples of the same rock type will form one cluster, based on the similarity among their spectra. The distribution of points in the PC space should correlate with the distribution in the QAPF diagram.

Principal components are conventionally used for multivariate quantification in a principal component regression (PCR) algorithm [25]. A PCA algorithm is applied to the data matrix ***X*** to truncate redundant information and to form a new data matrix. Then the regression vector ***y*** representing the content of an analyte in samples is regressed against the PC scores. Alternatively, PCA is used to model both the data and the regression vector (or matrix when more variables are considered) simultaneously; such an algorithm is called the partial least squares regression (PLSR) [25].

Results and discussion

No reliable calibration curve could be constructed from the complete data set (consisting of 27 igneous rocks) due to significant matrix effect and weak overall correlation between intensities and concentrations. Therefore we attempted the construction of partial calibration curves for each individual rock type. First, all samples were divided into three groups according to the class membership (ANS, DIO, and KP, see Table 1) determined by a geologist. Second, the system 1 and system 2 data set, respectively, was classified by the means of PCA based on the similarities in the sample matrices, *i.e.* similarities in the composition of matrix elements. Then partial calibration curves for each individual cluster were constructed. The procedure was applied to the data obtained with both LIBS systems in order to assess the independence of measurements on a particular instrument.

As stated above, the presence of lines of matrix elements Al, Ca, K, Na, and Si in LIBS spectra is important for reliable classification. Those matrix elements are conventionally used in geochemistry to fully discriminate igneous rocks [5]. 15 spectral lines of matrix elements were selected, table 2, for multivariate analysis. The experimental system 1 does not cover a spectral range of 200 - 300 nm, where several silicon lines are located; only Si 390.55 nm line was used to account for Si concentration. It was therefore expected that system 2 covering both the UV and VIS spectral ranges would yield the more reliable classification.

Table 1, list of samples.

Sample No	Rock Type		Alteration	Cu (%)
ANS1	Andesite	**ANS**	Phyllic	0.64
ANS2	Andesite	**ANS**	Phyllic	0.49
ANS3	Andesite	**ANS**	Phyllic	0.60
ANS4	Andesite	**ANS**	Phyllic	0.61
ANS5	Andesite	**ANS**	Phyllic	0.79
ANS6	Andesite	**ANS**	Phyllic	0.21
ANS7	Andesite	**ANS**	Phyllic	0.52
DIO1	Diorite	**DIO**	Potassic	0.40
DIO2	Diorite	**DIO**	Potassic	0.71
DIO3	Diorite	**DIO**	Potassic	0.73
DIO4	Diorite	**DIO**	Potassic	0.78
DIO5	Diorite	**DIO**	Potassic	0.32
DIO6	Diorite	**DIO**	Potassic	0.31
DIO7	Diorite	**DIO**	Potassic	0.30
DIO8	Diorite	**DIO**	Potassic	0.37
DIO9	Diorite	**DIO**	Potassic	0.73
KP1	Monzonite-Quartz Monzonite	**KP**	Potassic	0.24
KP2	Monzonite-Quartz Monzonite	**KP**	Potassic	0.18
KP3	Monzonite-Quartz Monzonite	**KP**	Potassic	0.21
KP4	Monzonite-Quartz Monzonite	**KP**	Phyllic	0.06
KP5	Monzonite-Quartz Monzonite	**KP**	Phyllic	0.51
KP6	Monzonite-Quartz Monzonite	**KP**	Phyllic	0.34
KP7	Monzonite-Quartz Monzonite	**KP**	Phyllic	0.45
KP8	Monzonite-Quartz Monzonite	**KP**	Potassic	0.48
KP9	Monzonite-Quartz Monzonite	**KP**	Potassic	0.71
KP10	Monzonite-Quartz Monzonite	**KP**	Potassic	0.56
KP11	Monzonite-Quartz Monzonite	**KP**	Potassic	0.30

Table 2, list of elemental lines selected for analysis.

Elemental line	**λ** nm	**E_i** eV	**E_i** eV	**Einstein coeff.** $10^8\ s^{-1}$
Al (I)	308.22	0	4.02	0.63
Al (I)	309.27	0.014	4.02	0.75
Al (I)	394.4	0	3.14	0.51
Al (I)	396.12	0.014	3.14	1.01
Ca (II)	393.37	0	3.15	1.47
Ca (II)	396.85	0	3.12	1.44
Ca (I)	422.7	0	2.93	2.18
Ca (I)	558.87	2.53	4.74	0.49
Na (I)	589	0	2.104	0.62
Na (I)	589.59	0	2.102	0.61
Si (I)*	251.43	0	4.93	0.74
Si (I)*	251.61	0.03	4.95	1.68
Si (I)*	251.92	0.01	4.93	0.55
Si (I)*	288.16	0.78	5.08	2.17
Si (I)	390.55	1.91	5.08	0.13
Cu (I)	324.75	0	3.82	1.37
Cu (I)	327.39	0	3.79	1.36
Cu (I)	515.32	3.79	6.19	0.6
Cu (I)	521.82	3.82	6.19	1.22

*Elemental lines used only when data from system 2 were utilized for multivariate analysis and linearization.

Calibration without PCA classification

The calibration curves for copper are constructed using the Cu (I) line at 521.8 nm for both LIBS systems data. This line is unlikely to get self-absorbed because of the low concentration range of Cu (<1 %) in analyzed samples and because of its high upper level of 6.19 eV. Prior the analysis, spectra of each sample were firstly normalized to spectra total intensities and then averaged to obtain one spectrum per a sample. Spectral lines selected for uni- and multivariate analysis were fitted with pseudo-Voigt profile and the area under such fitted line was adopted as the line intensity.

It is proposed to use individual calibration curves for analysis of individual rock types. Figure 2 shows the combined (for all rock types) and partial (for each rock type) calibration plots The linear regression performed on the whole data set (the combined plot) yields the coefficient of determination $R^2 = 0.49$ for measurements with system 1 and $R^2 = 0.23$ for measurements with system 2. Such low R^2's indicate high uncertainty and low accuracy of the calibration.

In contrary, the partial calibration plots (according to the rock type), although showing different slopes, reveal higher values for coefficients of determination, between 0.57 and 0.69 and between 0.68 and 0.87 for systems 1 and 2, correspondingly. Partial calibration curves for DIO and KP rock types result only in a moderate improvement of the coefficient of determination, i.e. $R^2 \sim 0.6$, for both LIBS systems. The coefficient of determination for ANS rock type is higher with system 2, $R^2 = 0.87$, than with system 1, $R^2 = 0.57$. Even though the partitioning of calibration curves in accordance with rock types led to an improved linear regression, the accuracy of analysis with such low R^2 partial calibration curves is expected to be moderate.

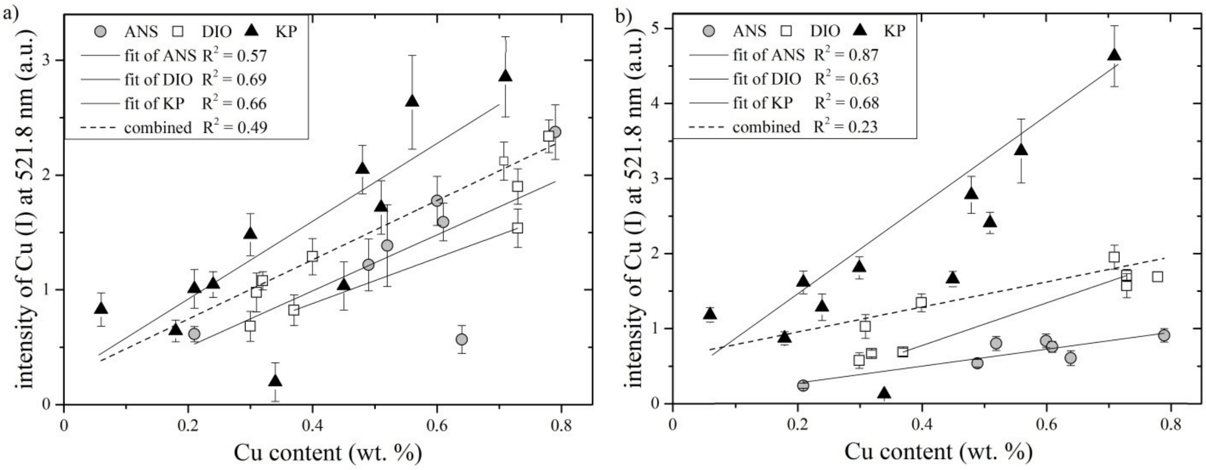


Figure 2, calibration curves for selected copper line 521.8 nm a) system 1 data set, b) system 2 data set. The dashed line refers to calibration plot for complete sample set (assigned as 'all' in the legend of the plot).

Conventionally, a further step to improve the quality (*i.e.* linearity) of the calibration is to normalize the intensity of an analytical line to the intensity of a matrix element line. With our current samples, the selection of an appropriate line of the matrix element was not an easy task due to the varying concentrations of these elements in the samples. Using lines of matrix elements Al, Ca, Na, and Si (table 2) to normalize the Cu (I) 521.8 nm line was not successful. The coefficients of determination for such linearized calibration curves were maximum at a $R^2 \sim 0.5$ level. In view of a moderate improvement of the calibration curve linearity, this approach of calibration curve linearization was abandoned in further analyses. Similar results were observed by Laville *et al.* [26] who analyzed mineral samples and tried to linearize calibration curves for Al, Ca, Mg, Fe and Ti by rationing the line intensities of these elements by the line intensity of Si. . The significant variation of $Si0_2$ (from 37.2 to 62.6 wt. %) in the samples was assumed to be the cause of the unsuccessful linearization.

The preliminary results of the multivariate quantitative analysis of the whole data set using the partial least squares regression (PLSR) and the principal components regression (PCR) are shown in figure 3. The data matrix $\boldsymbol{X}$ was constructed from all copper lines listed in table 3 plus selected lines of matrix elements listed in table 2. Spectra of each sample were firstly normalized to the sum of their total intensity and then averaged to obtain one spectrum per sample. The spectra were organized as rows and individual variables, matrix element lines from table 2, were assigned to columns of the data matrix. Each column was then mean-centered and normalized to unity in order to equalize the contribution of all variables to the estimation of principal components (PCs). The vector $\boldsymbol{y}$ for regression consists of certified copper contents, table 1. The whole data set (all samples) was used simultaneously. Copper lines were not expected to suffer from self-absorption in the limited range of Cu concentrations in the samples (from 0.06 wt. % to 0.79 wt. %). Therefore, the variations of interest were assumed to be linear and were analyzed by the linear multivariate regressions, *i.e.* PCR and PLSR [26]. In every step of the multivariate regression all spectra of one sample were left out from the PCR and PLSR model estimation and then the content of the analyte (Cu) in the left-out sample was predicted using the model built from the remaining samples , *i.e.* leave-one-sample-out (LOSO) prediction. It is noteworthy to mention the work of Death *et al.* [27]. In their study they proposed to introduce matrix lines into the computation of latent variables to enhance intergroup dissimilarities.

PCR and PLSR regressions were then applied for data sets obtained with both LIBS systems. To avoid under- or overfitting the model, 10 PCs were gradually added to the model and related mean square errors of prediction (MSEP) were estimated. The optimal number of PCs used for each particular model is given in the box in figure 3. In general, a lower number of PCs is needed in PLSR for its essential property to simultaneously model both the data matrix $\boldsymbol{X}$ and the regression vector $\boldsymbol{y}$. However, it is rather ambiguous to decide which multivariate regression algorithm is more parsimonious to the data. Moreover less PCs was used in the regression models of system 2 data. This may be a consequence of more variables reflecting the Si content in the samples. In both cases, PLSR provided higher values for R^2, namely $R^2 = 0.76$ for system 1 and $R^2 = 0.79$ for system 2 respectively, compared to $R^2 = 0.66$ and $R^2 = 0.7$ obtained with PCR. Nevertheless, such low coefficients of determination could not guarantee a reliable quantitative analysis. Thus, the multivariate calibration using the whole dataset must be recognized unsuccessful owing to the significant variation in the composition of sample matrices as well as the narrow range of copper concentrations.

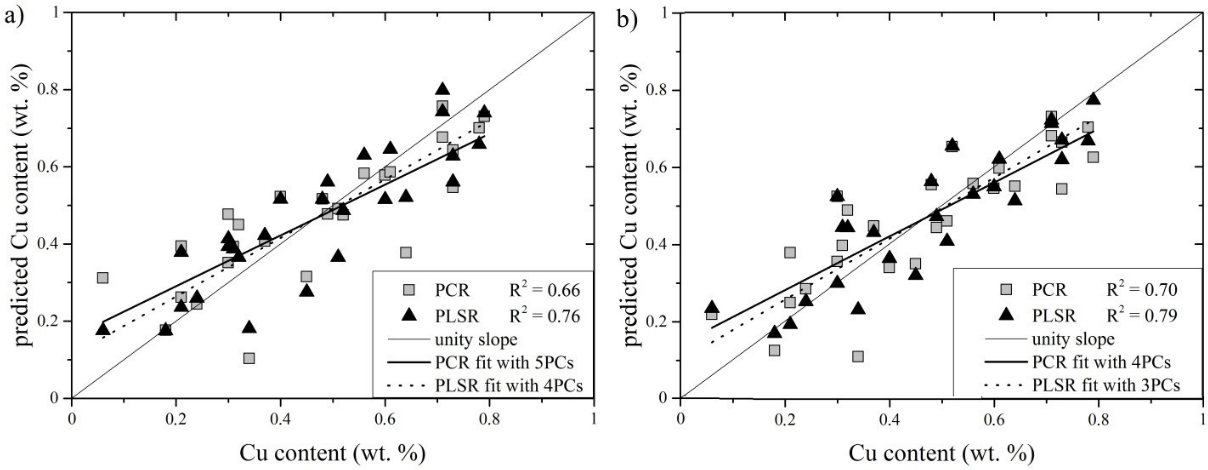


Figure 3, quantification of Cu content using PCR and PLSR for a) system 1 data, b) system 2 data.

Calibration with PCA classification

To improve the linearity and related accuracy/bias of quantitative analysis we attempted the classification of the samples based on the similarities in their matrix compositions. The data matrix was constructed from the lines of matrix elements, listed in table 2, and pretreated in the same manner as in the foregoing regression case. Though, 20 spectra of each sample obtained from system 1 were averaged to give 5 averaged spectra/data points per sample. Respectively, spectra from system 2 were averaged to give the same number of 5 averaged data points per sample. As stated before, the samples were classified by a geologist who inspected their physical properties. This approach relies entirely on the experience and qualification of the geologist; in some difficult cases it might yield erroneous results. It is highly desirable to have a more robust and human-independent classification technique, for example, PCA. The technique can be utilized for the detection of outliers (in this case false positives estimated by a geologist) and data clustering.

The results of the PCA analysis are shown in figure 4 andfigure 5 as the 2D projections of the data scores on the planes of first three principal components taken pair-wisely. The first three PCs covered 97.1 % (69.5 % + 23.2 % + 4.4 %) of the total variance in the data obtained with system 1 and 97.5 % (60.1 % + 32.9 % + 4.5 %) in the data obtained with system 2. The Gaussian clustering algorithm, built in the MATLAB Statistics Toolbox, was used to discriminate individual groups of samples according to their distributions in the PC space. Three PCs were used to construct a space to which clustering algorithm was applied. First, a selected number of ellipsoids (*e.g.* equal to the number of groups given by a geologist) is randomly placed into the PC space. Then those ellipsoids adapt their position and size with each iteration (in this case total number of 1000 iterations was utilized) to fit the distribution of the data points in the PC space. The ellipses in the projection plot mark the areas that cover 99% of the volume of the particular Gaussian sub-distributions. Some clusters can overlap for samples with similar compositions and further analysis is needed. One should note that the Gaussian clustering was used only for enclosing the points into separate groups. It is worth mentioning that data points do not have to necessarily obey the Gauss distribution in the PC space.

When PCA is applied to the system 1 data set, three distinct outliers are observed, see figure 4. Those outliers (ANS1, ANS7, and DIO1) are enclosed within one cluster and may be discarded from further classification while the composition of their matrices does not correspond to any other cluster in the data set. We may assume that such samples were wrongly assigned by the geologist. Then the rest of the data set is clustered into two distinct groups. The higher content of Si and Al in the KP rock type is responsible for its separation form the cluster of ANS and DIO rock types, as indicate the loadings plot in figure 6a. In this figure, loadings of particular PCs are crossplotted against each other to reveal the dependency of the PCA model on matrix elements. In parallel analysis, PCA was applied to the system 2 data set and resulted in similar distribution of the data in newly constructed PC space, figure 5, as in the foregoing case. However, in the system 2 case the outliers (ANS1, ANS7, and DIO1) are more distant from the rest of the data set. Except for several significantly outlying samples, rock types ANS, DIO and KP are distinctly separated. This is most probably the consequence of more variables (Si lines from the 200 to 300 nm spectral region) introduced to the computation of PCs. Furthermore, the KP rock type is separated into two clusters according to the alteration of the soil samples. Therefore it can be concluded that one Si line (390.55 nm) may not be sufficient for the direct classification of samples in the system 1 dataset. Nevertheless, the loadings plot in figure 6a reveals similar grouping of matrix lines of Al, Ca and Si that are distributed along the first principal component as in the system 2 case depicted in figure 6b.

It is seen from the given results that the data tend to cluster according to spectral lines of matrix elements and a consequent classification may be possible. In other words, spectral lines of matrix elements determine clustering patterns in subspaces of principal components. The main reason to use only the lines of matrix elements for classification is the reduction of datasets. The reduction of data matrices additionally accelerates PCA calculations that can be helpful for *on-line* analysis of data during the mining process. As suggested by Death *et al.* [27], data points of individual samples are distributed in more distinct groups in the PC space while only the important information is used, *i.e.* lines of matrix elements. Moreover, the PCA results for data sets from systems 1 and 2 are comparable. Those results support the reliability and reproducibility of measurements with different LIBS setups.

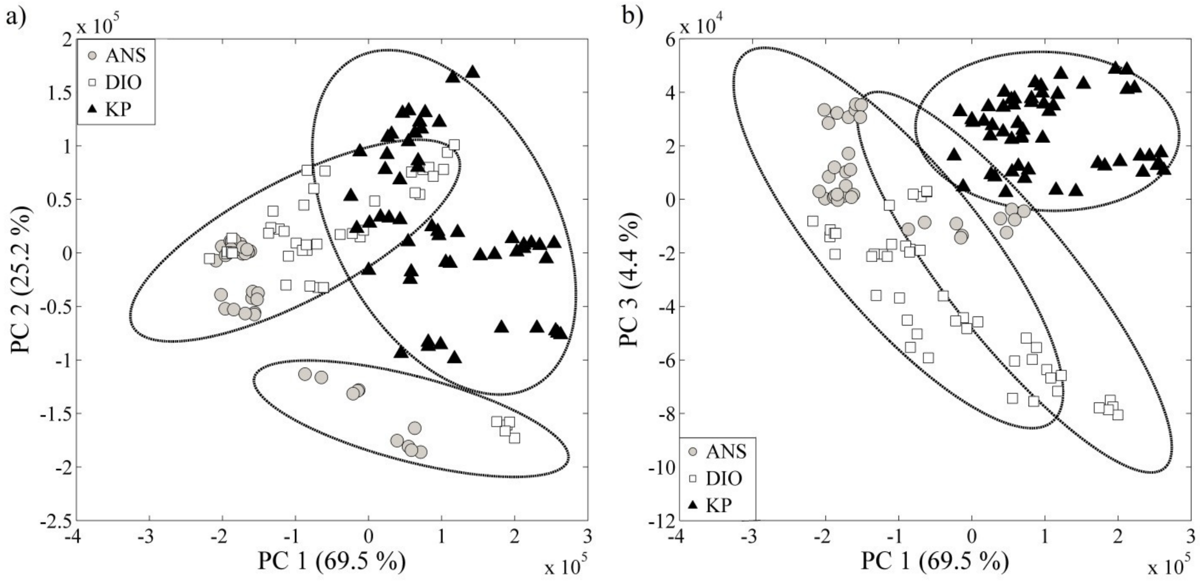


Figure 4, PCA applied on the system 1 data, a) first and second PC scores, b) first and third PC scores.

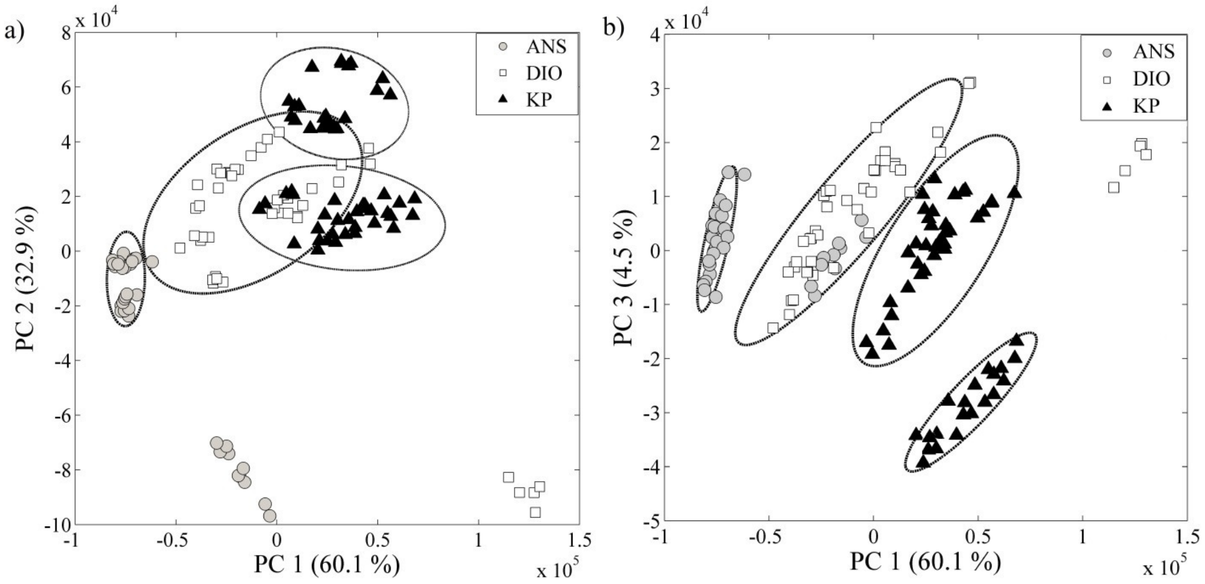


Figure 5, PCA applied on the system 2 data, a) first and second PC scores, b) first and third PC scores.

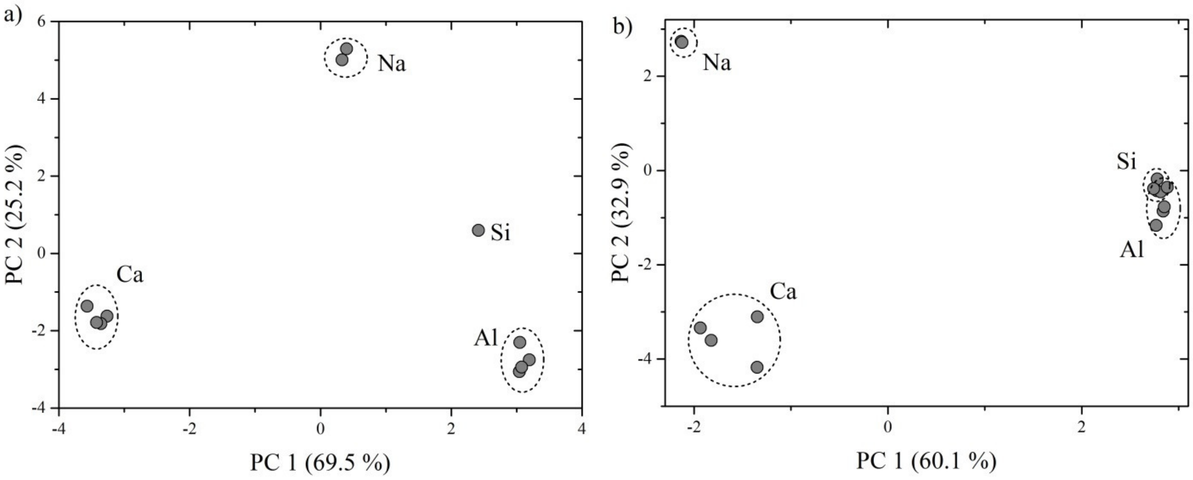


Figure 6, first two PC loadings a) system 1 data, b) system 2 data.

Though, the data set of system 1 was not completely separated as the one of system 2, comparing the discrimination of the data points in figure 4andFigure 5. Therefore we used the approach for data classification suggested by Multari *et al.* [28]. They suggested discarding well discriminated (tightly clustered) data points, *i.e.* samples/measurements, from the data matrix and then applying the PCA and consequent clustering algorithms again on the reduced data matrix. This step in the data analysis should emphasize the differences between groups which were formerly interfered. This procedure could be repeated until only well discriminated groups remain in the PC space. Based on the algorithm suggested by Multari *et al.* [28], figure 7a shows the clustering of the KP rock type which was well discriminated from the rest of the data in previous classification. Therefore we applied PCA to this rock type solely. Data points were then clustered according to the relevant rock alteration. Similarly, the PCA was applied to the reduced data matrices of ANS and DIO rock types (when the outliers were discarded). Figure 7b depicts two clusters which are distinguished from each other, *i.e.* enveloped with their own ellipses. The clusters are projected onto the first two PCs and are oriented in direction in PC space. Furthermore, DIO rock type was treated with the PCA once more to reveal two more outliers (DIO8 and DIO9), which not shown in figure 7.

Thus, the PCA helps in identification of outliers among the whole dataset. The outliers (ANS1, ANS7, DIO1, DIO8, and DIO9) were discarded from further quantitative analysis. After this, the partial calibration curves were re-plotted and are shown in figure 8. According to the classification of rocks in geochemistry, the rock type and alteration should correspond to each other. Therefore, a calibration curve was constructed only for the potassic alteration of the KP rock type while the phyllic alteration was excluded from the calibration. The difference in the composition was as well revealed in the PC space of both data set (system 1 and system2), figure 5 and figure 7a.

Significant improvements in the linearity of the partial calibration curves are obtained for all the rock types (ANS, DIO, and KP) as can be found in the comparison of figure 2 and figure 8. As a consequence of such data pretreatment the matrix effect may be avoided to such extent that it does not significantly affect the accuracy/bias of quantification using partial calibration curves.

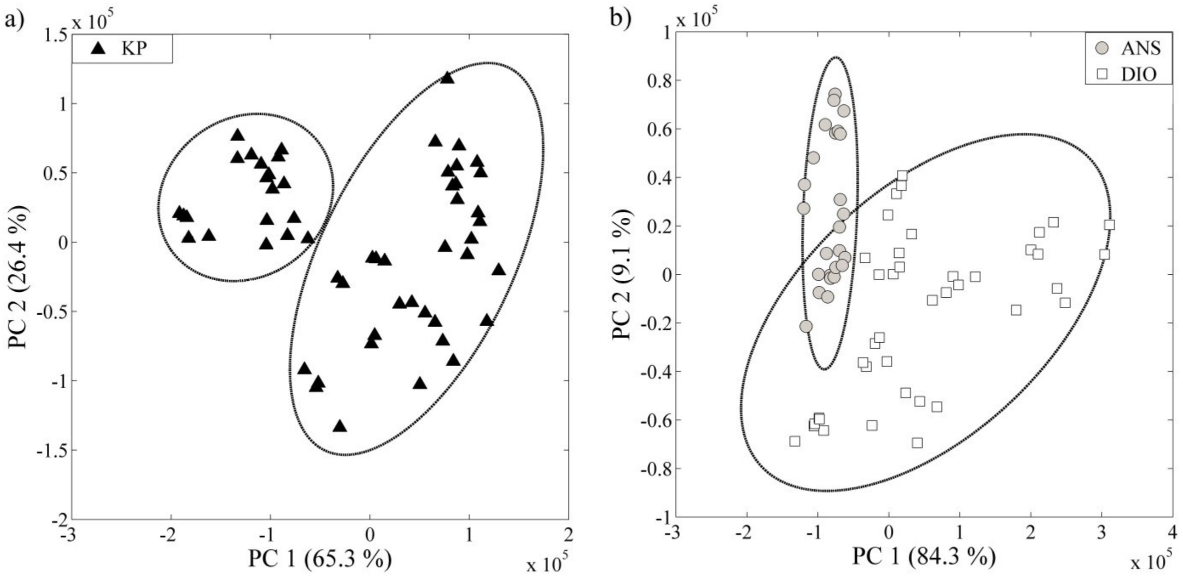


Figure 7, further classification of the system 1 data set a) classification of KP alterations, b) discrimination of ANS and DIO.

The improvement in prediction of copper concentrations using partial calibration curves was additionally proved by the leave-one-sample-out (LOSO) method described above. Table 3 provides the comparative results of LOSO using combined and partial calibration curves. The reference (certified) values of Cu concentrations were obtained from the ICP-MS measurements. The precision of the prediction can be estimated from bias expressed by following equation [29]:

$$\text{Bias}\ (\%) = \frac{|\bar{x}-\mu_0|}{\mu_0} \times 100,$$

where $\bar{x}$ is the predicted copper concentration and μ_0 is the reference value. Bias is a quantitative term describing the difference between the average of measurements made on the same object and its true value [29, 30]. The bias was calculated for two methods (using combined and partial calibrations) and for two LIBS systems. As one sees from table 3, the lower bias values, *i.e.* more accurate estimation of the copper content, are observed when partial calibration curves are used for the quantification. We used the median of the bias values to quantify the difference between the two methods. The median bias is 27.3% for the combined calibration and 7.7% for the partial calibration curves of system 1 data and 89% and 11% respectively for the system 2 data.

Figure 9 shows the correlation plots of the predicted versus certified concentrations for the combined (a) and partial (b) calibration curve methods. This visualization of the predicted copper contents strengthens the recommendation to use the partial calibration curves to obtain more reliable quantitative analysis, when the coefficient of determination of the measurement is R^2 = 0.955for system 1 and R^2 = 0.949 for system 2 respectively. Moreover, As seen from the table, the results for systems 1 and 2 are similar.

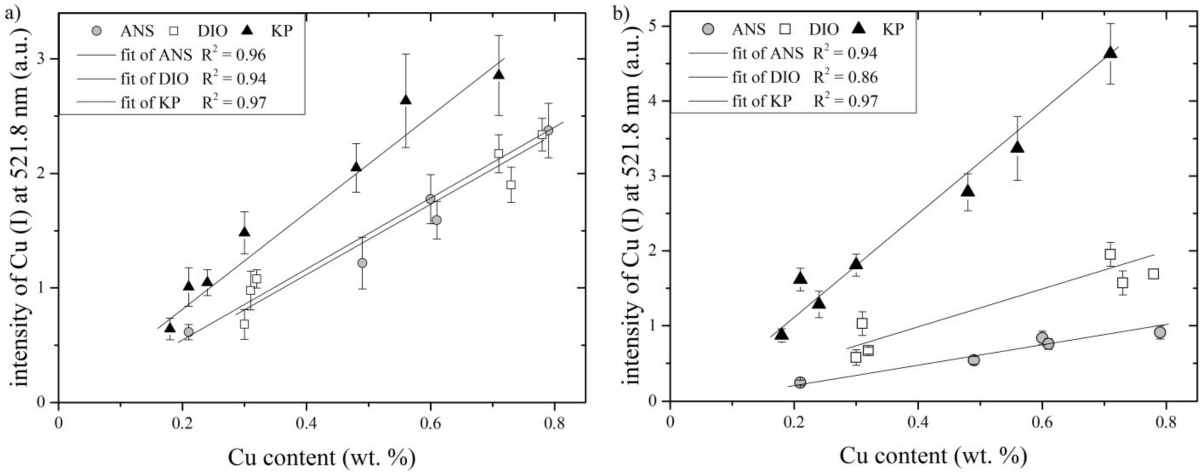


Figure 8, separate calibration curves formed after PCA analysis a) for system 1 data set, b) for system 2 data set.

Table 3, copper content calculated utilizing individual and combined calibration curves.

sample	Cu content / wt. %								
	certified	System 1				System 2			
		combined calibration	bias	partial calibration	bias	combined calibration	bias	partial calibration	bias
ANS2	0,49	0,38 ± 0,09	21,9%	0,45 ± 0,08	8,1%	0,96 ± 0,11	96,1%	0,43 ± 0,04	11,9%
ANS3	0,6	0,6 ± 0,08	0,3%	0,63 ± 0,07	5,7%	0,06 ± 0,06	90,0%	0,71 ± 0,08	19,0%
ANS4	0,61	0,52 ± 0,06	14,3%	0,55 ± 0,05	9,2%	0,01 ± 0,05	98,1%	0,62 ± 0,06	2,4%
ANS5	0,79	0,84 ± 0,09	6,0%	0,84 ± 0,08	6,4%	0,15 ± 0,05	81,0%	0,7 ± 0,06	10,8%
ANS6	0,21	0,14 ± 0,03	32,7%	0,23 ± 0,02	10,2%	0,5 ± 0,01	136,8%	0,19 ± 0,03	9,9%
DIO2	0,71	1,13 ± 0,07	59,5%	0,86 ± 0,05	21,0%	0,9 ± 0,12	26,1%	0,92 ± 0,08	29,2%
DIO3	0,73	0,64 ± 0,06	12,4%	0,68 ± 0,05	7,2%	0,58 ± 0,11	20,1%	0,65 ± 0,07	11,2%
DIO4	0,78	0,82 ± 0,06	5,4%	0,73 ± 0,04	6,3%	0,67 ± 0,05	13,9%	0,69 ± 0,03	11,1%
DIO5	0,32	0,33 ± 0,03	2,9%	0,34 ± 0,02	6,6%	0,65 ± 0,02	104,2%	0,31 ± 0,03	2,0%
DIO6	0,31	0,29 ± 0,07	7,0%	0,33 ± 0,05	6,6%	0,18 ± 0,12	40,6%	0,37 ± 0,05	19,4%
DIO7	0,3	0,16 ± 0,05	45,3%	0,26 ± 0,05	14,5%	0,77 ± 0,06	157,9%	0,26 ± 0,04	13,9%
KP1	0,24	0,32 ± 0,04	34,1%	0,23 ± 0,03	4,4%	0,51 ± 0,05	110,5%	0,26 ± 0,01	6,4%
KP2	0,18	0,16 ± 0,04	12,8%	0,17 ± 0,04	7,4%	0,38 ± 0,08	112,8%	0,16 ± 0,02	11,8%
KP3	0,21	0,31 ± 0,06	47,2%	0,23 ± 0,04	8,2%	0,31 ± 0,03	49,1%	0,23 ± 0,01	11,7%
KP8	0,48	0,71 ± 0,08	48,7%	0,47 ± 0,05	1,8%	0,91 ± 0,11	89,3%	0,44 ± 0,04	9,2%
KP9	0,71	1,07 ± 0,15	51,3%	0,77 ± 0,09	8,5%	1,15 ± 0,13	62,0%	0,77 ± 0,07	7,9%
KP10	0,56	0,96 ± 0,16	71,7%	0,61 ± 0,1	9,4%	1,14 ± 0,19	104,2%	0,53 ± 0,06	6,2%
KP11	0,3	0,49 ± 0,07	63,8%	0,33 ± 0,04	9,8%	0,49 ± 0,07	64,1%	0,29 ± 0,02	3,8%
		median	27.3%	median	7.7%	median	89.7%	median	11%

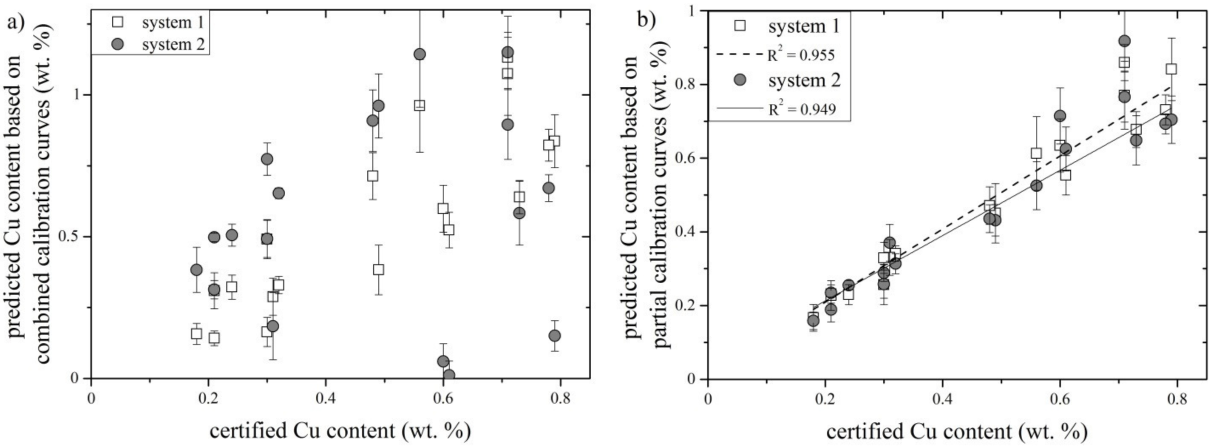


Figure 9, predicted Cu content for data of both LIBS systems from a) combined calibration curves, b) partial calibration curves.

Conlusion

Based on the presented study, LIBS can satisfy the needs of the mining industry when LIBS measurements are processed with PCA. In our study, two different LIBS systems were utilized to compare the compatibility of results obtained from suggested process for data analysis. It was demonstrated that PCA can be employed for the classification of samples and the detection of outliers prior to construction of partial calibration curves. It is shown that discrimination of the samples into classes based on the differences in the content of their matrix elements (Al, Ca, Na, and Si) improves the quality of quantitative analysis. The bias in concentrations predicted from partial calibration curves (constructed for individual rock types) was lower than that obtained with the calibration curve constructed from the whole data set. However, further study should be conducted in order to create more robust data sets. We recommend constructing the calibration curve only for samples with similar composition of matrix elements in order to avoid the matrix effect.

Acknowledgement

P.P. gratefully acknowledges prior and ongoing support from the AiF ZiM and the support of Brno University of Technology on the frame of grant FSI-S-11-22 (Application of advanced optical methods).

References

[1] N. J. McMillan, R. S. Harmon, F. C. DeLucia and A. W. Miziolek, “Laser-induced breakdown spectroscopy analysis of minerals: Carbonate and silicates,” *Spectroschimica Acta Part B,* vol. 62, pp. 1528-1536, 2007.

[2] L. Arenas, M. Ortega, M. J. García-Martínez, E. Querol and J. F. Llamas, “Geochemical characterization of the mining district of Linares (Jaen, Spain) by means of XRF and ICP-AES,” *Journal of Geochemical Exploration,* vol. 108, pp. 21-26, 2011.

[3] D. A. Cremers and L. J. Radziemski, Handbook of Laser-Induced Breakdown spectroscopy, New York, USA: John Wiley & Sons, Ltd., 2006.

[4] A. W. Miziolek, V. Palleschi and I. Schechter, Laser-Induced Breakdown Spectroscopy (LIBS) Fundamentals and Applications, Cambridge, UK: Camridge University Press, 2006.

[5] A. Streckeisen, “To each plutonic rock its proper name,” *Earth Science Reviews,* vol. 12, pp. 1-33, 1976.

[6] K. Novotný, J. Kaiser, M. Galiová, V. Konečná, J. Novotný, R. Malina, M. Liška, V. Kanický and V. Otruba, “Mapping of different structures on large area of granite samples using laser-ablation based analytical techniques, an exploratory study,” *Spectrochimica Acta Part B,* vol. 63, pp. 1139-1144, 2008.

[7] J. Lameyre and P. Bowen, “Plutonic rock types series: discrimination of various granitoid and related rocks,” *Journal of Volcanology and Geothermal Research,* vol. 14, pp. 169-186, 1982.

[8] R. S. Harmon, R. E. Russo and R. R. Hark, “Applications of laser-induced breakdown spectroscopy for geochemical and environmental analysis: A comprehensive review,” *Spectroschimica Acta Part B,* vol. 87, pp. 11-26, 2013.

[9] R. S. Harmon, F. C. DeLucia, C. E. McManus, N. J. McMillan, T. F. Jenkins, M. E. Walsh and A. W. Miziolek, “Laser-induced breakdown spectroscopy - An emerging chemical sensor technology for real-time field-portable, geochemical, mineralogical, and environmental applications,” *Applied Geochemistry,* vol. 21, pp. 730-747, 2006.

[10] R. S. Harmon, J. Remus, N. J. McMillan, C. McManus, L. Collins, J. L. Gottfried Jr., F. C. DeLucia and A. W. Miziolek, “LIBS analysis of geomaterials: geochemical fingerprint for the rapid analysis and discrimination of minerals,” *Applied Geochemistry,* vol. 24, pp. 1125-1141, 2009.

[11] R. T. Wainer, R. S. Harmon, A. W. Miziolek, K. L. McNesky and P. D. French, “Analysis of environmental lead contamination: comparison of LIBS field and laboratory instruments,” *Spectrochimica Acta Part B,* vol. 56, pp. 777-793, 2001.

[12] J. L. Gottfried, R. S. Harmon, F. C. DeLucia Jr. and A. W. Miziolek, “Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification,” *Spectrochimica Acta Part B,* vol. 64, no. 10, pp. 1009-1019, 2009.

[13] R. S. Harmon, K. M. Shughue, J. J. Remus, M. A. Wise, L. J. East and R. R. Hark, “Can the provenance of the conflict minerals columbite and tantalite be ascertained by laser-induced breakdown spectroscopy?,” *Analytical and Bioanalytical Chemistry,* vol. 400, no. 10, pp. 3377-3382, 2011.

[14] B. Bousquet, J. B. Sirven and L. Canioni, “Towards quantitative laser-induced breakdown spectroscopy analysis of soil samples,” *Spectrochimica Acta Part B,* vol. 62, no. 12, pp. 1582-1589,

2007.

[15] G. Vítková, K. Novotný, L. Prokeš, A. Hrdlička, J. Kaiser, J. Novotný, R. Malina and D. Prochazka, "Fast identification of biominerals by means of stand-off laser-induced breakdown spectroscopy using linear discriminant analysis and artificial neural networks," *Spectrochimica Acta Part B,* vol. 73, pp. 1-6, 2012.

[16] C. A. Munson, F. C. DeLucia Jr., T. Piehler, K. L. McNesby and A. W. Miziolek, "Investigation of statistic strategies for improving the discriminating power of laser-induced breakdown spectroscopy for chemical and biological warfare agent stimulants," *Spectrochimica Acta Part B,* vol. 60, pp. 1217-1224, 2005.

[17] M. Z. Martin, S. Allman, D. J. Brice, R. C. Martin and N. O. Andre, "Exploring laser-induced breakdown spectroscopy for nuclear analysis and in-situ applications," *Spectrochimica Acta Part B,* vol. 74, pp. 177-183, 2012.

[18] M. Z. Martin, N. Labbé, T. G. Rials and S. D. Wullschleger, "Analysis of preservative-treated wood by multivariate analysis of laser-induced breakdown spectroscopy spectra," *Spectrochimica Acta Part B,* vol. 60, pp. 1179-1185, 2005.

[19] S. I. Gornushkin, I. B. Gornushkin, J. M. Anzano, B. W. Smith and J. D. Winefordner, "Effective normalization technique for correction of matrix effects in laser-induced breakdown spectroscopy detection of magnesium in powdered samples," *Applied spectroscopy,* vol. 56, pp. 433-436, 2002.

[20] I. B. Gornushkin, A. Ruiz-Medina, J. M. Anzano, B. W. Smith and J. D. Winefordner, "Identification of particulate minerals by correlation analysis using a microscope laser induced breakdown spectrometer," *Journal of Analytical Atomic Spectrometry,* vol. 15, pp. 581-586, 2000.

[21] S. Verboven, "LIBRA: a MATLAB library for Robust Analysis," 2 Aug 2012. [Online]. Available: http://wis.kuleuven.be/stat/robust/Programs/LIBRA. [Accessed 9 Oct 2013].

[22] A. M. Ollila, J. Lasue, H. E. Newson, R. A. Multari, R. C. Wiens and S. M. Clegg, "Comparison of two partial least squares-discriminant analysis algorithms for identifying geological samples with the ChemCam laser-induced breakdown spectroscopy instrument," *Applied Optics,* vol. 51, pp. 130-142, 2012.

[23] R. G. Brereton, Chemometrics for pattern recognition, Bristol, UK: John Wiley and Sons, Ltd., 2009.

[24] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström and S. Wold, Multi- and megavariate data analysis. Part I: Basic principles and applications, Umeå, Sweden: Umetrics AB, 2006.

[25] H. Martens and T. Naes, Multivariate calibration, Chichester, UK: John Wiley & Sons Ltd, 1989.

[26] S. Laville, M. Sabsabi and F. Doucet, "Multi-elemental analysis of solidified mineral melt samples by Laser-Induced Breakdown Spectroscopy coupled with a linear multivariate calibration,"

Spectrochimica Acta B, vol. 62, pp. 1557-1566, 2007.

[27] D. Death, A. P. Cunningham and L. J. Pollard, “Multi-element and mineralogical analysis of mineral ores using laser induced breakdown spectroscopy and chemometric analysis,” *Spectrochimica Acta B,* vol. 64, pp. 1048-1058, 2009.

[28] R. A. Multari, D. A. Cremers, J. M. Dupre and J. E. Gustafson, “The use of laser-induced breakdown spectroscopy for distinguishing between bacterial pathogen species and strains,” *Applied spectroscopy,* vol. 64, pp. 750-759, 2010.

[29] S. L. R. Ellison, V. J. Barwick and T. J. D. Farrant, Practical statistics for the analytical scientist, 2. ed., Cambridge, UK: RSC Publishing, 2009.

[30] NIST/SEMATECH, “e-Handbook of Statistical Methods,” National Institute of Standards and Technology, 30 10 2013. [Online]. Available: http://www.itl.nist.gov/div898/handbook/. [Accessed 3 7 2014].